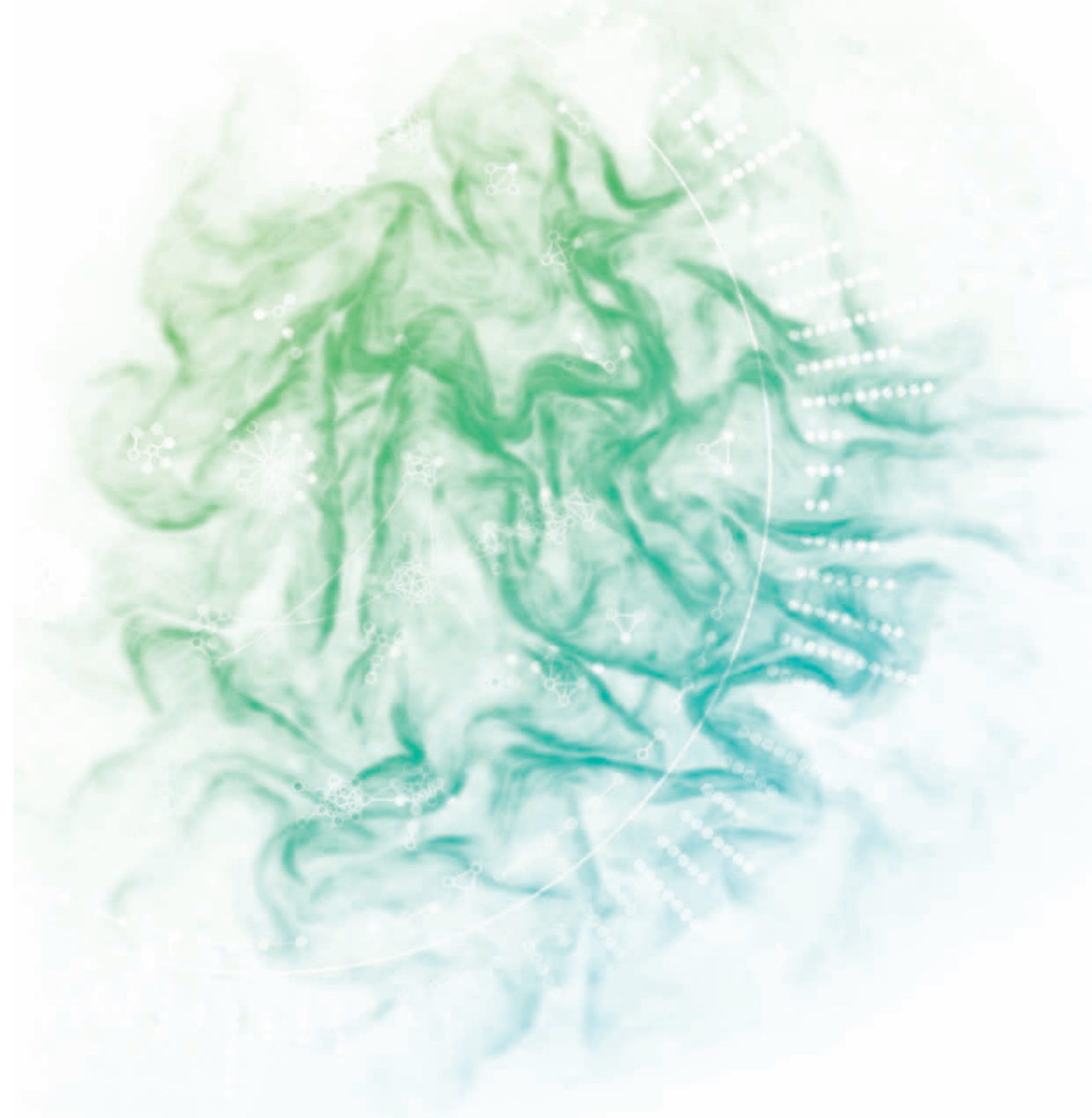


European Molecular Biology Laboratory

Research at a Glance 2011

EMBL

European Molecular Biology Laboratory

Research at a Glance 2011

EMBL

Contents

Foreword by EMBL's Director General

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 has been continuously ranked as one of the top research institutes worldwide.

EMBL is Europe's only intergovernmental laboratory in the life sciences and as such its missions extend beyond performing cutting-edge research in molecular biology. It also offers services to European scientists, most notably in the areas of bioinformatics and structural biology, provides advanced training to researchers at all levels, develops new technologies and instrumentation and actively engages in technology transfer for the benefit of scientists and society. It is the synergy between these diverse missions and the stringent, external quality control to which we submit all our activities regularly, that allows EMBL to stay at the forefront and achieve world-class standards.

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 while drawing on the diverse techniques and expertise available in the institute. But what really distinguishes EMBL is the large number of inter-unit collaborations, bringing people with common interests but distinct skills together to tackle ambitious projects. Increasingly, our young scientists come with physics, chemistry, engineering, mathematics and computer science backgrounds, contributing new perspectives and the complementary expertise required to unravel the complexity of living systems.

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, an important, diverse and interdisciplinary range of core facilities as well as 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 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 organised 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 in 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 organisationand 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 membrane is continuous with the ER and perforated by nuclear pore complexes (NPCs). 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 attached to 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 systems, we are using somatic mammalian cells for mitosis, as well as mouse oocytes for meiosis, in which we study the asymmetric division they undergo to become a fertilisable egg.

In somatic cells, we could show that nuclear membrane formation originates from the endoplasmic reticulum. We found that mitotic nuclear breakdown and reformation is initiated by the ordered dis- and reassembly of NPCs. Interestingly, we found that NPCs are assembled along a different pathway assembly during nuclear growth in interphase. Recently, we have identified many new cell division genes by screening the entire human genome using time-lapse microscopy (figure 1). In animal oocytes, we discovered that asymmetric transport of chromosomes to the cell surface is mediated by a contractile F-actin network rather than by microtubules.

Future projects and goals

The objective of our future work is to gain comprehensive mechanistic insight into cell cycle remodelling of the nucleus. For the interphase nucleus, we are currently focusing on the mechanism of nuclear growth, as well as chromosome architecture and dynamics, where we recently proposed a fractal organisation of chromosomes in interphase nuclei (figure 2). For mitosis, we are aiming to achieve a systems level understanding and assay all relevant proteins identified in our genome-wide survey of human mitotic genes in living cells. To this end we are continuously automating advanced fluorescence imaging techniques. In oocytes we are pursuing the molecular mechanism of actin-mediated chromosome transport as well as homologous chromosome segregation.

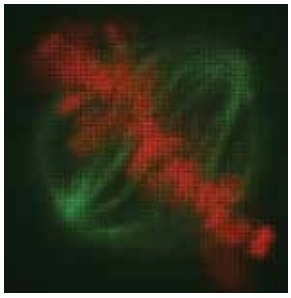
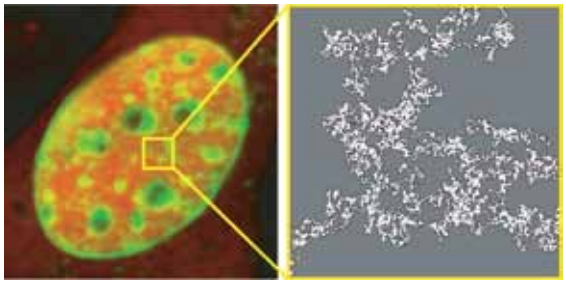


Figure 1: Mitotic spindle composed of thousands of microscopy images of human cells in which individual genes were silenced. Chromosomes (red) are made of images from genes that affect their segregation, while the mitotic spindle (green) is composed of images from genes affecting its assembly.

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).



Selected references

Dultz, E. & Ellenberg, J. (2010). Live imaging of single nuclear pores reveals unique assembly kinetics and mechanism in interphase. *J. Cell. Biol.*, 191, 15-22

Neumann, B., *et al.* (2010). Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature*, 464, 721-727

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

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.

Claude Antony

Cellular electron tomography of cells

Previous and current research

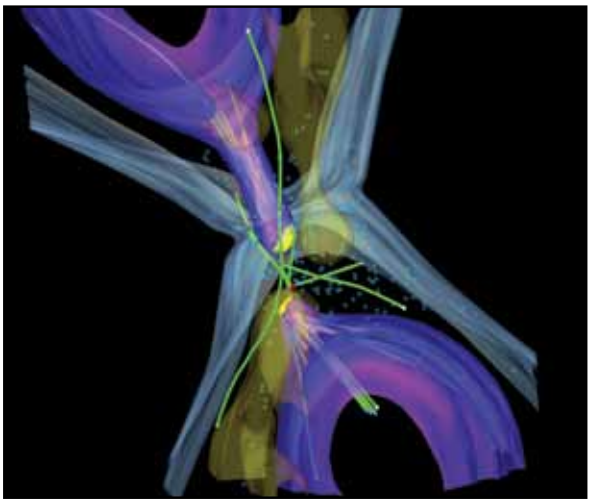
The focus of interest in the team is the organisation of microtubular cytoskeleton arrays in both fission yeast and budding yeast as well as in the *Xenopus* mitotic spindle. For this purpose we use electron tomography (ET), which not only allows the reconstruction, modelling and quantification of subcellular elements, but also enables the visualisation of a number of fine structural features which would not be detectable by conventional EM. We are equipped with a Tecnai F30 tomography microscope (FEI), which is mostly devoted to plastic electron tomography projects.

Our work has focussed on key areas, such as resolving the 3D organisation of microtubule arrays 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 analysed the morphology of the microtubule ends by ET so as to determine the type of ends of the growing microtubules. This analysis was carried out by Johanna Höög in collaboration with Damian Brunner's group (both formerly EMBL). We also investigated microtubule bundling factors in fission yeast, Ase1 (a non-motor homodimer protein) and further identified Dis1p (XMAP215 homologue) as an alternative microtubule bundling factor that 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. This work was done by Helio Roque in collaboration with Damian Brunner's group.

Another project is to investigate the role and organisation of microtubules in the budding yeast mating pathway, led by Romain Gibeaux (PhD student) in collaboration with Michael Knop's group (EMBL). The project concerns the morphological and molecular analysis of the karyogamy process in the budding yeast mating pathway. We recently focused on the nuclear congression by resolving the detailed microtubule organisation and microtubule connection with cells' organelles. These experiments are carried out by electron tomography. In parallel we are searching for molecular players associated with microtubules that are also involved in force generation by designing a set of mutant analysis coupled with live-cell fluorescence microscopy. This work is currently ongoing.

Future projects and goals

In collaboration with in-house and external research groups (Nédélec group, page 14) and Rebecca Heald, UCB) we run a major project aiming to reconstruct, at EM resolution, the *Xenopus laevis* mitotic spindle assembled from egg extracts. Samples are cryofixed by high pressure freezing and prepared for tomography acquisition. The large-scale reconstruction of such a huge structure, or parts of it, is being 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, we are interested in understanding the structural organisation of microtubules in the midzone of the spindle and at the poles. Xavier Heiligenstein (PhD student) and Erin Tranfield (postdoc) are currently working on this project.



Nuclear congression in the budding yeast. 3D reconstruction of mating yeast cells after plasma (blue-green) fusion has occurred. The two nuclei are brought together by forces exerted by microtubules emanating from the two SPBs of opposite cells. (Picture by Romain Gibeaux; collaboration with Michael Knop's group. Bar, 300 nm).

Selected references

Tängemo, C., Ronchi, P., Colombelli, J., Haselmann, U., Simpson, J. C., Antony, C., Stelzer E.H.K., Pepperkok, R. and Reynaud É.G. (2011). A Novel Laser Nanosurgery Approach Provides Support for *de novo* Golgi Biogenesis in Mammalian Cells. *J.Cell Sc*. (in press)

Höög, J.L., Huisman, S.M., Sebö-Lemke, S, Sandblad, L., McIntosh J.R., Antony, C. and Brunner, D. (2011). Electron Tomography Reveals a Flared Morphology on Growing Microtubule Ends. *J.Cell Sc.* (in press).

Kitamura, E., Tanaka, K., Komoto, S., Kitamura, Y., Antony, C. & Tanaka, T.U. (2010). Kinetochores Generate Microtubules with Distal Plus Ends: Their Roles and Limited Lifetime in Mitosis. *Dev. Cell*, 18, 248-259

Roque, H., Ward, J.J., Murrells, L., Brunner, D. & Antony, C. (2010). The fission yeast XMAP215 homolog Dis1p is involved in microtubule bundle organization. *PLoS One*, 5, e14201

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 led 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 mathematical models that will accurately simulate this complex *in vivo* morphogenesis process.

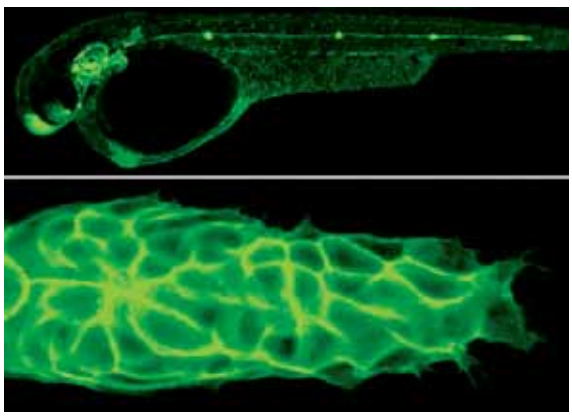


Figure 1: The zebrafish migrating lateral line organ allows collective migration to be easily studied in vivo.

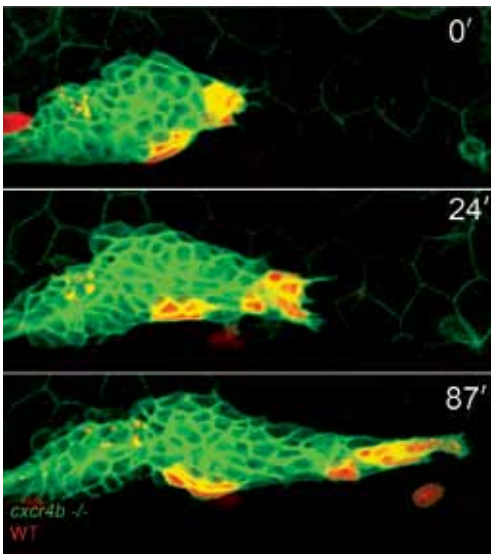


Figure 2: Transplanted wild-type cells (red) rescue the migration of cxcr4 mutant primordia (green).

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 cohesin'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?

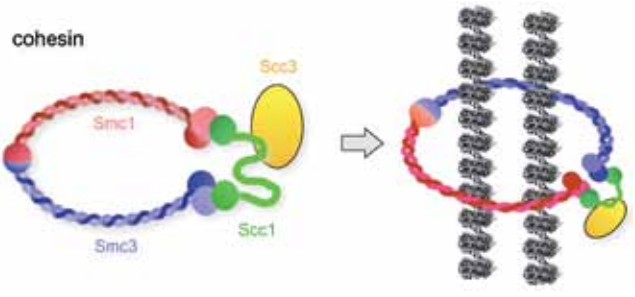


Figure 1 (top): Model of the cohesin ring holding sister chromatids together.

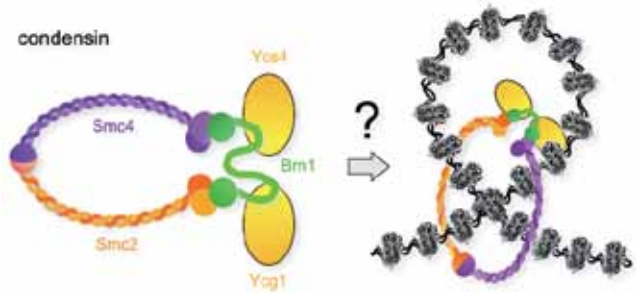


Figure 2 (bottom): Model of the condensin ring structuring chromosomes.

Selected references

Kurze, A., Michie, K.A., Dixon, S.E., Mishra, A., Itoh, T., Khalid, S., Strmecki, L., Shirahige, K., Haering, C.H., Lowe, J. & Nasmyth, K. (2010). A positively charged channel within the Smc1/Smc3 hinge required for sister chromatid cohesion. *EMBO J.*, 30(2), 364-78

J, Nasmyth, K. & Haring, 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

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.

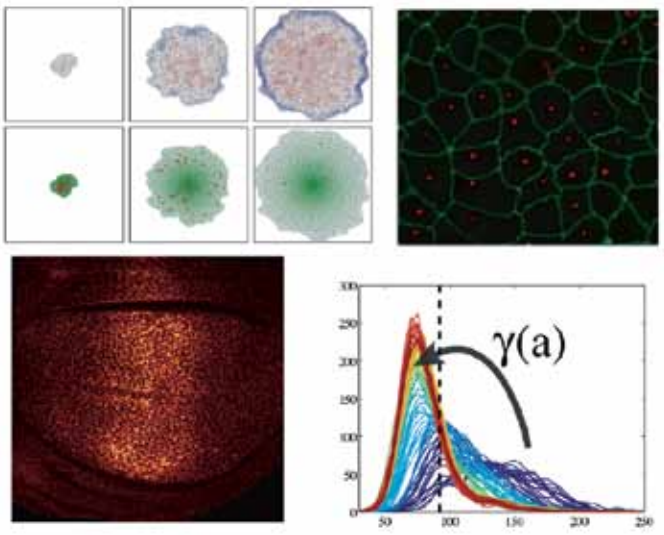
We have recently investigated the interplay between the growth of *Drosophila* wing imaginal discs and the formation of the Dpp morphogen gradient. Our results suggest a new scenario of size determination, where disc size is determined relative to the fixed morphogen distribution. Our model shows that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

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 a modular organisation, where the different modules form dynamic molecular machines that drive processes such as signaling, 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 that is composed of over 50 different protein species and of several thousand individual protein molecules. We aim to understand the processes that regulate the assembly of the endocytic machinery, and the mechanisms that determine the location and timing of endocytic events in the cell.

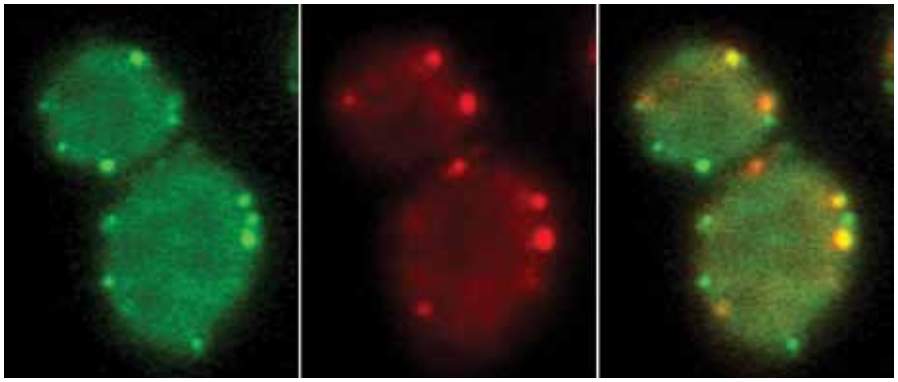
Our main experimental organism is budding yeast *Saccharomyces cerevisiae*. In our studies we use quantitative live-cell imaging methods (e.g. particle tracking, FRAP, FCS/FCCS, high-throughput microscopy) in combination with the powerful yeast genetics. We also use correlated light and electron microscopy to gain nanometer-scale information about the endocytic structures, and biochemistry to characterise protein-protein and protein-lipid interactions.

Future projects and goals

We are interested in the mechanisms that initiate the assembly of the endocytic machinery and regulate the precise timing of the sequential stages of the assembly. The spatial distribution of the endocytic events is tightly coupled to the cell cycle and to the overall polarity of the cell. The spatially regulated initiation of endocytic events is critical for determining the cellular distribution of endocytosis.

We are also studying the mechanisms of selective recruitment of cargo molecules into the endocytic vesicles. The recruitment of cargo proteins is tightly regulated by a family of endocytic adaptors. We want to understand how this adaptor system integrates environmental and intracellular signals in deciding which cargoes to recruit.

Furthermore, we want to understand how actin functions to promote endocytic vesicle budding. In yeast endocytosis is strictly dependent on actin polymerisation, but the mechanisms by which actin drives vesicle budding are not well understood. We are currently studying the molecular basis of the coupling between the actin cytoskeleton and the endocytic membrane. The core membrane trafficking events, such as the clathrin-mediated endocytosis, are elemental cellular 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 mechanisms that we unravel in yeast cells will be applicable to eukaryotes in general.



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

Selected references

Brach, T., Specht, T. & Kaksonen, M. (2011). Reassessment of the role of plasma membrane domains in the regulation of vesicular traffic in yeast. *J. Cell. Sci.*, 124, 328-337

Kukulski, W., Schorb, M., Welsch, S., Picco, A., Kaksonen, M. & Briggs, J.A. (2011). Correlated fluorescence and 3D electron microscopy with high sensitivity and spatial precision. *J. Cell. Biol.*, *192, 111-9*

Gallego, O. *et al* (2010). A systematic screen for protein-lipid interactions in Saccharomyces cerevisiae. *Mol. Syst. Biol.*, 6, 430

Péter Lénárt

PhD 2004, EMBL and University of Heidelberg

Postdoctoral research at the Institute of Molecular Pathology (IMP), Vienna

Staff scientist at EMBL since 2008

Group leader since 2011

Cytoskeletal dynamics and function in oocytes

Previous and current research

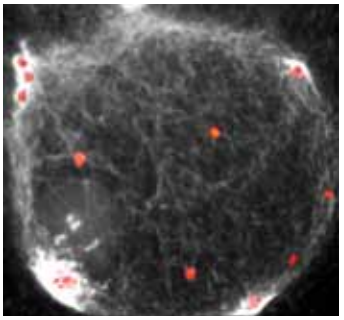
All animal life begins with the fusion of sperm and egg. Our research is focused on the egg cell, specifically investigating how the fertilisable egg develops from the oocyte through meiotic divisions. Oocytes are exceptionally large cells, with diameters up to millimetres in size, because they store large amounts of nutrients to support embryonic development. Therefore, in oocytes and eggs, the cytoskeleton has to transport organelles, separate chromosomes, and in general organise cellular architecture in a very large cytoplasm. How the cytoskeleton adapts to this unusual size, and how these mechanisms are different from those in small somatic cells, is largely unknown.

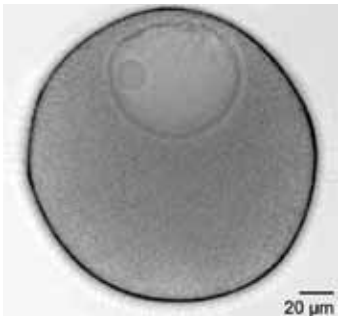
We use starfish oocytes as a model system, because they are easy to handle, complete meiosis rapidly, and develop simply in sea water at room temperature. Most importantly, starfish oocytes are transparent, ideal for high resolution imaging of cytoskeletal dynamics in live cells, and available in large quantities and synchronous populations - ideal for biochemical analysis. We use confocal microscopy to image live oocytes and employ computational image analysis tools to extract quantitative parameters from these 3D time-lapse datasets. Parameters such as local concentrations or velocities of cellular components then provide a quantitative assay for the biological process and at the same time serve as inputs for computational models of cytoskeletal dynamics. Model predictions are then tested in perturbation experiments using physical (e.g. laser microsurgery) or molecular (e.g. RNAi knock-down) manipulations. Biochemistry, in combination with the imaging assays, is used to identify the key molecular components in the process.

We have recently shown that meiotic chromosomes scattered in the large oocyte nucleus are collected by an actin meshwork and transported to the spindle, whose short microtubules cannot reach the chromosomes directly, as they do in somatic cells. This novel actin-based chromosome transport system forms as the nuclear envelope breaks down and fills the nuclear space with an actin meshwork physically entrapping chromosomes. We could show that the actin meshwork contracts homogeneously, however because it is mechanically anchored to the cell cortex this homogeneous contraction is translated into directional transport towards the cortex where the spindle forms. By understanding the mechanism of chromosome transport essential to oocyte division and fertility, our studies revealed a novel design principle for a cytoskeletal 'transport machine' that is very different from previously known mechanisms of actin-driven intracellular transport.

Future projects and goals

Our immediate goals include determining the detailed structure of the F-actin meshwork, understanding the molecular mechanism of meshwork contraction and identifying the mechanisms by which chromosomes attach to the meshwork. We plan to employ high-resolution imaging methods, including electron tomography and super-resolution light microscopy, to resolve single actin-filaments. In addition, we will identify, localise and perturb molecules regulating actin filament dynamics that combined with high-resolution imaging will allow us to address the underlying molecular mechanisms. Longer term, we are interested in related cytoskeletal processes that occur in oocytes, eggs and early embryos, with the overall goal to mechanistically understand the organisational principles of the actin and microtubule cytoskeleton.





Right: The actin filament network (gray) embedding the chromosomes (red).

Far right: Transparent starfish oocytes are uniquely suited for imaging meiotic divisions.

Selected references

Lénárt, P., Petronczki, M., Steegmaier, M., Di Fiore, B., Lipp, J.J., Hoffmann, M., Rettig, W.J., Kraut, N., Peters, J.M. (2007) The small-molecule inhibitor BI 2536 reveals novel isights into mitotic roles of polo-like kinase 1. *Curr. Biol.*, 17, 304-15

Lénárt, P., Bacher, C.P., Daigle, N., Hand, A.R., Eils, R., Terasaki, M., Ellenberg, J. (2005) A contractile nuclear actin network drives chromosome congression in oocytes. *Nature*, 436, 812-8

Lénárt, P., Rabut, G., Daigle, N., Hand, A.R., Terasaki, M., Ellenberg, J. (2003) Nuclear envelope breakdown in starfish oocytes proceeds by partial NPC disassembly followed by a rapidly spreading fenestration of nuclear membranes. *J. Cell. Biol.*, 1 60, 1055-68

François Nédélec

PhD 1998, Université Paris 11.
Postdoctoral research at EMBL.
BioMS Group Leader since 2005.
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, F. (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, while the chromosomes are stable, all the other components of a spindle – polar filaments called microtubules and associated proteins – are in rapid turnover. Microtubules grow, shrink and disappear in a matter of minutes and proteins continuously and stochastically bind and unbind. The resulting assembly is highly dynamic and yet stable and remarkably precise; it can remain steady for hours, until it eventually applies the balanced forces necessary to position and segregate the chromosomes exactly.

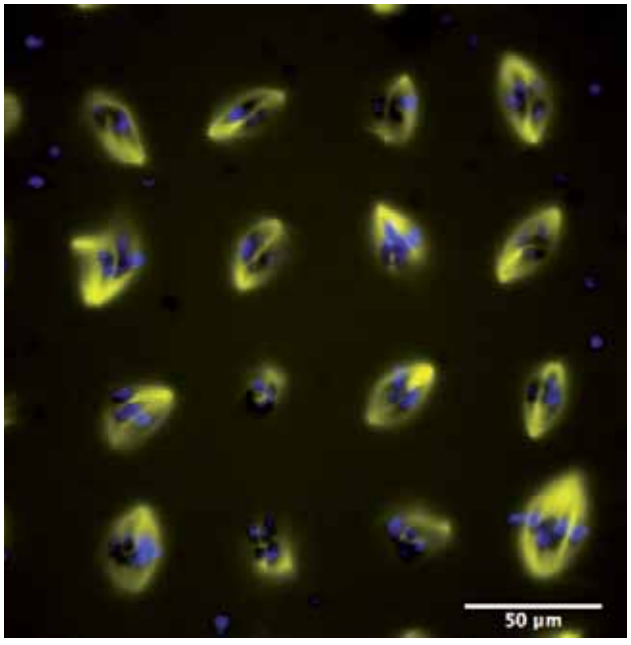
The spindle is thus a fascinating structure, which illustrates a central question in biology: how can the uncoordinated and inevitably imperfect actions of proteins and other molecules collectively fulfill the biological needs with the required accuracy?

Today, understanding biological phenomena from their multiple biological components - systems biology - is a cutting-edge research topic. The collective behaviour in biology is more than a simple statistical average. It is a challenging problem for many reasons: 1) the diversity of molecular players is enormous; 2) their interactions are often dynamic and out-of-equilibrium; 3) the properties of the constituents have been selected by natural evolution.

We approach this topic 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 specific proteins, 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 also 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.

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 modelling 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.



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

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.

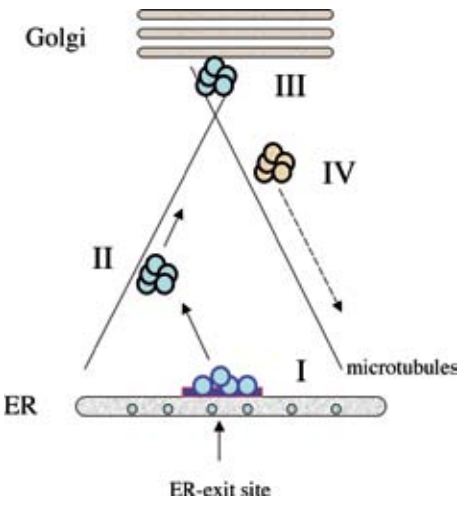
The specific questions we are presently addressing in this context are: 1) What are the mechanisms underlying the regulation of ER-exit sites biogenesis and function? 2) How are ER exit and microtubule mediated ER to Golgi transport coupled at the molecular level? 3) What are the mechanisms of Golgi biogenesis? 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 are able to 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 returned 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

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

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

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

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

Mall, M.A., Button, B., Johannesson, B., Zhou, Z., Livraghi, A., Caldwell, R.A., Schubert, S.C., Schultz, C., *et al.* (2010). Airway surface liquid volume regulation determines different airway phenotypes in liddle compared with betaENaC-overexpressing mice. *J. Biol. Chem.*, 285, 26945-26955

Schultz, C. (2010). Challenges in studying phospholipid signaling. *Nat. Chem. Biol.*, 6, 473-475

Subramanian, D., Laketa, V., Muller, R., Tischer, C., Zarbakhsh, S., Pepperkok, R. & Schultz, C. (2010). Activation of membrane-permeant caged PtdIns(3)P induces endosomal fusion in cells. *Nat. Chem. Biol.*, 6, 324-326

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

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 such as cyclic nucleotides, inositol phosphates and phosphoinositides to membrane-permeant, bioactivatable derivatives ('prodrugs').

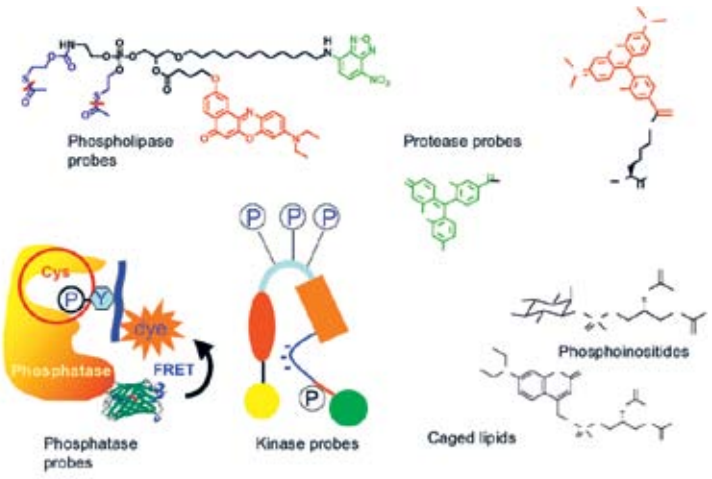
At EMBL, our interest has shifted to the basic signalling network underlying epithelial secretion as well as receptor endocytosis and recycling. We have developed a wide range of fluorescent reporter molecules, either genetically encoded or as small molecule fluorescent probes (see figure). With these sensors, we hope to provide a more complete picture of the signalling network and to help find compounds that might be beneficial for unravelling 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 novel 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 2011, 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.*, 2010). Vesicle trafficking and endocytosis is investigated in collaboration with Rainer Pepperkok's group (page 15). Finally, we are interested in how the plasma membrane is repaired after physical impact. We combine fluorescence microcopy of tagged proteins with electron microscopy (correlative microscopy), the latter a 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.



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.

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

Besse, F., Lopez de Quinto, S., Marchand, V., Trucco, A. and Ephrussi, A. (2009) Drosophila PTB promotes fomation of high-order RNP particles and represses oskar translation. *Genes and Development*, 23, 195-208

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

Chekulaeva, M., Hentze, M.W. and 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. and Ephrussi A. (2004). Splicing of *oskar* RNA in the nucleus is coupled to its cytoplasmic localization. *Nature,* 28, 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 localised in the egg specify the body axes and patterning of the future embryo. During oogenesis, 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.

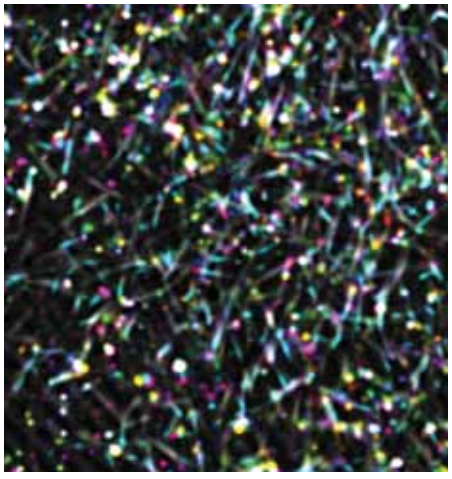
Drosophila is ideally suited for genetic, biochemical, and cell biological investigation of the processes of cell polarisation, mRNA localisation and translational control. We make use of this model system to study: 1) cytoskeletal polarisation; 2) the assembly of the RNA transport complexes and their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

Future projects and goals

Combining genetics, biochemistry and a broad spectrum of cell biological and imaging approaches we are investigating:

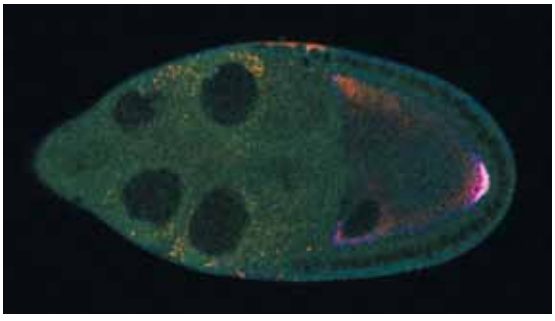
- The mechanisms underlying cell polarisation;
- the role of the cytoskeleton and motors in mRNA transport;
- the architecture of transport RNPs: the cis-acting RNA elements and interacting proteins, and how they assemble and associate with their motor proteins to form functional RNA transport complexes;
- 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 in patterning the embryo.



Left: oskar *mRNA on the move. Time projection of a squash of ooplasm from a stage 9 oocyte imaged with TIRF microscopy.* oskar *mRNA (labelled with MS2-MCPGFP, shown in rainbow colours) utilises microtubules (labelled with mCherry-a1-tubulin and EB1-Cherry, shown in gray with cyan tips, indicating plus ends) to take fast, long linear runs.*

Below: 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 great remaining 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.

Our lab has chosen to investigate a new molecular animal model, the marine annelid *Platynereis dumerilii*. As a 'living fossil', *Platynereis* represents an ideal connecting link between vertebrates and the fast evolving protostome models, *Drosophila* and *Caenorhabditis*. Genomic resources and molecular techniques have been generated that make it a model marine invertebrate for ocean biology and for organismal systems biology. As characteristic for the *Platynereis* life cycle with different stages exploring different ecological niches, environmental influences impact directly on the organismal state (eco-devo) or are sensed via the nervous system (organismal neurobiology) and are reflected by the variation on genome, transcriptome or any other level of the cellome.

Platynereis is amenable to high-throughput imaging techniques and functional interference approaches. For example, first genetic knock-out lines have been generated. With the recent development of the PrImR (Profiling by Image Registration) resource, *Platynereis* has become the first animal model for which gene expression profiling data can be obtained in cellular resolution for the whole organism. Building on these resources, we have discovered that the *Platynereis* brain harbours sensory-associative brain parts and a neurosecretory brain centre that correspond to the vertebrate pallium and hypothalamus, respectively. These findings revolutionise our current understanding of brain evolution.

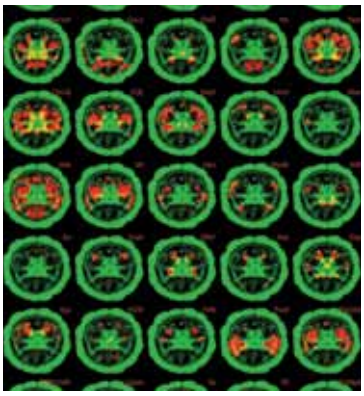
In an attempt to broaden our comparative approach, we have introduced two new model species to the lab, the lancelet amphioxus and the sea anemone *Nematostella*, representing distinct divisions of the animal kingdom: chordates and cnidarians. Amphioxus has a very simple brain uniting invertebrate- and vertebrate-like features. The *Nematostella* nervous system is very simple and thus represents a good proxy for a very early stage of brain evolution.

Future projects and goals

A clear picture is emerging that the *Platynereis* brain harbours many cell types so far known only in vertebrates, but in a much more simple and 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 brain. In a comparative eco-evo-devo approach, our aim is to gain a systems view of the *Platynereis* brain and to track the evolutionary history of all constituent cell types by identifying and investigating their evolutionary counterparts in amphioxus and sea anemone. This will involve investigations of cell type-specific gene regulatory networks as well as neurobiological and behavioural approaches.

In collaboration with the Janelia Farm Research Campus, we plan to extend the PrImR protocol to key stages of the *Platynereis* larval development and life cycle, in order to generate the first cellular resolution expression atlas for a whole animal, involving early developmental as well as differentiation stages.

Finally, we are also interested in exploring population genetics and the variability of development and differentiation in different habitats and conditions. To this end, we are collecting strains of *Platynereis* and amphioxus as part of the Tara Oceans expedition.



Above: Platynereis dumerilii *(Polychaeta, Annelida, Lophotrochozoa). Right: A collection of gene expression patterns in the* Platynereis *brain, aligned to a standardised scaffold in cellular resolution by PrImR*

Selected references

Jékely, G., Colombelli, J., Hausen, H., Guy, K., Stelzer, E., Nédélec, F., Arendt, D. (2008) Mechanism of phototaxis in marine zooplankton. *Nature*, 456, 395-9.

Christodoulou, F., Raible, F., Tomer, R., Simakov, O., Trachana, K., Klaus, S., Snyman, H., Hannon, G. J., Bork, P. and Arendt, D. (2010). Ancient animal microRNAs and the evolution of tissue identity. *Nature,* 463, 1084-88

Tomer, R., Denes, A. S., Tessmar-Raible, K. and Arendt, D. (2010). Profiling by Image Registration reveals common origin of annelid mushroom bodies and vertebrate pallium. *Cell,* 142(5), 800-9

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.

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

Timing of mammalian embryogenesis

Previous and current research

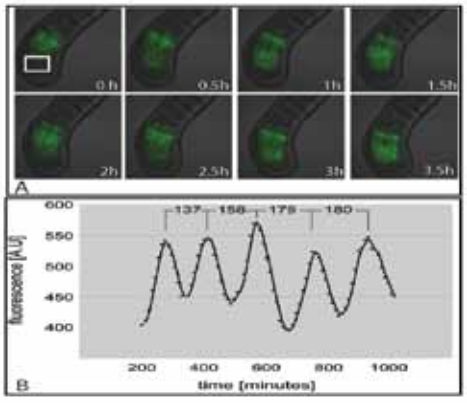
In situ *hybridisation of mouse embryo at day 9 of development. Uncx4.1 mRNA is visualised in formed somites, while Wnt3a mRNA is expressed in the posterior embryo.*

During an embryo's journey from a single cell to a complex organism, countless patterning processes unfold with remarkable precision, both spatially but also in respect to their temporal sequence, or timing. This temporal aspect of embryonic development is 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 approach further and are establishing a novel and 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 real-time 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 that are responsible for controlling these oscillations within the segmentation process. The insight gained from studying this specific oscillation phenomenon will be combined with our efforts that address the mechanisms that control the 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;
- discovery of novel oscillatory phenomena during embryogenesis.

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 intracellular trafficking during tissue morphogenesis

Previous and current research

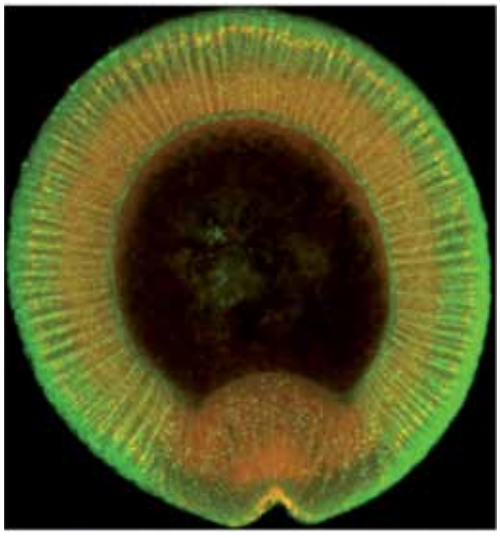
Our research focuses on how cells reorganise their cyto-architecture during tissue morphogenesis. Using a combination of imaging techniques and biochemistry, the group aims to understand how changes in protein and membrane dynamics regulate changes in cell morphology and tissue movement during development.

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 signalling 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., Elemento, O., Tavazoie, S. & Wieschaus, E.F. (2007). Unmasking activation of the zygotic genome using chromosomal deletions in the Drosophila embryo. *PLoS Biol.*, 5, e117

De Renzis, S., Yu, J., Zinzen, R. & Wieschaus, E. F. (2006). Dorsal-ventral pattern of Delta trafficking is established by a Snail-Tom-Neuralized pathway. *Dev. Cell,* 10, 257-264

De Renzis, S., Sonnichsen, B. & Zerial, M. (2002). Divalent Rab effectors regulate the sub-compartmental organization and sorting of early endosomes. *Nat. Cell Biol*., 4, 124-133

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.

Selected references

Heisler, M.G., Hamant, O., Krupinski, P., Uyttewaal, M., *et al.* (2010) Alignment between PIN1 polarity and microtubule orientation in the shoot apical meristem reveals a tight coupling between morphogenesis and auxin transport. *PLoS Biol*., 8(10),

Hamant, O., Heisler, M.G., Jonsson, H., Krupinski, P., Uyttewaal, M., Bokov, P., Corson, F., Sahlin, P., Boudaoud, A., Meyerowitz, E.M., Couder, Y. & Traas, J. (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

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 and materials 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.

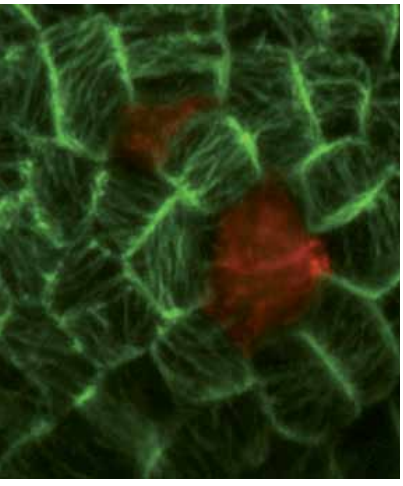
Our prime interest is in understanding patterning and how distinct patterning processes coordinate. Lateral organ formation in the model plant species *Arabidopsis thaliana* provides an ideal system for investigating such questions since organ formation involves the coordination of cell polarity, gene expression and morphogenesis. Our recent work reveals that patterns of cell polarity control both morphogenesis at the cellular level as well as at the tissue level. This integration occurs through the co-alignment of microtubule arrays with the polar localisation patterns of the auxin efflux carrier PIN1. The microtubule cytoskeleton regulates growth direction at the cellular level while PIN1 works to concentrate the hormone auxin at the tissue level to localise growth. Our data so far suggests a role for mechanical stresses in orienting these factors and current work is aimed at further investigating this possibility. Interestingly, we have also found that the patterns of cell polarity associated with organogenesis correlate spatially with particular patterns of gene expression normally associated with the dorsal and ventral cell types of lateral organs. This raises the question of whether these expression domains play a causal role in organising cell polarity patterns and, in turn, whether these polarity patterns influence dorsiventral gene expression. This rich interplay is one of our prime focuses.

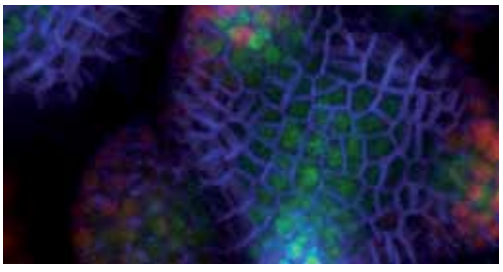
Future projects and goals

Establishment and function of dorsiventral boundaries (ERC funded): Previously we developed confocal based methods for image growing plant tissues, enabling us to obtain dynamic high-resolution data for protein localisation and gene expression (making full use of the different GFP spectral variants). By incorporating such data directly into mathematical models we aim to develop an explicit understanding of the complexity underlying the patterning processes associated with dorsiventral cell-type specification. The main questions we wish to address include: How do dorsiventral gene expression boundaries regulate organ morphogenesis and positioning (e.g. cell polarity patterns)? How are dorsiventral gene expression boundaries established and regulated?

Like animals, plants can also re-pattern their tissues in response to wounding. For instance vascular pathways get respecified around wound sites and meristems can re-reposition their stem cells and associated niches. Wounding also causes dramatic changes to dorsiventral patterning, although the mechanisms by which this occurs remain unknown. Our recent results show that cell polarity patterns respond dramatically to wounds, suggesting this cellular response may play an important role in tissue reorganisation. We aim to investigate this possibility using two-photon induced ablation and DSLM microscopy.

***In toto* transcriptomics:** Another focus for the lab is in developing single-cell transcriptomics as a tool for understanding development at the systems level. Our goal is to integrate this approach with high-resolution 3D microscopy and microfluidics techniques in order to associate spatial information with genome-wide expression data at single-cell resolution. If successful, our approach should enable a broad but detailed view of development and serve as a tool for understanding gene function on a cell-by-cell basis.





Far 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.*

Takashi Hiiragi

PhD 2000 Kyoto University, Japan

Postdoctoral research at the Max Planck Institute of Immunobiology, Freiburg, Germany

Group leader at the Max Planck Institute of Immunobiology 2002-2007

Independent group leader at the Max Planck Institute for Molecular Biomedicine, Muenster, 2007-2011

Group leader at EMBL since 2011

Systems-level understanding of early mammalian development

Previous and current research

A fundamental question in developmental biology is the mechanism by which the embryonic asymmetry is established during development. Early mammalian development is characterised by formation of the embryonic polarity and initial cell lineages in the blastocyst, composed of the inner cell mass (ICM) surrounded by one-cell layer of the trophectoderm (TE). The former lineage creates the embryo proper, while the latter yields an extra-embryonic tissue specific to mammals, the placenta. Despite its importance for understanding mammalian development and for stem cell research, the mechanism of blastocyst morphogenesis and patterning has long been elusive. When and how does the embryonic polarity develop from the totipotent egg? Is the mechanism comparable to that of non-mammalian species?

We have established a live-imaging system for mouse pre-implantation development, demonstrating high dynamicity of blastocyst morphogenesis. Our recent studies characterised key principles underlying early mammalian development: 1) absence of polarity in the egg (i.e. localised determinants play little or no role in generating the asymmetry); 2) mechanical and structural context plays a key role in morphogenesis and embryonic patterning (see figure 1); 3) stochastic processes generate dynamic heterogeneity during the patterning process (see figure 2). Collectively, an attractive hypothesis is that early mammalian embryogenesis may be, at least to some extent, a stochastic process in a particular structural context that eventually leads to self-organisation.

These features suggest that, in order to fully understand early mammalian development and evaluate our hypothesis, it will be essential to address how diverse inputs acting on every individual cell are integrated in the embryo at the systems level. Therefore we have recently established necessary tools and multi-disciplinary strategies, including: fluorescence-based gene-trap mouse lines that visualise molecular dynamics during embryonic patterning; gene expression profiling of every single cell in the mouse pre-implantation embryo; and computer simulation of the blastocyst morphogenesis (see figure 1). Mouse pre-implantation embryo is suitable for a systems-level analysis, because a) it is composed of a limited number of cells (up to 64) and of cell populations (four); b) it is an isolated system, with no influence from the oviduct or uterus; c) the development can be recapitulated *in vitro* and be visualised by live-imaging; d) the hypothesis can be tested by experimental manipulation. We aim at understanding principles underlying early mammalian development.

Future projects and goals

We adopt a wide variety of experimental strategies, including micromanipulation of the embryo, molecular biology, genetics, live-imaging, and modelling, in order to address fundamental questions in developmental and cell biology at a molecular, cellular and systems level. Our goals are:

- Understanding the mechanism of the first lineage establishment, ICM and TE;
- understanding the formation of epithelial cell polarity and its potential impact on ICM vs TE differentiation;
- development of a new 4D-imaging microscopy to digitally reconstruct blastocyst morphogenesis and patterning, and establish a null hypothesis;
- evaluating the significance of dynamic cellular heterogeneity in embryonic patterning.

Figure 1 (left): Computer simulation of blastocyst morphogenesis.

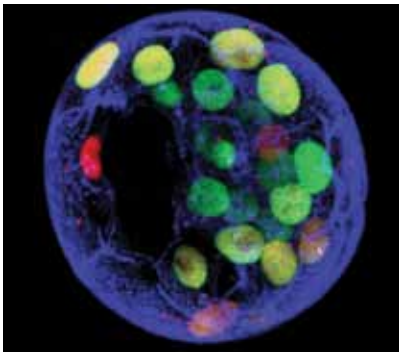


Figure 2 (above): Unprecedented molecular heterogeneity during mouse blastocyst patterning. Cells expressing Nanog (green), Cdx2 (red) or both (yellow).

Selected references

Honda, H., Motosugi, N., Nagai, T., Tanemura, M. and Hiiragi, T. (2008) Computer simulation of emerging asymmetry in the mammalian blastocyst. *Development*, 135(8), 1407-1414.

Dietrich, J. E. and Hiiragi, T. (2007) Stochastic patterning in the mouse pre-implantation embryo. *Development*, 134(23), 4219-4231.

Motosugi, N., Bauer, T., Polanski, Z., Solter, D. and Hiiragi, T. (2005) Polarity of the mouse embryo is established at blastocyst and is not prepatterned. *Genes Dev.*, 19(9),1081-1092.

Hiiragi, T. and Solter, D. (2004) First cleavage plane of the mouse egg is not predetermined but defined by the topology of the two apposing pronuclei. *Nature*, 430(6997), 360-364.

PhD 2002, University of Cologne.

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

Group leader at EMBL since 2008.

Francesca Peri

Microglia: the guardians of the developing brain

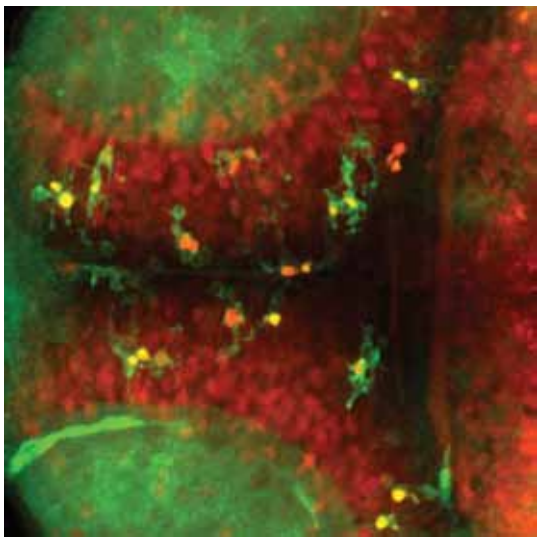
Previous and current research

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

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

Future projects and goals

Despite the importance of microglia in several neuronal pathologies, many fundamental questions concerning microglial neuronal 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), 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

Schlegelmilch, T., Henke, K. & Peri, F. (2010). Microglia in the developing brain: from immunity to behaviour. *Curr. Opin. Neurobiol.* (epub ahead of print)

Peri, F. (2010). Breaking ranks: how leukocytes react to developmental cues and tissue injury. *Curr. Opin. Genet. Dev.*, 20, 416-419

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

While the major part of vertebrate genomes is not coding for proteins, it nevertheless contains functionally important information. In particular, the tightly regulated expression of many developmental genes is achieved through the coordinated action of multiple cis-regulatory elements, which could be located hundreds of kilobases away from the gene they control. Notably, many recent genome-wide association studies have identified disease susceptibility intervals that do not overlap with any protein-coding gene, suggesting that they instead contain elements influencing the expression of distant genes. These findings emphasise the critical role of non-genic sequences in controlling gene activities and the importance of defining precisely the location of such elements and their function. Importantly, the specific distribution of these elements along the chromosomes seems to play a major role in directing their activities, as shown by the dramatic consequences of chromosomal rearrangements and the frequent preservation of chromosomal syntenies during evolution.

Our lab aims to understand the mechanisms that define the regulatory structure of the mammalian genomes and control gene expression. Specifically, we aim to understand how the complex intermingled arrays of genes and cis-regulatory elements found in several loci are translated into gene-specific expression programs. To address this, our lab has developed several experimental approaches in the mouse that allow us to explore and assess the regulatory organisation of the mammalian genome both at a large scale and in great detail. In addition to mouse genetic and transgenic approaches, we have established a simple and efficient *in vivo* transposition system that enables us to characterise systematically the different kinds of regulatory activities present along a chromosome. In particular, we focus on intervals associated with developmental abnormalities, where we use chromosomal engineering techniques to generate series of chromosomal rearrangements that reproduce those found in human patients. Besides supplying mouse models to study the molecular etiology of these genomic disorders, these approaches provide insights into the regulatory architecture of these loci.

Future projects and goals

Chromatin, chromosomal conformation and gene expression: In the nucleus, chromosomes adopt different spatial organisations depending on their transcriptional activities. Distant genomic regions could be brought in close physical proximity by the formation of large chromosomal loops, to favour functional interactions between remote enhancers and their target genes. However, the genomic elements and protein complexes that determine the formation of such specific long-range chromatin interactions are still poorly understood. By combining our advanced genomic engineering approaches with chromatin profiling and conformation analyses (using next generation sequencing and imaging), we aim to understand how specific chromatin structures and conformations are established at defined loci and determine their functional significance in the context of a developing embryo.

Genome regulatory architecture, structural variations and evolution: The tools we have established greatly facilitate the functional exploration of the non-coding part of the genome. We are interested in further expanding these approaches, notably to understand the phenotypic consequences of structural variations or chromosomal aneuploidies found in humans. We are also interested in comparing the regulatory architecture of developmental gene loci between different species, to trace back its emergence and establish how it could have contributed to evolution of body forms.

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

Tschopp, P., Tarchini, B., Spitz, F., Zakany, J. & Duboule, D. (2009). Uncoupling time and space in the collinear regulation of Hox genes. *PLoS Genet.*, 5 (epub)

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

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

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. The development and integration of microfludic devices and the recent advances in next generating sequencing will facilitate major advances in these areas in the coming years. 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.

Regulatory networks driving cell fate decisions: dissecting the logic

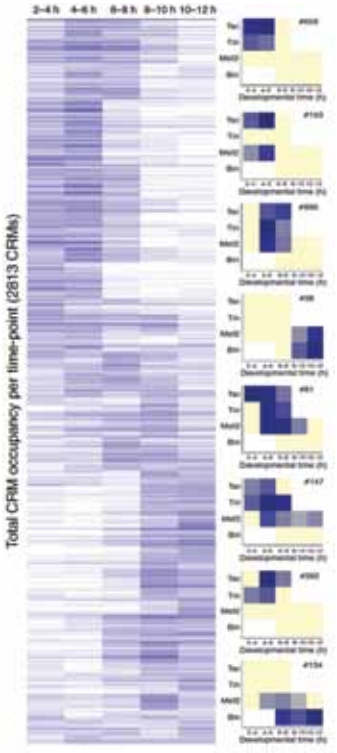
Previous and current research

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, the *Drosophila* fruit fly, there are no predictive models for a transcriptional network leading to cell fate specification.

The main aim of our research is to understand how gene regulatory networks control development and how network perturbations lead to specific phenotypes. To address this we integrate functional genomic, genetic and computational approaches to make predictive models of developmental progression. We 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.

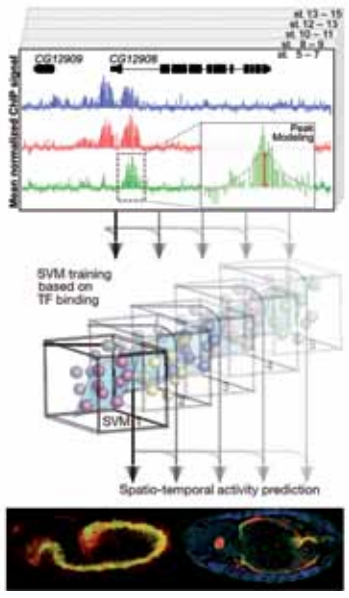


Evolution of cis-regulatory networks: To gain a better understanding of the core functional features of the transcriptional network driving mesoderm specification we will take advantage of the fact that all of the key transcription factors involved are highly conserved at both a sequence and phenotypic 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. 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 2 (left): Dynamic enhancer occupancy reflects a temporal map of developmental progression (Wilczynski & Furlong, Mol Systems Biol*, 2010).*

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



Selected references

Wilczynski, B. & Furlong, E.E. (2010). Dynamic CRM occupancy reflects a temporal map of developmental progression. *Mol. Syst. Biol.*, 6, 383

Cunha, P.M., Sandmann, T., Gustafson, E.H., Ciglar, L., Eichenlaub, M.P. & Furlong, E.E. (2010). Combinatorial binding leads to diverse regulatory responses: Lmd is a tissue-specific modulator of Mef2 activity. *PLoS Genet.*, 6, e1001014

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

Lars Steinmetz

PhD 2001, Stanford University.

Postdoctoral research at the Stanford Genome Technology Center.

Group leader at EMBL since 2003.
Joint Head of Unit since 2009.

Selected references

Xu, Z., Wei, W., Gagneur, J., Clauder-Muenster, S., Smolik, M., Huber, W. & Steinmetz, L.M. (2011). Antisense expression increases gene expression variability and locus interdependency. *Mol. Systems Biology* (In press)

Zheng, W., Zhao, H., Mancera, E., Steinmetz, L.M. & Snyder, M. (2010). Genetic analysis of variation in transcription factor binding in yeast. *Nature*, 464, 1187-91

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., *et al* (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature*, 457, 1033-1037

Systems genetics

Previous and current research

Individuals differ at thousands of positions in the genome. These differences interact with each other and 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 establishing the necessity of 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 at the level of the genome, transcriptome and proteome.

Current projects include elucidating the genetic basis of resistance to malaria parasites in mosquitoes at the level of single alleles; studying the function and mechanism of pervasive transcription of non-coding RNAs; 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 of all sequence variants between two genomes in a single step. Ultimately, by integrating genetics, genomics, systems biology and computational modelling with high-throughput sequencing and microarrays, we aim to develop approaches that will enable personalised and preventative medicine across the world.

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

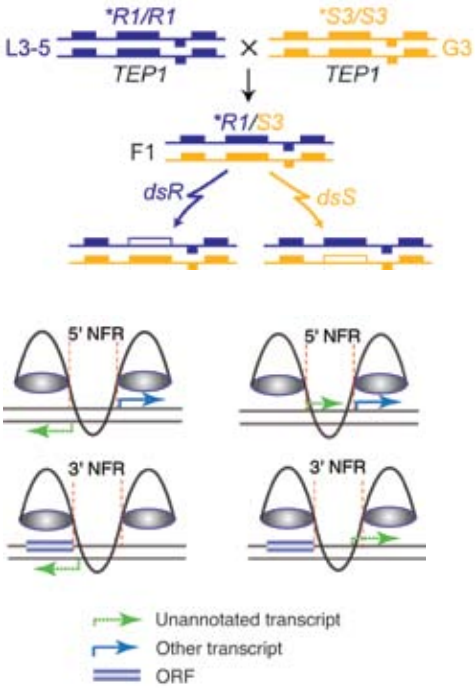
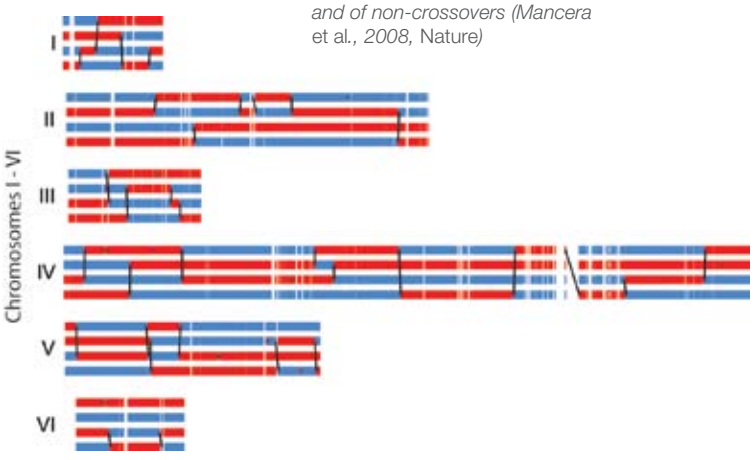


Figure 1 (left): Reciprocal allele-specific RNAi identifies the mosquito allele TEP1 as a major contributor to resistance to malaria parasites (Blandin et al.*, 2009,* Science*).*

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*)*



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.

Computational biology and genomics

Previous and current research

The group develops mathematical and statistical methods for the understanding of genomic data. We apply these methods to the design and analysis of novel, cutting-edge experimental approaches in genetics and cell biology. Our aim is to understand genetic and phenotypic variation on a genome-wide scale: how is the information that is encoded in the genome expressed? How is it processed in networks of interacting molecules in space and time? How does it differ between individuals and how does that affect their condition? How can we predict and tweak a biological system's behaviour from such information?

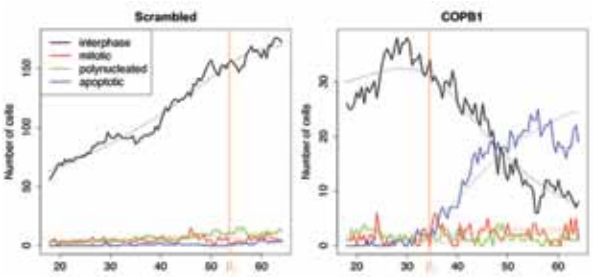
The group brings together expertise from quantitative disciplines – mathematics, statistics, physics, and computer science – with the design and analysis of genomic experiments and their biological interpretation. Computational and statistical methods are often at the heart of systematic, large-scale experimental approaches. Our aim is to develop high-quality methods of general applicability that can be widely used in genomic research. We regard the publication of scientific software as an integral part of the publication of new methodical approaches and contribute to the Bioconductor project (www.bioconductor.org).

Future projects and goals

Progress in biology will continue to be driven by advances in technology. Sequencing now allows us to know the genomes of individual people and is providing transcription and DNA-protein interaction data for many different cellular systems at unprecedented precision. Light microscopy of single, live cells is providing data on molecular interactions and life cycles within the cell and is becoming increasingly automated. We can observe the dynamics of signalling, the cell cycle, and of cell migration both under normal conditions and many different perturbations (RNAi, drugs), and we start to understand the basis of biological variation between cells.

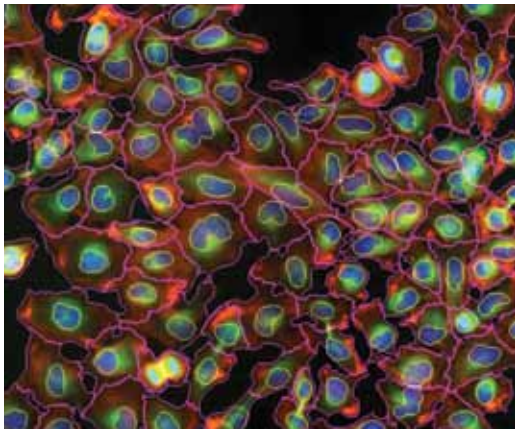
An emphasis of our work is on project-oriented collaborations with experimenters. We aim to develop the computational methods in statistics, probabilistic modelling, image analysis and bioinformatics that are needed to make new types of experiments feasible, and to turn the data into biology. Current projects include:

- Automated phenotyping from high-throughput microscopy: pattern recognition, machine learning, inference of dynamical models;
- DNA-, RNA- and ChIP-Seq and their applications to genetic variation and transcription regulation: data analysis, statistical modelling, regression and testing;
- genetic interactions by large-scale RNAi: design and analysis of combinatorial experimental approaches, sparse model building;
- protein expression and turnover from imaging: parameter estimation, model selection;
- open source software for the analysis of sequencing data, for high-throughput phenotyping from automated microscopy, and for integrative bioinformatics.



Left: Time courses of cellular states observed by live cell imaging are mathematically modelled by a dynamical system with differential equations of motion

Below: Example image from a large-scale RNAi screen on populations of human cells stained for DNA (blue), tubulin (green) and actin (red). Images are automatically segmented, quantitative cell descriptors are computed and analysed for biological phenotypes by machine learning methods.



Selected references

Anders, S. & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.*, 11, R106

Bourgon, R., Gentleman, R. & Huber, W. (2010). Independent filtering increases detection power for high-throughput experiments. *Proc. Natl. Acad. Sci. USA*, 107, 9546

Fuchs, F., Pau, G., Kranz, D., Sklyar, O., Budjan, C., Steinbrink, S., Horn, T., Pedal, A., Huber, W. & Boutros, M. (2010). Clustering phenotype populations by genome-wide RNAi and multiparametric imaging. *Mol. Syst. Biol.*, 6, 370

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

Maja Köhn

PhD 2005 MPI for Molecular Physiology, Dortmund.
Postdoctoral research at Harvard University, Cambridge, Massachusetts.
Group leader at EMBL since 2007.

Selected references

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

Investigation of phosphatases using chemical biology tools

Previous and current research

Protein and second messenger dephosphorylation by phosphatases is fundamental to a vast number of cellular signalling processes and thus to physiological functions. Impairment of these processes contributes to the development of human diseases such as cancer and diabetes. The investigation of phosphatases is challenging, mainly due to their broad substrate specificity and the lack of tools to selectively study particular phosphatases. Despite major accomplishments in the field, understanding of phosphatase function, regulation and substrate interaction is in general still limited. Our main interest is thus to control and investigate phosphatases with the help of chemical tools, based on phosphoinositide (PIP) and peptide synthetic organic chemistry as well as protein semisynthesis, and also with molecular biology approaches. Thereby, we are focusing on phosphatases that promote diseases.

We are working on the design of inhibitors/modulators for phosphatases based on chemical modification of substrates and interacting proteins. PIPs are a major substrate class of interest in our research. We have developed a novel solid phase synthesis strategy that accelerates access to these compounds and their analogues. One goal is to achieve a detailed picture of substrate specificities of lipid phosphatases in biochemical structure-activity relationship (SAR) studies using a library of PIP analogues. Information about specific substrate preferences will help to design specific inhibitors of lipid phosphatases.

The PRL family of phosphatases is of particular interest to us because it is involved in several types of cancer. We apply protein semisynthesis, imaging, and molecular and genome biology approaches to obtain information about natural substrates, regulation and networks of these oncogenic phosphatases. Furthermore, in collaboration with Mathieu Bollen (KU Leuven, Belgium), we are employing new ways to modulate the ubiquitous phosphatase PP1.

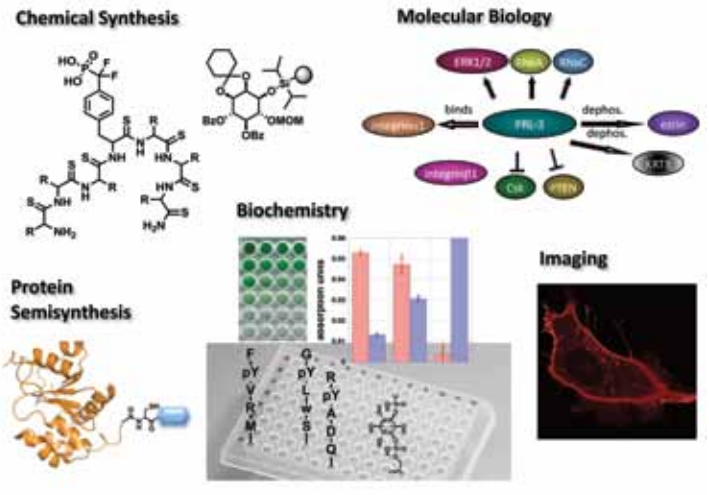
A general understanding of phosphatase and kinase networks is still very incomplete. In collaboration with Matthias Wilmanns (page 101), and Janet Thornton (page 62), we use computational, biochemical and structural approaches to view, predict and validate these networks.

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 lipid phosphatases in cells by applying the modulators resulting from our SAR studies. Developing modulators for highly non-specific serine/threonine phosphatases is a long-term goal and we already successfully developed modulators for PP1 and have begun to pursue this for PP2C in collaboration with José Márquez (page 95).

The lab consists of an equal number of molecular biologists and organic chemists at both 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.



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.

Genome dynamics, evolution, and structural variation

Previous and current research

Members of our group apply experimental and/or computational approaches to study the extent, functional impact and mutational origins of genetic variants, particularly genomic structural variants (SVs). SVs, also known as copy-number variants (CNVs), are among the least well studied classes of genetic variation, despite the fact that their net effect on the human genome (in terms of affected base pairs) is higher than the effect of single nucleotide polymorphisms (SNPs). Recent technological advances (e.g. tiling arrays and next-generation DNA sequencing) are enabling us to decipher the impact of SVs through the analysis of entire genomes.

A recent focus of our group has been the development of next-generation sequencing based approaches for SV discovery and genotyping, including paired-end mapping based SV discovery (PEM), breakpoint junction based SV genotyping (BreakSeq), and read-depth based SV genotyping (CopySeq). As a member of the structural variation analysis group of the 1000 Genomes Project, our group recently integrated computational and experimental methods to construct the largest and highest resolution map of SVs developed to date. The study involved the sequencing and analysis of 185 human genomes (figure 1). Based on a new SV clustering approach, we identified over 50 SV formation hotspots - genomic regions in which distinct DNA rearrangement processes (e.g. recombination) appear to operate at an increased rate. Furthermore, we recently assessed the effect of genetic variants (SVs and SNPs) on gene expression regulation, based on coupling chromatin immunoprecipitation sequencing (ChIP-Seq) in several human cell-lines with a computational analysis framework (figure 2). Additional research projects in our lab focus on the *de novo* formation mechanisms and phenotypic impact of somatic genomic rearrangements occurring in cancer, specifically prostate cancer and pediatric brain tumours.

Future projects and goals

The extent to which genomes differ due to SVs, the impact SVs have on the phenotype, and the mutational processes underlying *de novo* SV-formation are poorly understood compared to other variation classes (such as SNPs). We believe that SVs commonly cause phenotypic variation by perturbing tightly regulated biological systems. The long-term goals of our group include the study of SV *de novo* formation mechanisms and the assessment of the impact of SVs on biological systems.

Furthermore, together with collaborators involved in the 1000 Genomes Project and the International Cancer Genome Consortium, we are continuing the development of approaches for SV detection. With human genome sequencing becoming a routinely applied research tool, we foresee that in the not-so-distant future personal genome sequencing will be widely applied to predict specific medical outcomes in patients diagnosed with diseases. Thus, our approaches may in the future have an impact on personalised medicine by enabling predictions relating to the effectiveness of personalised drug treatments.

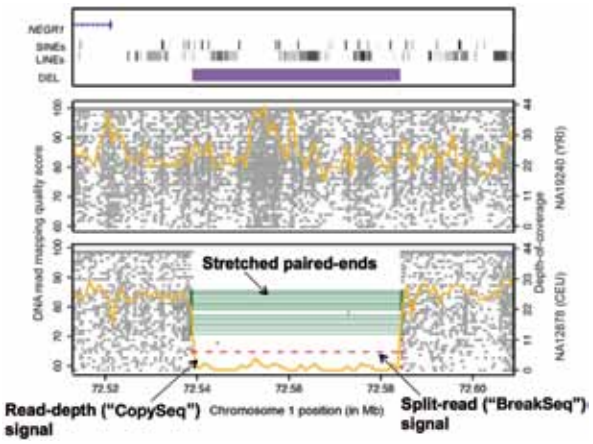
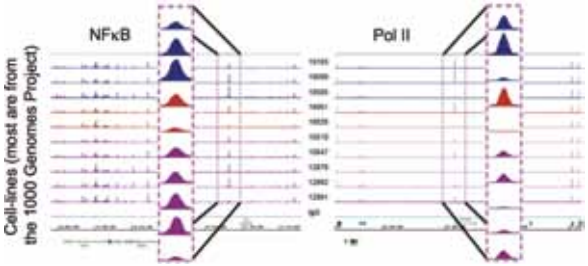


Figure 1 (left): Integrating computational methods with population-scale sequencing enables global SV analysis (adapted from Mills et al.*,* Nature *2011).*

Figure 2 (below): Protein-DNA binding associated gene regulation displays marked variation among humans, frequently involving SNPs or SVs (Kasowski et al.*,* Science *2010).*



Selected references

The 1000 Genomes Project Consortium (2010). A map of human genome variation from population-scale sequencing *Nature*, 467, 1061-1073

Kasowski, M. *et al*., (2010). Variation in transcription factor binding among humans. *Science*, 328, 232-235

Waszak, S.M., Hasin, Y., Zichner, T., Olender, T., Keydar, I., Khen, M., Stütz, A.M., Schlattl, A., Lancet, D. & Korbel, J.O. (2010). Systematic inference of copy-number genotypes from personal genome sequencing data reveals extensive olfactory receptor gene content diversity. *PLoS Comput. Biol.*, 6, e1000988

Mills, R. *et al*. (2011). Mapping copy number variation by population-scale sequencing. *Nature*, 470, 59-65

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.

Selected references

Van Hoof, D. *et al* (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*, 8(7), 1566-78

Dormeyer, W., van Hoof, D., Braam, S.R., Heck, A.J., Mummery, C.L. & Krijgsveld, J. (2008). Plasma membrane proteomics of human embryonic stem cells and human embryonal carcinoma cells. J. *Proteome Res.*, 7, 2936-2951

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

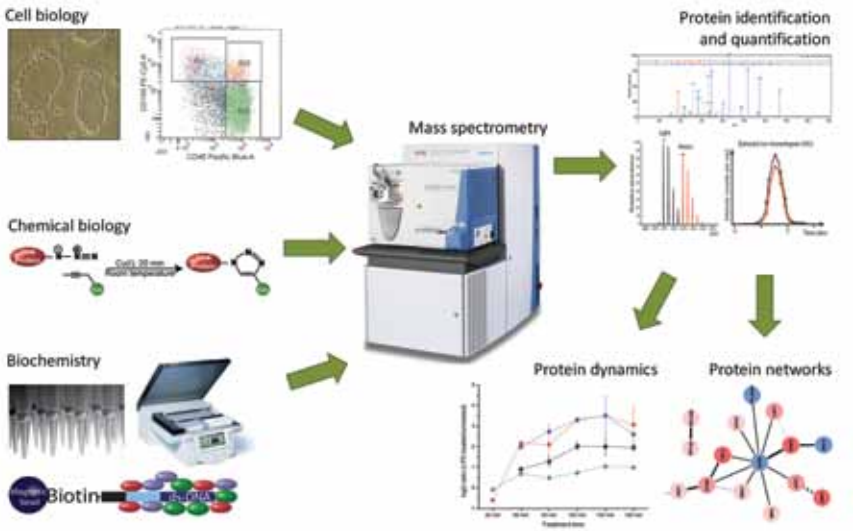
Quantitative proteomics

Previous and current research

Proteins fulfil most of the functions that are crucial for cellular survival. In addition, it is becoming increasingly clear that proteins rarely act alone, but that they constitute intricate networks, both among themselves and with other biomolecules. This system is both robust and dynamic, allowing a cell to respond to external cues and an organism to develop from an embryonic to a mature state. Our interest is to understand cellular behaviour from this perspective, realising that one needs to study proteins collectively rather than in isolation, and dynamically rather than under a static condition.

Our research is centred on quantitative proteomics, combining biochemistry, analytical chemistry, mass spectrometry and bioinformatics. Mass spectrometry coupled to liquid chromatography has matured to the stage that thousands of proteins can be analysed simultaneously, encompassing entire proteomes for relatively simple organisms (e.g. yeast) while covering decent portions of more complex proteomes (e.g. mammals). Our lab is equipped with state-of-the-art mass spectrometric technology (Thermo Orbitrap Velos, Bruker Maxis Qq-Tof) that we use for the development of quantitative proteomic techniques using stable-isotope labelling (e.g. SILAC and chemical approaches), and their application to study protein dynamics and interactions.

Our biological interest focuses on three main topics. The first is in developmental biology, with an emphasis on stem cell biology. We apply quantitative proteomics to the initial steps in haematopoiesis, studying FACS-sorted haematopoietic stem cells and multipotent progenitors isolated from mice. This should provide molecular and mechanistic clues as to how stem cells progress from a quiescent to an activated state. Knowing that proteins change in expression does not necessarily explain how this change is regulated, and therefore, secondly, we are interested in the underlying phenomenon of protein turnover, defined by protein synthesis and degradation. Using chemical biological tools, we can selectively capture proteins that are newly synthesised upon cellular stimulation, isolating them from the background of 'old' (pre-existing) proteins. Profiling these newly synthesised proteins quantitatively over time provides a valuable link between genome regulation and protein output. A third research topic is in the area of transcriptional regulation, where we are interested in identifying proteins that interact with DNA in a sequence-specific manner. This is complementary to the concept of chromatin IP, where we don't ask the question where a particular protein binds to the genome, but rather what proteins bind to a defined genomic region. We have implemented the tools to identify proteins interacting with enhancer elements in fruit flies, identifying candidate regulatory proteins that are now being tested for their functionality in embryonic development.



Future projects and goals

Our future work can be divided into three major areas:

- Studying the changing proteome during gain and loss of pluripotency. We will focus on differentiation of haematopoietic stem cells and on reprogramming of fibroblasts to iPS cells.
- investigating protein turnover in mammalian cells and yeast, using various perturbations and growth conditions;
- studying protein-DNA and protein-RNA interactions to identify proteins modulating transcription and translation.

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.

Professor of physiological chemistry in the Medical Faculty of the University of Munich since October 2010

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 group 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 instance, 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.

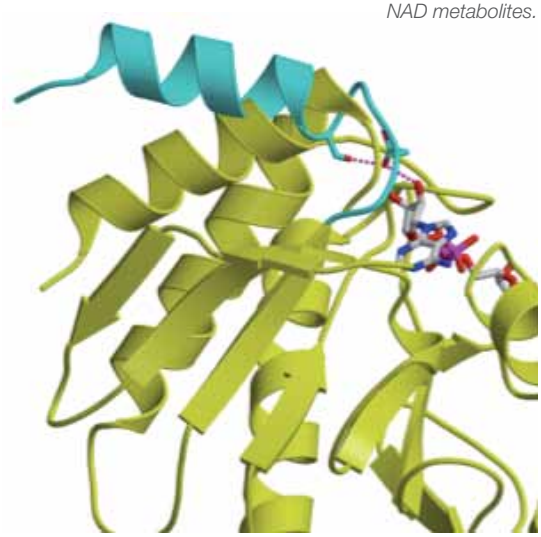
Our current research is focused on three complementary areas of chromatin plasticity. We are dissecting the structure and function of important chromatin remodelling enzymes using a combination of high-resolution X-ray crystallography, biochemistry, protein engineering and *in vivo* approaches. Secondly, 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.

Macrodomains rapidly respond upon DNA damage activation by relocalising to nuclear sites of PARP1 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.



Selected references

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. B., *et al.* (2009). A macrodomain-containing histone rearranges chromatin upon sensing PARP1 activation. *Nat. Struct. Mol. Biol.*, 16(9), 904-7

Stuwe, T., Hothorn, M., Lejeune, E., Rybin, V., *et al*. (2008). The FACT Spt16 "peptidase" domain is a histone H3-H4 binding module. *Proc. Natl. Acad. Sci. USA*, 105, 8884-8889

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

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.

Christoph A. Merten

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.

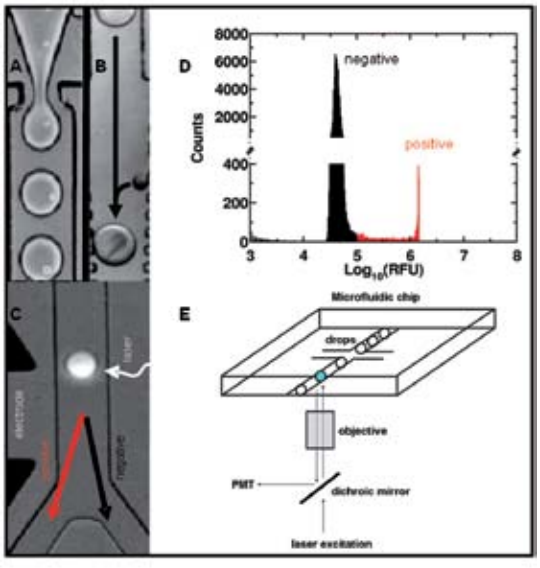
The research in our group focuses on novel microfluidic approaches for biological and chemical applications. Recently, we have established droplet-based microfluidic platforms for the incubation and screening of human cells and multicellular organisms. In these systems, aqueous droplets within an immiscible 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

Vyawahare, S., Griffiths, A.D. & Merten, C.A. (2010) Miniaturization and Parallelization of Biological and Chemical Assays in Microfluidic Devices. *Chem. Biol.*, 17(10), 1052-65.

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,* 10(10), 1302-7.

Granieri, L., Baret, J.C., Griffiths, A.D. & Merten, C.A. (2010) High-throughput screening of enzymes by retroviral display using droplet-based microfluidics. *Chem. Biol.,* 17(3), 229-35.

Koster, S., Angile, F.E., Duan, H., Agresti, J.J., Wintner, A., Schmitz, C., Rowat, A.C., Merten, C.A., Pisignano, D., Griffiths, A.D. & Weitz, D.A. (2008) Droplet-based microfluidic devices for encapsulation of single cells. *Lab Chip,* 8(7), 1110-5.

Structural and Computational Biology Unit

Our 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 and NMR with medium-range resolution from single particle electron microscopy, and cellular imaging obtained by electron tomography and light microscopy. Biochemistry, chemical biology and single molecule fluorescence spectroscopy complement the structural biology activities and, together with a wide range of innovative computational biology activities, integrate them into a comprehensive description of biological function.

Within the unit, there is a continuing interplay between groups with expertise in different methodologies, reflecting the belief that a combination of structural and functional studies is the most rewarding route to an understanding of the molecular basis of biological function, and that computational biology is essential to integrate the variety of tools and heterogeneous data into a comprehensive spatial and temporal description of biological processes. In this way, groups in the Structural and Computational Biology Unit pursue a few common large projects. One example is the 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. An example for structural systems biology is the analysis of a thermophilic fungus where combined experimental and computational approaches enable insight into eukaryotic thermophily at the molecular, cellular and organismal level.

Currently, the unit consists of twelve research groups with broad methodological expertise. It covers electron microscopy (three groups), X-ray crystallography (two groups), NMR (one group), chemical biology (two groups) and computational biology (four groups). In addition, two groups based in different units have shared appointments with the unit (the Ladurner group, Genome Biology (page 33) 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 for single particle cryo-electron microscopy and cryo-electron tomography. The unit also has facilities for single-molecule light microscopy, isothermal calorimetry, circular dichroism, static and dynamic light scattering and analytical ultracentrifugation, as well as for large scale growth of prokaryotic and eukaryotic cells. The computing environment offers access to about 3000 CPU cores whereby large central clusters and separate workstations are 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.

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:

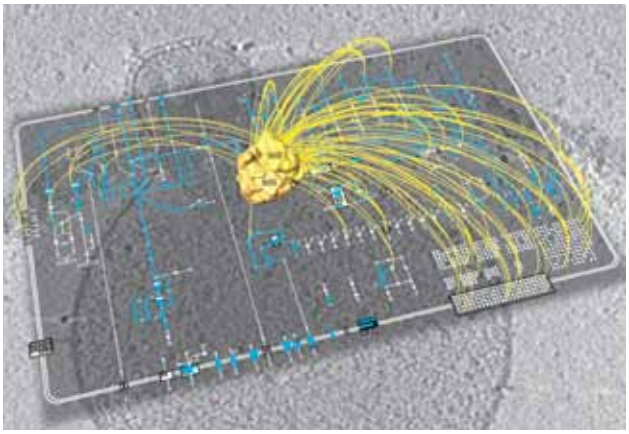
- genes, proteins and small molecules;
- networks and cellular processes;
- phenotypes and environments, often related to diseases.

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 lifestyle 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 and cancer, 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.

Future projects and goals

The main system we will study over the next years is the human gut, but we will also take part in many collaborations studying various other systems, such as the Tara Oceans project to explore the biodiversity on Earth. In the gut, we aim to understand biological processes upon drug treatment, but also using large scale perturbation data and several cellular readouts to identify drug targets. We will explore networks between proteins and chemicals such as lipids or carbohydrates and link them to phenotypic data such as drug side effects. We will also look at our 2kg or so bacterial in our intestinal system, study them as communities and explore their impact on colorectal cancer and various other diseases in the context of lifestyle and other parameters. We also want to understand how these communities evolve in us, how frequently they are transmitted parentally or horizontally, and how they communicate with each other and with our cells.

The group is partially associated with the Max Delbruck Center for Molecular Medicine in Berlin and with the Molecular Medicine Partnership Unit (MMPU) 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* Iconnectors*) and a metabolic reconstruction in which the correspondence to operon organisation is shown (blue).*

Selected references

Qin, J., Li, R., Raes, J., *et al.* (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature,* 464, 59-65

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

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

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 in 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.

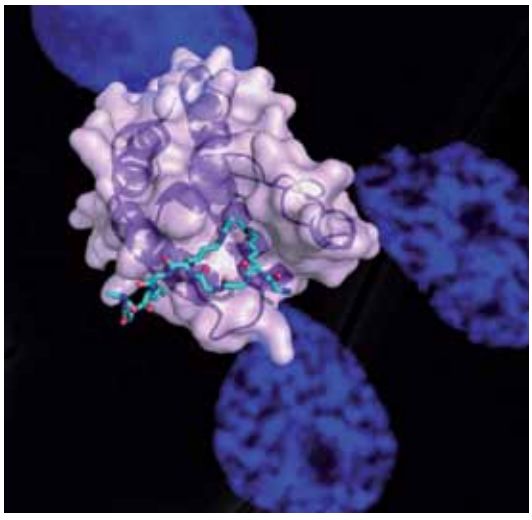
To achieve these goals, we use structural information obtained by X-ray crystallography and electron microscopy combined with other biophysical and biochemical techniques. Systems currently under investigation include transcription factor/DNA complexes, yeast RNA polymerase III and multi-protein complexes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

Fernandez-Tornero, C., Bottcher, B., Rashid, U.J., Steuerwald, U., Florchinger, B., Devos, D.P., Lindner, D. & Muller, C.W. (2010). Conformational flexibility of RNA polymerase III during transcriptional elongation. *EMBO J.*, 29(22) 3762-72

J, Stirnimann, C.U., Petsalaki, E., Russell, R.B. & Muller, C.W. (2010). WD40 proteins propel cellular networks. *Trends Biochem. Sci.*, 35(10), 565-74

Moriniere, J., Rousseaux, S., Steuerwald, U., Soler-Lopez, M., Curtet, S., *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(13), 1965-77

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

Controlled DNA rearrangements are essential for survival on all levels of life from individual cells to populations. Our lab is interested in understanding how DNA recombination is carried out on the molecular and cellular levels. We mainly focus on DNA transposons, a class of mobile genetic elements, which can autonomously 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 be engineered to carry desired genetic information, and offer stable and heritable modifications of a target genome. Consequently, these 'jumping' DNA elements offer attractive tools for genetics and human gene therapy. To support the future development of transposon based genetic tools, we study their mechanism of movement. We strive to understand their molecular mechanism of transfer, target site selection and cellular control. We currently study: i) the movement of various DNA transposons; and ii) RNA-based regulatory pathways that control the efficiency of transposition. Our techniques include structural biology (mainly X-ray crystallography), molecular biology, biochemistry and cell culture assays.

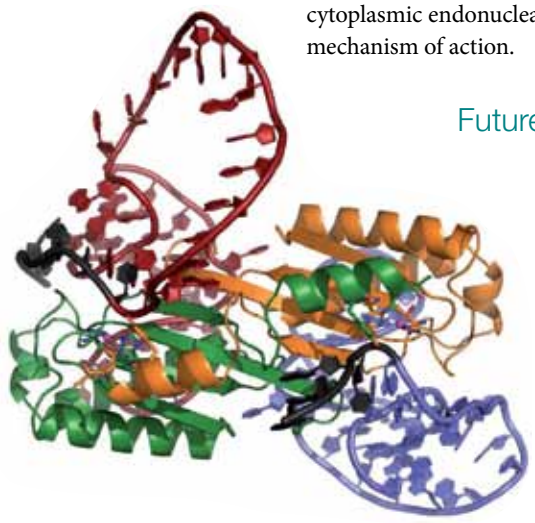
The reactivated transposon *Sleeping Beauty* has recently become a favoured genetic tool used for forward mutagenesis screens, identification of oncogenes, mapping regulatory landscapes, chromosomal engineering (Ivics *et al*, 2009) and even in gene therapy clinical trials (Williams, 2008). However, the mechanism of SB transposition is poorly understood. Our on-going structural and functional studies will offer a mechanistic understanding and invaluable insights for rational design of transposition cassettes.

Site-specific elements: One of the main obstacles of gene therapy is integration of the therapeutic gene at unwanted genomic locations. Site-specific recombinases may offer a solution. Our recent work revealed the mechanism of the bacterial Insertion Sequence, IS608, that integrates site-specifically at short DNA sequences (Barabas *et al*, 2008). We found that target site recognition is achieved by the transposon DNA base pairing with target DNA. The site of insertion can be easily altered by changing a few nucleotides in the transposon (Guynet *et al*, 2009). We are currently investigating if this target recognition strategy can be expanded to target potentially unique genomic sites. We are also studying a newly discovered mobile element, the plasticity zone transposon (TnPZ) from *Helicobacter pylori* (Kersulyte *et al*, 2009). This element shares features with Xer recombinases and may move via bacterial conjugation. We want to learn how this transposon moves and selects its 7nt long specific target site sequence.

In eukaryotes transposon silencing is mainly achieved by regulatory piRNAs (PIWI-interacting RNAs). They facilitate transposon silencing by triggering both transposon DNA methylation and mRNA degradation by cytoplasmic endonucleases. We investigate various proteins involved in this pathway to reveal their role and mechanism of action.

Future projects and goals

- To understand transposition in the cellular context, we will identify host proteins that interact with the transposase protein or transposon DNA. The effect of host factors on transposition efficiency will also be analysed.
- We will select representative targets from several transposon and recombinase families to capture a broader picture of recombination pathways applied by nature.



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

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*, 34, 612-619

Kersulyte, D., Lee, W., Subramaniam, D., Anant, S., Herrera, P., Cabrera, L., Balqui, J., Barabas, O., Kalia, A., Gilman, R.H. & Berg, D.E. (2009). Helicobacter Pylori's plasticity zones are novel transposable elements. *PLoS One*, 4, e6859

Guynet, C., Hickman, A.B., Barabas, O., Dyda, F., Chandler, M. & Ton-Hoang, B. (2008). *In vitro* reconstitution of a single-stranded transposition mechanism of IS608. *Mol. Cell*, 29, 302-312

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*, 132, 208-220

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 step-wise 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.

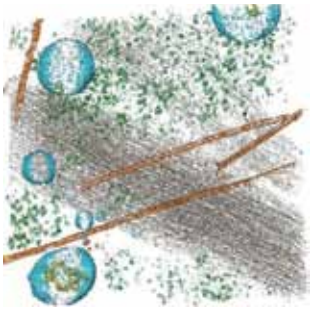


Figure 1. Cryo-electron tomogram of a fraction of the cytoplasm of a human cell. Microtubules are colored in orange, stress fibres in grey, protein complexes in green, membranes in cyan and vesicular contents in yellow.

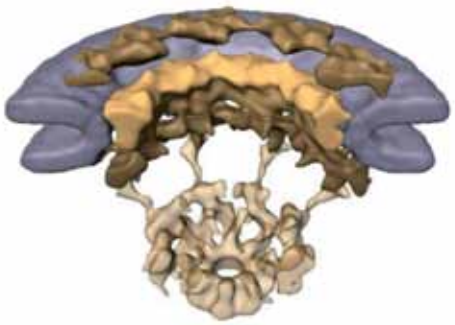


Figure 2. Structure of the nuclear pore complex. Membranes are colored in grey, the scaffold structure in yellow and the nuclear basket in transparent brown.

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 the mechanisms of assembly and budding of enveloped viruses and coated vesicles. We aim to understand how proteins collect together the cargo of the virus or vesicle, and define and manipulate the shape of the membrane to cause budding. To explore these questions we are studying a range of different cellular and viral specimens using cryo-electron microscopy and tomography.

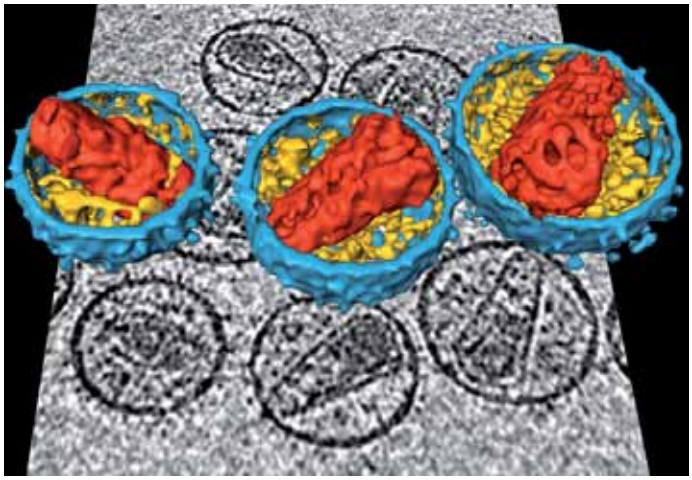
Cryo-electron microscopy techniques are particularly appropriate for studying vesicles and viruses because they allow membrane topology to be observed in the native state, while maintaining information about the structure and arrangement of associated proteins. Computational image processing and three-dimensional reconstructions are used to extract and interpret this information.

A particular emphasis of our research is the structure and life-cycle of asymmetric membrane viruses such as HIV and Marburg virus. The structure and assembly of the virus particles offers insights into general features of membrane budding.

We take a step-by-step approach to understanding the native structure of budding events. Correlative fluorescence and electron microscopy methods can be used to locate and characterise features of interest. Three-dimensional reconstructions of these features can be obtained using 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.

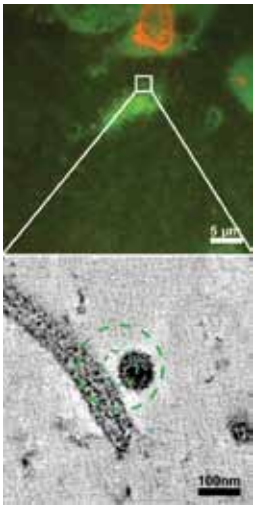
Future projects and goals

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? How does the curvature of a membrane influence its interaction with particular proteins? We are developing and applying novel microscopy and image processing approaches to address these questions.



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

Correlative fluorescence and electron microscopy can be used to locate an individual fluorescent virus particle at the surface of a cell (Kukulski et al. *2011)*



Selected references

Kukulski, W., Schorb, M., Welsch, S., Picco, A., Kaksonen, M. & Briggs, J.A. (2011). Correlated fluorescence and 3D electron microscopy with high sensitivity and spatial precision. *J. Cell Biol.,* 192(1), 111-9

de Marco, A., Muller, B., Glass, B., Riches, J.D., Krausslich, H.G. & Briggs, J.A. (2010). Structural analysis of HIV-1 maturation using cryo-electron tomography. *PLoS Pathog.,* 6, e1001215

Welsch, S., Kolesnikova, L., Krahling, V., Riches, J.D., Becker, S. & Briggs, J.A. (2010). Electron tomography reveals the steps in filovirus budding. *PLoS Pathog.,* 6, e1000875

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(27), 11090-5

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 involved in RNA metabolism and methodology development for drug design

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 overcome challenges relating to 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) structure-activity and dynamics-activity relationships of RNP complexes and catalytic RNAs involved in RNA processing; and 2) the interaction of small drugs with cellular receptors.

Our work aims at describing the features of RNA-protein recognition in RNP complex enzymes and at characterising the structural basis for their function. Currently, we are investigating the nucleolar multimeric box C/D RNP complex responsible for the methylation of the 2'-O-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 Pillai group (page 98) to understand the structure, function and assembly control of RNP complexes involved in the regulation of gene expression through the piRNA pathway (figure 1).

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. In a second area of research, 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 allow a ligand-based reconstruction of the receptor binding pocket (figure 2). The most prominent example of our activities in this field is INPHARMA, a novel approach to structure-based drug design that does not require high-resolution structural data on the receptor-drug complex. We apply our methods to study the functional mechanisms of anti-cancer drug-leads, designed as inhibitors of kinases, proteasome and membrane receptors.

Future projects and goals

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.

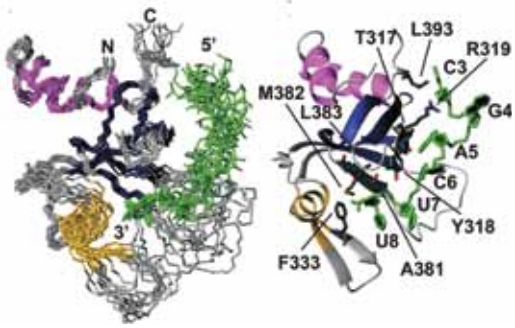


Figure 1: Overview of the structure of the complex between the Miwi-PAZ domain and the 3'-end 2'-O-methylated piRNA.

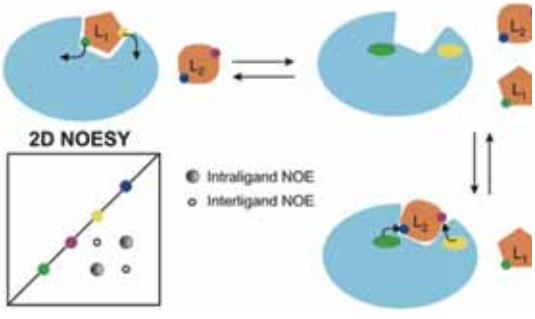


Figure 2: Schematic representation of the principle of the INPHARMA NOEs.

Selected references

Falb, M., Amata, I., Gabel, F., Simon, B. & Carlomagno, T. (2010). Structure of the K-turn U4 RNA: a combined NMR and SANS study. *Nucleic Acids Res.*, 38(18), 6274-85

Simon, B., Kirkpatrick, JP., Eckhardt, S., Reuter, M., Rocha, EA., Andrade-Navarro, MA, Sehr, P., Pillai, R., Carlomagno, T. (2010). Recognition of 2'-O-methylated 3'-end of piRNA by the PAZ domain of a Piwi protein (in press)

Orts, J., Tuma, J., Reese, M., Grimm, S.K., *et al*. (2008). Crystallography-independent determination of ligand binding modes. *Angew Chem. Int. Ed. Engl.,* 47, 7736-7740

Liu, S., Li, P., Dybkov, O., Nottrott, S., Hartmuth, K., Luhrmann, R., Carlomagno, T. & Wahl, M.C. (2007). Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. *Science,* 316, 115-120

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.

Biomolecular networks

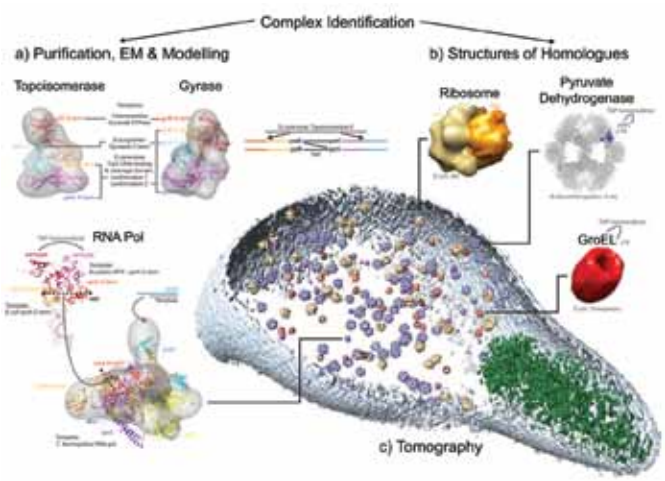
Previous and current research

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.

Selected references

Gallego, O., Betts, M.J., Gvozdenovic-Jeremic, J., Maeda, K., Matetzki, C., Aguilar-Gurrieri, C., *et al.* (2010) A systematic screen for protein-lipid interactions in *Saccharomyces cerevisiae*. *Mol. Syst. Biol.*, 6, 430.

Gehlenborg, N., O'Donoghue, S.I., Baliga, N.S., Goesmann, A., Hibbs, M.A., *et al.* (2010) Visualization of omics data for systems biology. *Nat. Methods.* 7(3 Suppl), S56-68

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

Yus, E., Maier, T., Michalodimitrakis, K., van Noort, V., Yamada, *et al.* (2009) Impact of genome reduction on bacterial metabolism and its regulation. *Science*, 326,1263-8

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

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. Decisions are made by in-complex molecular switching. The deeply misleading 'kinase cascade' metaphor needs to be retired and the sooner, the better. Regulatory proteins make remarkable numbers of interactions, with the corollary that they also have highly modular architectures.

We and collaborators develop and deploy the Eukaryotic Linear Motif resource ELM for investigating 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). Viral proteomes are rich in LMs that are used for hijacking cell systems required for viral production (see figure). The ELM resource data are now being used by many bioinformatics groups to develop and benchmark LM predictors.

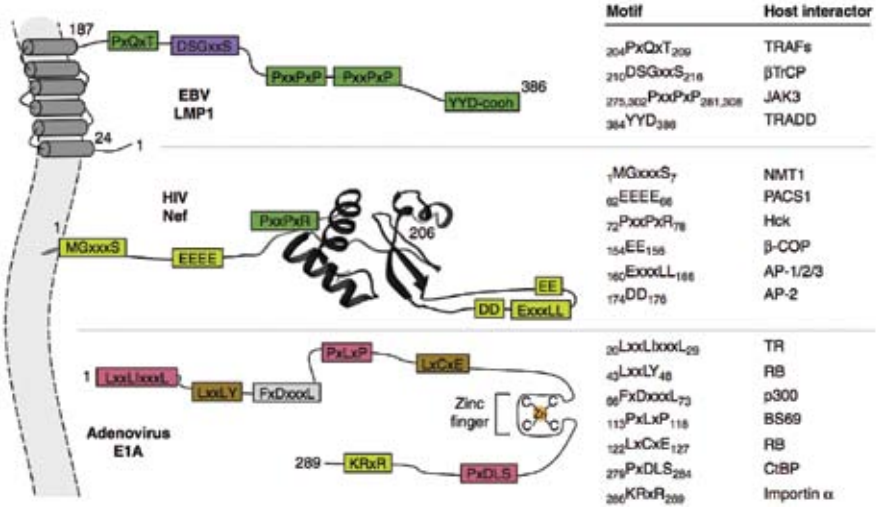
We are now actively hunting for new LM candidates. For example, we recently proposed new candidate KEN boxes, a sequence motif that targets cell cycle proteins for destruction in anaphase, as well as KEPE, a motif of unknown function that is superposed on many sumoylation sites. We look to collaborate with experimental groups undertaking validation experiments.

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 Clustal W and Clustal X programs that are widely used for multiple sequence alignment. We also provide public web servers for Phospho.ELM, a collection of some 42 000 reported phosphorylation sites, and EpiC, a tool to aid in targeting epitopes for antibody selection.

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. We are now working to improve the way that bioinformatics standards represent cooperative molecular interactions. 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.



Selected LM-rich viral proteins and their cellular partners (Davey, 2010).

Selected references

Davey, N.E., Trave, G. & Gibson, T.J. (2010). How viruses hijack cell regulation. *Trends Biochem. Sci.* (in press)

Haslam, N.J. & Gibson, T.J. (2010). EpiC: an open resource for exploring epitopes to aid antibody-based experiments. *J. Proteome Res.*, 9, 3759-3763

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., Weatheritt, R.J., Budd, A., Hughes, T., Pas, J., Rychlewski, L., Trave, G., Aasland, R., Helmer-Citterich, M., Linding, R. & Gibson, T.J. (2009). ELM: the status of the 2010 eukaryotic linear motif resource. *Nucleic Acids Res.*, D:167-80

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

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

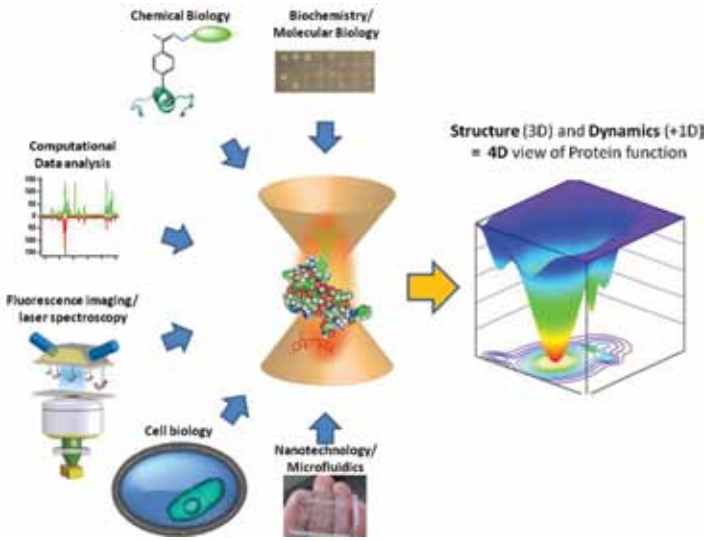
Structural light microscopy – single molecule spectroscopy

Previous and current research

Our research combines 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 signalling hubs central to the interactome (the whole set of molecular interactions in the cell). Their ability to adopt multiple conformations is considered a major driving force behind their evolution and enrichment in eukaryotes.

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

In contrast, single molecule techniques, which directly probe the distribution of molecular events, can reveal important mechanisms that otherwise remain obscured. 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 special fluorescent dyes, which still hampers the broad application of this technique. In our laboratory we are utilising a large spectrum of chemical biology and state-of-the-art protein engineering tools to overcome this limitation, with genetically encoding unnatural amino acids one of our primary strategies. With a focus on studying biological questions, we also continue to develop new methods and recruit techniques from other disciplines (such as super-resolution microscopy and microfluidics) whenever they promise to assist our overall goal of improving our biological understanding.



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 an absolute critical role in the 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 natural complex environment.

Selected references

Gambin Y, Vandelinder V, Ferreon ACM, Lemke EA, Groisman A, Deniz AA, (2011). Visualizing a one-way protein encounter complex by ultrafast single-molecule mixing. *Nature Methods,* epub Feb 2011

Lemke, E.A., Gambin, Y., Vandelinder, V., Brustad, E.M., *et al.* (2009). Microfluidic device for single-molecule experiments with enhanced photostability. *J. Am. Chem. Soc.*, 131, 13610-13612

Brustad, E.M., Lemke, E.A., Schultz, P.G. & Deniz, A.A. (2008). A general and efficient method for the site-specific dual-labeling of proteins for single molecule fluorescence resonance energy transfer. *J. Am. Chem. Soc.*, 130, 17664-17665

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

Kiran Patil

M. tech. (Chemical engineering) 2002, Indian Institute of Technology, Bombay.

PhD 2006, Technical University of Denmark.

Assistant Professor, 2006–2010, Technical University of Denmark.

Group leader at EMBL since 2010.

Architecture and regulation of metabolic networks

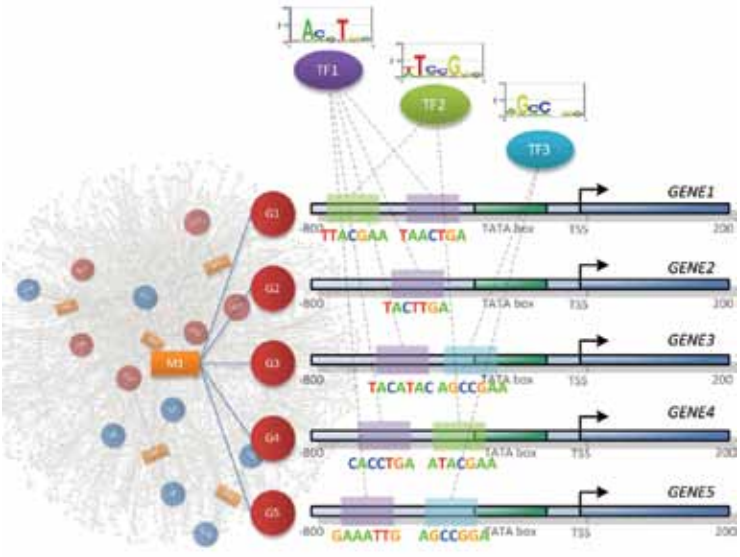
Previous and current research

Regulation of metabolic network activity in response to environmental and genetic changes is fundamental to the survival and evolution of organisms. Disorders and malfunctions of metabolic networks are at the root of complex, systemic diseases such as diabetes and obesity. On the other hand, microbial metabolic capabilities are crucial for sustainable production of chemicals and pharmaceutical compounds of socio-economic importance. What are the thermodynamic and regulatory principles underlying the architecture and operation of metabolic networks? What are the mechanisms by which metabolic responses are linked to sensing and signalling networks? Biochemical principles dictating the metabolic phenotype are emerging through various genome-wide molecular abundance and interaction studies. At the scale of genome and evolutionary time-span, mechanistic answers to these questions, however, have still remained largely elusive. A main goal of our group is to tackle these questions through a combination of modelling, bioinformatics and experimental approaches.

We develop *in silico* models and design algorithms for quantitatively predicting metabolic phenotypes given a certain genotype. These models exploit the principle of conservation of mass as well as our understanding of the biological objective functions underlying the network functionality. Several microbial metabolic engineering problems have been used by our group for successful *in vivo* testing of the *in silico* model-guided predictions. To further the predictive power of metabolic models, we are actively researching the integration of genomic, transcriptomic, proteomic and metabolomic information. This has led to the discovery of new regulatory principles and, in some cases, underlying mechanisms. For example, we have previously shown that the transcriptional regulation within a metabolic network is organised around perturbation-specific key metabolites crucial for adjusting the network state (see figure). Using such integrative data analysis approaches, we are also studying the human metabolic network, working towards the development of a framework for rationally designing clinical intervention strategies and diagnostics for type-2 diabetes.

Future projects and goals

Designing novel modelling strategies for incorporating non-linear regulatory constraints into genome-scale metabolic models will be a major goal of our future projects. Understanding of metabolic changes during development and adaptive evolution is another aspect that we wish to investigate in order to gain insight into the dynamic nature of metabolic network operation in these fundamental biological processes. To this end, we are actively seeking collaborative projects within EMBL and elsewhere.



Reporter algorithm integrates omics data with metabolic network and thereby identifies metabolic regulatory hot-spots. M1 - metabolite; G1-5 - upregulated genes; purple/green/blue circles & squares - transcription factors and corresponding binding motifs.

Selected references

Brochado, A.R., Matos, C., Moller, B.L., Hansen, J., Mortensen, U.H. & Patil, K.R. (2010). Improved vanillin production in baker's yeast through *in silico* design. *Microb. Cell Fact*, 9, 84

Zelezniak, A., Pers, T.H., Soares, S., Patti, M.E. & Patil, K.R. (2010). Metabolic network topology reveals transcriptional regulatory signatures of type-2 diabetes. *PLoS Comput. Biol.*, 6, e1000729

Cakir, T., Patil, K.R., Onsan, Z., Ulgen, K.O., Kirdar, B. & Nielsen, J. (2006). Integration of metabolome data with metabolic networks reveals reporter reactions. *Mol. Syst. Biol.*, 2, 50

Patil, K.R. & Nielsen, J. (2005). Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc. Natl. Acad. Sci. USA*, 102, 2685-2689

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.

Carsten Sachse

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

Previous and current research goals

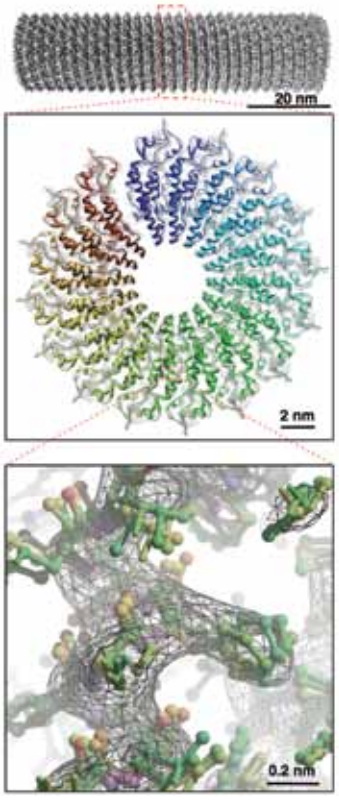
The molecular hallmark of neurodegenerative diseases, such as Alzheimer's and Parkinson's, is the formation of large protein aggregates called amyloid fibrils. The net build-up of these fibrillar aggregates is a result of an imbalance in the cellular production and clearance of misfolded polypeptides. In our group, we are investigating the molecular structures involved in these aberrant processes as they provide fundamental insights to our understanding of ageing and neuronal dysfunction.

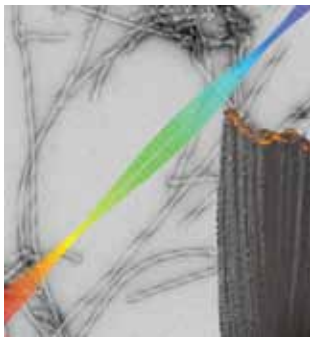
We are visualising the molecules by electron cryo-microscopy (cryo-EM), because large macromolecular structures and multi-protein complexes can be studied in their near-native environment without the need for crystals. Small amounts of material are sufficient to obtain snapshots of "single particles" in the electron cryo-microscope and subsequent computer-aided image processing enables 3D image reconstruction. To realise the promising potential of the technique, the scientific community is still in great need of hardware-based improvements and software enhancements. Therefore, we are also interested in developing techniques including sample preparation and data processing to ultimately increase the resolution of single-particle cryo-EM. We would like to make it a routine tool for structural biology research of large macromolecules.

Future projects and goals

Autophagy (from 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:

- How are protein deposits structurally linked to autophagy?
- What are the shapes of these multiprotein assemblies at the membrane?
- How do they give rise to the cellular structure of the autophagosome?





Far left: High-resolution helical reconstruction of tobacco mosaic virus at near-atomic resolution using single-particle cryo-EM. Top: helical rod. Center: cross section. Bottom: close-up of side-chain density.

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

Selected references

Sachse, C., Grigorieff, N. & Fandrich, M. (2010). Nanoscale flexibility parameters of Alzheimer amyloid fibrils determined by electron cryo-microscopy. *Angew. Chem. Int. Ed. Engl.*, 49, 1321-1323

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

Sachse, C., Fandrich, 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

Sachse, C., Chen, J.Z., Coureux, P.D., Stroupe, M.E., Fandrich, M. & Grigorieff, N. (2007). High-resolution electron microscopy of helical specimens: a fresh look at tobacco mosaic virus. *J. Mol. Biol.*, 371, 812-835

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 are 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.

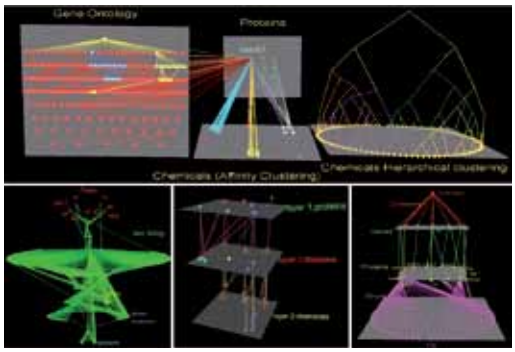
The principal aim of the group is to capture and centralise the knowledge generated by the scientists in the several divisions, and to organise that knowledge such that it can be easily mined, browsed and navigated. By providing access to this resource to all scientists in the organisation, it will foster 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.

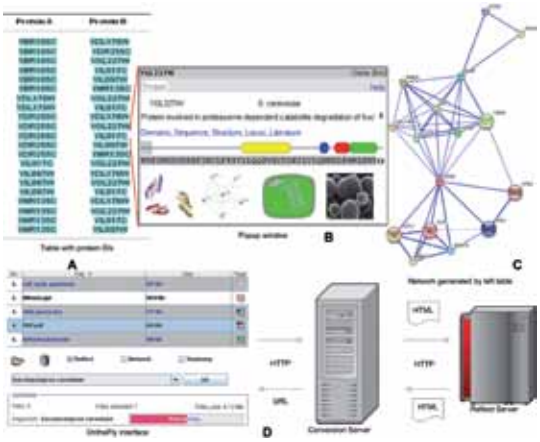
Future projects and goals

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.



Left: 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.

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



Selected references

O'Donoghue, S.I., Horn, H., Pafilis, E., Haag, S., Kuhn, M., Satagopam, V.P., Schneider, R., Jensen, L.J. (2010). Reflect: A practical approach to web semantics. *Journal of Web Semantics*, SI 2-3 (8) 182-189

Gehlenborg, N., O'Donoghue, S.I., Baliga, N.S., Goesmann, A., Hibbs, M.A., Kitano, H., Kohlbacher, O., Neuweger, H., Schneider, R., Tenenbaum, D. & Gavin, A.C. (2010). Visualization of omics data for systems biology. *Nat. Methods*, 7, S56-S68

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

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 made progress in understanding how Ran controls NPC assembly and has identified other protein and lipid kinases and phosphatases involved in regulating NE assembly. Their detailed mechanisms of action are under study. In addition, although it is known that Ran regulates where NE assembly occurs in the cell, it is not known 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.

The RanGTPase as a spatial regulator

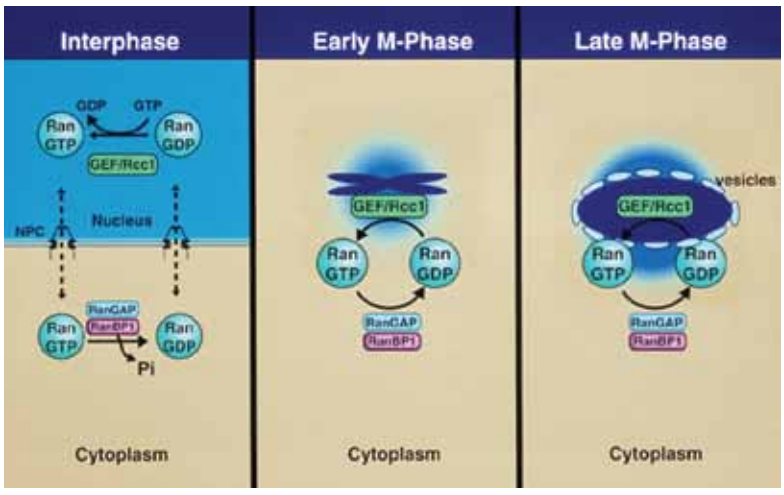
Previous and current research

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.

Selected references

Mattaj, I.W. & Muller, C.W. (2010). Solving the NES problem. *Nat. Struct. Mol. Biol.*, 17, 1288-1289

Yavuz, S., Santarella-Mellwig, R., Koch, B., Jaedicke, A., Mattaj, I.W. & Antonin, W. (2010). NLS-mediated NPC functions of the nucleoporin Pom121. *FEBS Lett.*, 584, 3292-3298

Mattaj, I.W. (2010). Coordinated action needed for Europe's research programmes. *Nature*, 465, 1005

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

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, including the regulation of mRNA translation and stability. We are elucidating these regulatory mechanisms executed by RNA-binding proteins and/or miRNA complexes. We use mostly biochemical approaches and mammalian, yeast and *Drosophila* model systems. We have also initiated a system-wide exploration of the scope and functions of 'REM networks' (Hentze and Preiss, 2010), which we expect to connect cell metabolism and gene expression in previously unrecognised ways (figure 1).

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, University Hospital Heidelberg). We also study the role of miRNAs in cancer and other diseases (with Andreas Kulozik and Martina Muckenthaler, University Hospital Heidelberg). 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.

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 explore, define and understand REM networks;
- 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 online at: www.embl.de/research/partnerships/mmpu.

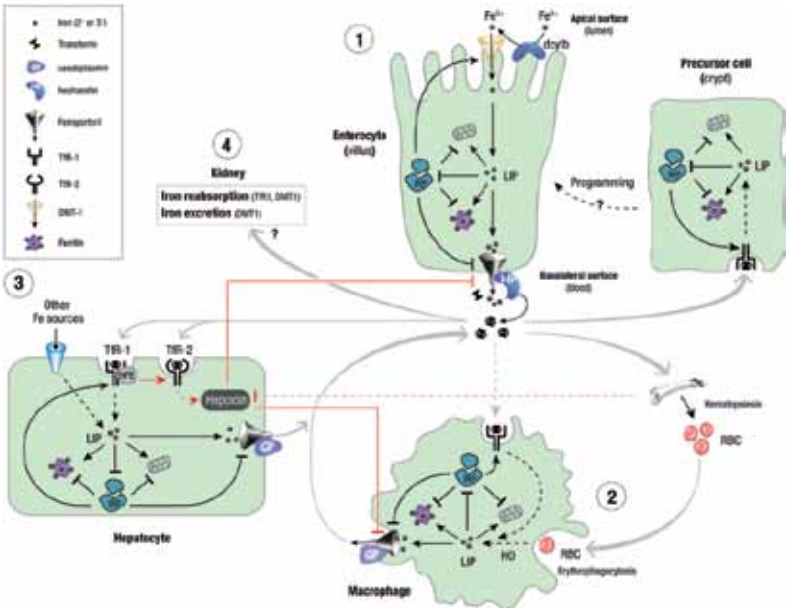


Figure 2: Systems biology of mammalian iron metabolism.

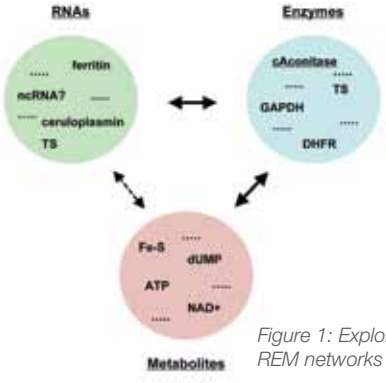


Figure 1: Exploring REM networks

Selected references

Galy, B., Ferring-Appel, D., Sauer, S.W., Kaden, S., Lyoumi, S., Puy, H., Kolker, S., Grone, H.J. & Hentze, M.W. (2010). Iron regulatory proteins secure mitochondrial iron sufficiency and function. *Cell Metab.*, 12, 194-201

Hentze, M.W. & Preiss, T. (2010). The REM phase of gene regulation. *Trends Biochem. Sci.*, 35, 423-426

Hentze, M.W., Muckenthaler, M.U., Galy, B. & Camaschella, C. (2010). Two to tango: regulation of Mammalian iron metabolism. *Cell*, 142, 24-38

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.

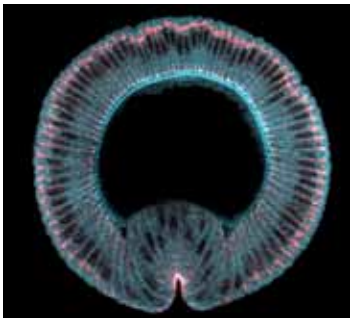
Terminal cells, like neurons, are extensively branched. The mechanisms involved in establishing their cellular architecture, especially the oxygen-carrying lumen, are poorly understood, as is also the case for related events in vertebrate tissues. Tracheal cell morphology depends on environmental signals 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.

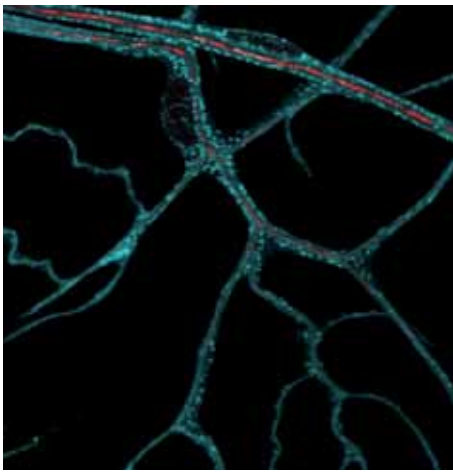
Future projects and goals

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.



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.*



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).

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.*, 8, 2017-27

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

EMBL's Core Facilities play a crucial role in enabling scientists to achieve ambitious research goals in a cost effective way. Following the establishment of a small set of facilities in 2001, the support of EMBL Council has enabled significant expansion, with the development of a number of high-level support teams that help focus diverse sets of expertise and multiple expensive technologies on specific biological problems. Currently, facilities cover the following areas: Advanced Light Microscopy, Chemical Biology, Electron Microscopy, Flow Cytometry, Genomics, Monoclonal Antibodies, Protein Expression and Purification, and Proteomics. In line with EMBL's mission to provide services to Member States, Core Facilities are open to both internal and external scientists, who benefit significantly from our contributions and advice and are able to conduct research at and beyond normal state-of-the-art.

Core Facilities are staffed by technology experts who focus entirely on service provision, delivering technologies to be used in research projects designed and run by others. Each is run by a Head of Facility who is responsible for daily operations and ensuring high user satisfaction. Close attention is given to the delivery of quality services, fast reaction times to user demands, affordable prices and the complete integration of Core Facilities with the scientific objectives of EMBL.

Such attributes are enhanced by a user committee, which consists of representatives of EMBL's research units. The committee helps to ensure that support activities are tailored to the demands of the research community, supports the introduction of new services, helps to define future strategies and provides valuable feedback on current operations. Regular internal user surveys have revealed a high level of user satisfaction. Furthermore, Core Facilities were externally reviewed (SAC and external experts) for the second time in 2010 and the overall performance was perceived as excellent and the technologies and services offered were described as 'of the highest quality'.

The EMBL model for Core Facilities has developed a first-rate reputation in the European life sciences community. They contribute significantly to internal and external training courses and workshops, often in collaboration with industrial partners. Moreover, institutions in Member States frequently seek our advice and guidance in setting up their own core facilities and services to enhance the efficiency and effectiveness of their scientific research.

Christian Boulin
Director, Core Facilities and Services

Rainer Pepperkok

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

Advanced Light Microscopy 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*
- Cytoo*
- Eppendorf*
- Lambert Instruments*
- Leica Microsystems*
- Olympus Europe*
- Perkin Elmer*
- PicoQuant
- Pro Cellcare
- Scientific Volume Imaging
- Visitron*

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

Selected references

Neumann, B., Walter, T., Heriche, J.K., Bulkescher, J., Erfle, H., Conrad, C., Rogers, P., Poser, I., Held, M., *et al*. (2010) Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature,* 464, 721-727

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

Joe Lewis

PhD 1991, Institute of Molecular Pathology, Vienna.
Postdoctoral research at EMBL.
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. Our aim 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) 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

Our new 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: as 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.

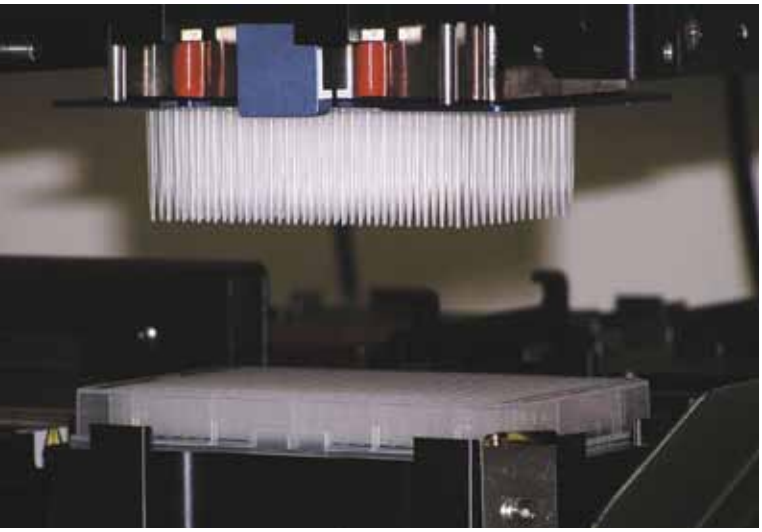
Further services include:

- 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.

Technology partners

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

Parallel pipetting of samples in 384-well format.



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

The EMCF gives EMBL scientists the opportunity to learn sample preparation for EM and provides training on advanced electron microscopes and specialised instrumentation, in particular the electron tomography setup. These techniques can be applied and adapted to various projects across the different units to achieve EM resolution at the level of cell organisation. The facility is also developing correlative microscopy approaches for cellular and developmental biology questions.

Major projects and accomplishments

Our electron tomography equipment, operational since 2008, includes a new microscope and computing set-up with programs for 3D reconstruction and cellular modelling. The microscope is a FEI F30 (300 kV microscope with a Field Emission Gun and Eagle FEI 4K camera) and is used mostly for cellular tomography of plastic-embedded samples. Nevertheless, it can also be used as a cryo-microscope. The F30 is managed by specialised EM engineers with expertise in tomography data acquisition and processing. Training is provided for researchers in handling the electron tomography microscope and its applications for cellular structure modelling. In 2010 we welcomed Rachel Mellwig, EMCF operations manger, as a new staff member.

Correlative Microscopy technology has been a fruitful approach to study the biogenesis of the Golgi apparatus (see Tangemo *et al* 2011). 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 18) are ongoing projects based on EM investigations involving electron tomography. External collaborations include the study of SPB duplication in meiotic fission yeast (K. Tanaka, University of Leicester, UK), bacterial protein expression in fission yeast (M. Balasubramanian, Singapore), and *in vivo* dengue virus replication (R. Bartenschlager group, University of Heidelberg). We have also looked at SPBs and centrioles using immunolabelling techniques: SPB-microtubule interactions in budding yeast (E. Schiebel group, DKFZ–ZMBH, Heidelberg) and centriole duplication (I. Hoffmann group, DKFZ, Heidelberg).

Services provided

- Up-to-date knowledge of EM methods for cell biology, immunocytochemistry, cryo-sectioning and cryofixation applied to various cell types or organisms.
- Maintaining 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
- Leica Microsystems
- Zeiss
- Martin Wohlwend GmbH (CH)

Selected references

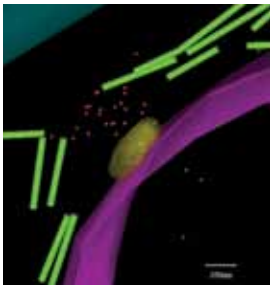
Tängemo C., Ronchi P., Colombelli J., Haselmann U, Simpson J. C., Antony C., Stelzer EHK, Pepperkok, R. and Reynaud É.G. (2011). A Novel Laser Nanosurgery Approach Provides Support for *de novo* Golgi Biogenesis in Mammalian Cells. *J.Cell Sc.*(in press)

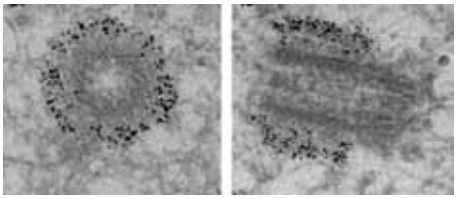
Cizmecioglu, O., Arnold, M., Bahtz, R., Settele, F., Ehret, L., Haselmann-Weiss, U., Antony, C. & Hoffmann, I. (2010). Cep152 acts as a scaffold for recruitment of Plk4 and CPAP to the centrosome. *J. Cell Biol.*, 191, 731-739

Araki, Y., Gombos, L., Migueleti, S.P., Sivashanmugam, L., Antony, C. & Schiebel, E. (2010). N-terminal regions of Mps1 kinase determine functional bifurcation. *J. Cell Biol.*, 189, 41-56

Roque, H., Ward, J.J., Murrells, L., Brunner, D. & Antony, C. (2010). The fission yeast XMAP215 homolog Dis1p is involved in microtubule bundle organization. *PLoS One*, 5, e14201

Meiotic SPB-Microtubule interaction in S.pombe *cells from a serial section 3D reconstruction and modelling using IMOD (Univ. Boulder CO). Picture by C. Funaya, collaboration with K. Tanaka.*





Immunogold labelling of Cep152, a centriolar protein in Hela cells. Picture by U. Haselmann, collaboration with I. Hoffmann J Cell Biol. *191:731-9.*

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

We offer 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 Heidelberg University'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 EMBL's scientific activities.

Technology partners

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

Deflection illumination for calibrating droplet break-off point.

Selected references

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

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

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

Vladimír Beneš

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

Genomics Core Facility 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 operated by highly-qualified staff. Training is a crucial aspect of our work and staff tutor individual researchers and organise practical courses on complementary subjects.

Major projects and accomplishments

The acquisition of new-generation sequencing technology was a vital step in ensuring EMBL remains at the forefront of European research. Since 2010, our massively parallel sequencing (MPS) suite has expanded considerably and now consists of the Illumina HiSeq2000, cBot and three GAIIx instruments. Preparation of MPS libraries is facilitated by a robust framework of other instruments (e.g. Covaris, Bioanalyzer, Qubit).

GeneCore provides the following analyses in a single-read or a paired-end sequencing mode, including multiplexing and mate-pair libraries:

- 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.

GeneCore continues to establish new protocols enabling the processing of challenging samples such as low input or metagenomics samples. For analysis of MPS data, we provide access to Genome Analyzer and Mining Station by Genomatix and work intensively with EMBL's bioinformatics community on the development of in-house tools. To date, GeneCore has generated more than one terabase of MPS data for its users.

Services provided

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

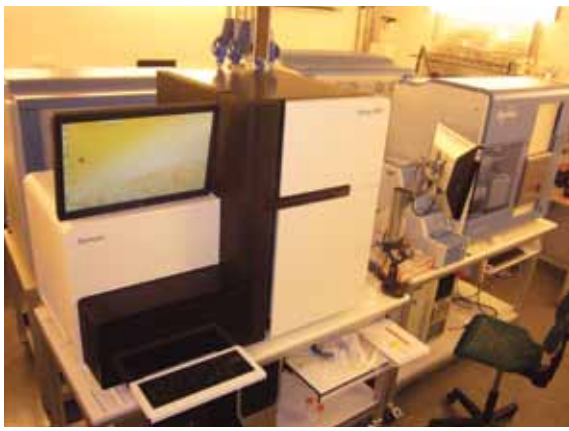
We offer processing of samples for a range of applications (mRNA, miRNA and other ncRNA expression profiling, comparative genome hybridisation, occupancy profiling) suitable for hybridisation to various microarray formats including Affymetrix and Agilent and customised arrays upon demand.

Three qPCR instruments managed by GeneCore are primarily used for gene transcript quantification to corroborate microarray results as well as detailed DNA occupancy profiling after chromatin immunoprecipitations or microRNA expression profiling. This valuable platform is complemented by an opened miQPCR system, co-invented with Mirco Castoldi (MMPU, see page 50). With our assistance EMBL researchers analysed around 150 000 qPCR assayed points in 2010. We also implemented a new application, known as high-resolution melting analysis, which enables the determination of sample homogeneity and the analysis of proteins.

Technology Partners

MPS is still a very dynamic and rapidly evolving technology. Various companies develop new products that need to be tested in our workflows.

Massively parallel sequencing suite.



Selected references

Kahramanoglou C., Seshasayee A.S., Prieto A.I., Ibberson D., Schmidt S., Zimmermann J., Benes V., Fraser G.M., Luscombe N.M. (2010) Direct and indirect effects of H-NS and Fis on global gene expression control in Escherichia coli. *Nucleic Acids Res* (epub ahead of print)

Gavriouchkina D., Fischer S., Ivacevic T., Stolte J., Benes V., Dekens MP. (2010) Thyrotroph embryonic factor regulates light-induced transcription of repair genes in zebrafish embryonic cells. *PLoS One*, 5(9), (epub)

Corbo J.C., Lawrence K.A., Karlstetter M., Myers C.A., Abdelaziz M., Dirkes W., Weigelt K., Seifert M., Benes V., Fritsche L.G., Weber B.H., Langmann T. (2010) CRX ChIP-seq reveals the cis-regulatory architecture of mouse photoreceptors. *Genome Res.*, 20(11), 1512-25

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

BSc 1988, Paddington College, London, UK.

Work at the Dept. of Molecular Oncology, Harvard Medical School, USA.

Facility head at EMBL since 2001.

Alan Sawyer

Monoclonal Antibody Core Facility

Major projects and accomplishments

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

Services provided

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

Technology Partners

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

Several biotechnology companies have licensed technologies developed at the MACF.

Microarray at the Monoclonal Antibody Core Facility

Selected references

Chiarella, P., Edelmann, B., Fazio, V.M., Sawyer, A.M. & de Marco, A. (2010). Antigenic features of protein carriers commonly used in immunisation trials. *Biotechnol. Lett.*, 32, 1215-1221

Taussig, M.J., Stoevesandt, O., Borrebaeck, C.A., *et al.* (2007). ProteomeBinders: planning a European resource of affinity reagents for analysis of the human proteome. *Nat. Methods*, 4, 13-17

Yan, X.J., Albesiano, E., Zanesi, N., Yancopoulos, S., Sawyer, A., *et al.* (2006). B cell receptors in TCL1 transgenic mice resemble those of aggressive, treatment-resistant human chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. USA*, 103, 11713-11718

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

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, mammalian cells and sera using a variety of chromatographic methods. Following each purification, we can perform biophysical analyses to ensure the quality of the purified proteins in terms of correct folding and stability. Our group also develops or evaluates new techniques or advanced protocols for protein production and purification. Most of our efforts are dedicated to developing time-saving solutions for these activities. Moreover, we are keeping stocks of a large number of expression vectors and bacterial strains for the users as well as preparing a collection of frequently used enzymes for general use which helps to considerably reduce the expenses of our users.

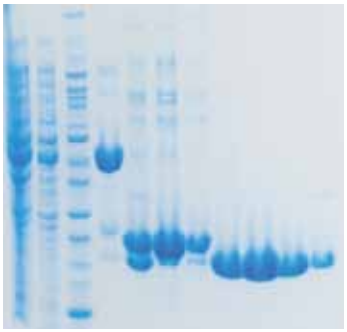
Major projects and accomplishments

We have evaluated new variants of our widely used pETM-series expression vectors for *E. coli* that can now be used for sequence and ligation-independent cloning (SLIC). We have adapted vectors for insect and mammalian cells for the same cloning protocol. Using a single PCR product with the gene of interest, it is possible to 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 different expression systems in parallel and avoid re-designing of inserts for restriction-based cloning.

We have established new vectors for expression of fusion protein based on the small SUMO proteins 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 developed a protocol for proteolytic cleavage of the urea-denatured fusion protein with the robust protease under conditions where other proteases show a poor performance. We could obtain pure, untagged proteins that were otherwise difficult to express or purify and could be used, for example, as antigens for immunisation.

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 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 website.
- 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 service capabilities of our facility. We are frequently approached by company representatives when new products become available, which we test according to the needs of our users.

Selected references

Grimm, C., Matos, R., Ly-Hartig, N., Steuerwald, U., *et al*. (2009). Molecular recognition of histone lysine methylation by the Polycomb group repressor dSfmbt. *EMBO J.*, 8;28(13): 1965-77

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

Fiedler, M., Sanchez-Barrena, M.J., Nekrasov, M., Mieszczanek, J., Rybin, V., Muller, J., Evans, P. & Bienz, M. (2008). Decoding of methylated histone H3 tail by the Pygo-BCL9 Wnt signaling complex. *Mol. Cell*, 30, 507-518

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

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 proteomics infrastructure for the identification and characterisation of proteins. This is centered around state-of-the-art mass spectrometry for MS and LC-MSMS experiments, and includes chromatographic and electrophoretic systems for protein and peptide separation.

Major projects and developments

- Molecular weight determination of intact proteins.
- Routine identification of proteins from coomassie and silver-stained gels.
- Identification of post-translational modifications.
- Nano-flow liquid chromatography coupled to high-resolution mass spectrometry: (LC-MSMS) for the identification of proteins in complex mixtures.
- Protein quantification by stable-isotope labelling (e.g. SILAC).

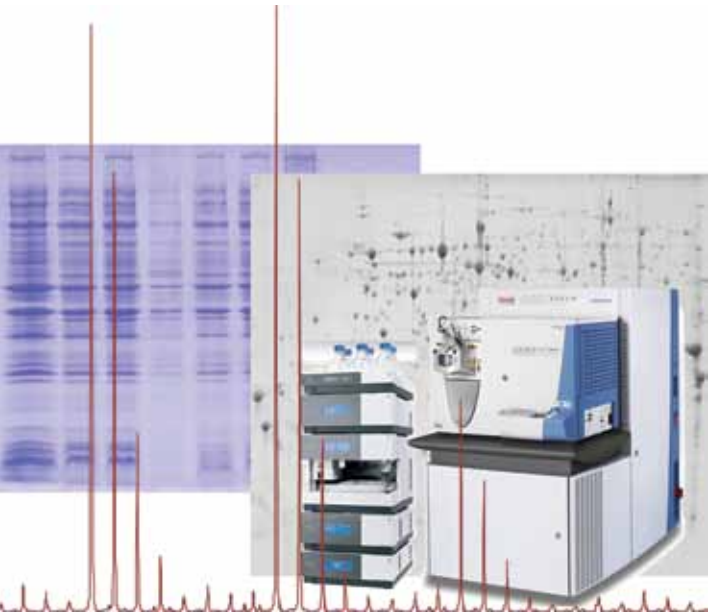
Services provided

Analysis of intact proteins:

- Molecular weight determination of intact proteins by ESI mass spectrometry (Bruker Maxis and Waters Q-tof).
- Determination of N- and C-termini of proteins and products of limited proteolysis.
- Protein separation by 1- and 2-dimensional electrophoresis using gel systems from 7 to 24 cm.
- Various gel staining protocols: Coomassie, silver, fluorescent staining by Sypro Ruby or Flamingo.

Proteomics:

- Pipeline for differential proteome analysis by 2D gels (incl. DIGE).
- Protein digestion in gel or in solution, using a variety of proteases.
- Enrichment of phosphopeptides (TiO2 and IMAC).
- Identification of post-translational modifications.
- Protein identification by MALDI peptide mass finger printing.
- Nano-flow reversed-phase chromatography (Dionex 3000 and Waters nano-Acquity systems) coupled in-line with ESI mass spectrometry.
- Ion trap (Bruker HCT) MS and MSMS for routine identification of proteins from coomassie and silver-stained gels.
- High-resolution and high mass-accuracy MS and MSMS (Thermo Orbitrap Velos) for identification and quantification of proteins in complex mixtures.
- Multi-dimensional peptide separation (isoelectric focusing and liquid chromatography);
- Protein quantification by stable-isotope labelling (SILAC and dimethyl labelling).



Technology partner

- BioRad

Selected references

Lannert, H; Franz, T; Able, T; Leicht, S; Lenze, M; Eckstein, V; Saffrich, R; Ho, AD (2008). Regulation of migration - Oncoprotein (Op18, Stathmin) overexpression in G-CSF mobilized stem cells. *Onkologie*, 31

Koyama, S., Hata, S., Witt, C.C., Ono, Y., Lerche, S., Ojima, K., *et al*. (2008). Muscle RING-finger protein-1 (MuRF1) as a connector of muscle energy metabolism and protein synthesis. *J. Mol. Biol.*, 376, 1224-1236

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

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 wide uptake of next-generation sequencing and other ultra-high throughput technologies by life scientists with interests spanning fundamental biology, medicine, agriculture and environmental science has led to unprecedented growth in data generation. Life scientists can now carry out experiments at rates previously undreamt of: for example, the International Cancer Genome Consortium plans to sequence 25 000 cancer genomes.

This has put the need for unrestricted access to biological data at the centre of biology. This is reflected in the use of EMBL-EBI's services, which exceeds 4.6 million requests a day. EMBL-EBI has a mandate to provide biomolecular data resources of universal relevance to biological and medical research; our services include the provision of biological databases and tools to explore them.

EMBL-EBI's constituency includes academic and commercial researchers throughout Europe and beyond. The EBI is at the sharp end of spectacular improvements in the speed, capacity and affordability of DNA sequencing, with submission rates exceeding half a million bases per second.

Our research groups, which comprise around a quarter of EMBL-EBI personnel, perform computational research into many different biological questions, ranging from genome evolution and transcriptional regulation to systems modelling of basic biological processes and disease. Bioinformatics continues to diversify, often led by the development of new technologies that generate the need for new methods for data handling and interpretation.

Our research groups are compact, typically with two or three students and externally funded postdocs. Their research complements the broad remit of EMBL-EBI's service provision, benefitting from the in-house technical expertise provided by the larger service teams and in turn helping to identify current challenges for researchers using our data resources. Several service teams also incorporate a small research and development component.

While EMBL- EBI's services focus on gathering and presenting comprehensive collections of information, much of our research addresses how that information is used in living organisms to choreograph the processes of life. Some approaches adopt a genome-wide approach, whereas others zoom in on specific processes.

EMBL-EBI's research aims to develop new ways to understand biology through bioinformatics. Some of this research involves the development of new resources that enable our research groups to answer these biological questions. In the EBI's spirit of open access, these resources are made openly available to other researchers.

As well as serving the biological research community by providing Europe's core biological data resources, EMBL- EBI coordinates Europe's bioinformatics service providers, effectively adding value by distributing effort. One of our most significant projects in this respect is ELIXIR, the nascent European life sciences infrastructure for biological information. ELIXIR is working to create a sustainable infrastructure for biological information in Europe. This is pivotal for academic and industrial research, and is of vital importance if Europe is to retain a competitive pharmaceutical sector.

The research we do seeks new methods for handling and analysing biological data, and perhaps most important of all, creates the mechanism to allow the generation of new hypotheses to understand biological processes. These in turn lead to a much better and deeper insight into how biology works and will provide the bedrock for new discoveries in the future.

Janet Thornton
Director, EMBL-EBI

Janet Thornton

PhD King's College and National Institute for Medical Research, London, UK, 1973.

Postdoc at the University of Oxford, NIMR and Birkbeck College. Lecturer, Birkbeck College, 1983-1989. Prof. of Biomolecular Structure, University College London (UCL) since 1990. Bernal Prof. at Birkbeck College, 1996-2002. Director of the Centre for Structural Biology, Birkbeck College and UCL, 1998-2001. Director of EMBL-EBI since 2001.

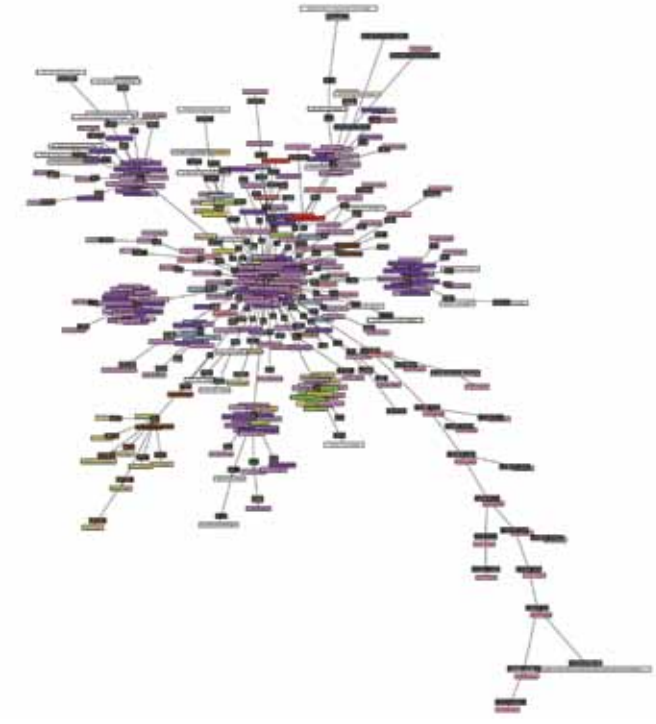
Computational biology of proteins: structure, function and evolution

Previous and current research

Our goal is to understand more about how biology works at the molecular level, with a particular focus on proteins and their three-dimensional (3D) structure and evolution. Currently we are exploring how enzymes perform catalysis and how these molecules recognise their cognate ligands. This involves gathering relevant data from the literature and developing novel software tools to characterise enzyme mechanisms and to navigate through catalytic and substrate space. In parallel we are investigating the evolution of these enzymes to discover how one enzyme can evolve new mechanisms and new specificities. This involves the integration of heterogeneous data with phylogenetic relationships within protein families, which are based on protein-structure classification data derived by colleagues at University College London (UCL). A further collaborative study on membrane proteins has focused on the structure of transmembrane pores and how this determines their different specificities. The practical goal of this research is to improve the prediction of function from sequence and structure and to enable the design of new proteins or small molecules with novel functions. The group is also interested in gaining a deeper understanding of the molecular basis of ageing in different organisms through a strong collaboration with experimental biologists at UCL. Our role is to analyse functional genomics data from flies, worms and mice and relate these observations to effects on life span by combining information on function, context (i.e. pathways and interactions) and evolutionary relationships.

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 impacts on structure, function and disease. We will seek 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. 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.



ArchSchema representation of multi-domain architectures containing a NAD(P)-binding, Rossmann-like domain. Each primary node in the graph shows a set of coloured bars, each corresponding to a structural domain, as identified by Gene3D. The NAD(P)-binding, Rossmann-like domains are shown as green bars. The central architecture of a single green domain represents all protein sequences in UniProt that contain just this domain. Linked to this are architectures having progressively more domains added. A red bar below any domain(s) indicates which parts of the proteins have 3D structural information in the Protein Data Bank (PDB). The coloured satellite nodes represent different enzymes' functions. Each colour corresponds to a different Enzyme Commission class; the deeper the hue, the more protein sequences there are. The network shows how function arises from addition of specific domains, or where a wide range of functions can be achieved by a single architecture.

Selected references

Fischer J.D., *et al.* (2010) The structures and physicochemical properties of organic cofactors in biocatalysis. *J. Molecular Biol.* (in press)

Oliva, R., *et al.* (2010) Electrostatics of aquaporin and aquaglyceroporin channels correlates with their transport selectivity. *Proc. Nat. Acad. Sci.,* 107, 4135-4140

Tamuri, A.U. and Laskowski, R.A. (2010) ArchSchema: a tool for interactive graphing of related Pfam domain architectures. *Bioinformatics,* 26, 1260-1261

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

Paul Bertone

PhD Yale University, 2005.
At EMBL–EBI since 2005.
Joint appointments in Genome Biology and Developmental Biology Units.

Pluripotency, reprogramming and differentiation

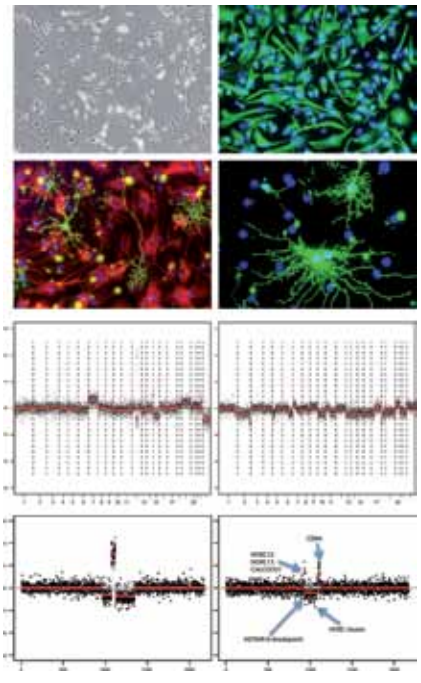
Previous and current research

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

A number of molecular pathways involved in embryonic development have been elucidated, including those that influence 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, only a small number of components have been identified and large-scale characterisation of cellular commitment and terminal differentiation to specific cell types remains incomplete. Our research group applies the latest high-throughput technologies to investigate the functions of key regulatory proteins and their influence on the changing transcriptome. We focus on early lineage commitment of ES cells, neural differentiation and nuclear reprogramming. The generation of large-scale data from functional genomic and proteomic experiments will help to identify and characterise the regulatory influence of key transcription factors, signalling genes and non-coding RNAs involved in early developmental pathways, leading to a more detailed understanding of the molecular mechanisms of vertebrate embryogenesis.

Future projects and goals

We will continue working to understand the molecular mechanisms that support pluripotency in ground-state embryonic stem cells, and to map the transition between the pluripotent state and early lineage commitment. 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 the progressive restriction of 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 will also further characterise the molecular properties of neural cancer stem cells, and assess the role of genetic aberrations and variation across individuals in the multipotent capacity of cell lines of different origins. This will involve genome and transcriptome sequencing, time-course expression profiling and functional experiments to identify alterations in disease versus normal cell types.



Neural cancer stem cells propagate indefinitely in culture (top) and can differentiate into the major cell types of the central nervous system, such as astrocytes and oligodendrocytes (second row). Array CGH and genome resequencing identify chromosomal abnormalities (third row) and the disruption of genes affected by them (bottom row).

Selected references

Git, A., *et al*. (2010) Systematic comparison of microarray profiling, real-time PCR and next-generation sequencing technologies for measuring differential microRNA expression. *RNA,* 16, 991-1006

Salmon-Divon, M., *et al*. (2010) PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. *BMC Bioinformatics,* 11, 415

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

Anton Enright

PhD in Computational Biology, University of Cambridge, 2003.

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

At EMBL-EBI since 2008.

Functional genomics and analysis of small RNA function

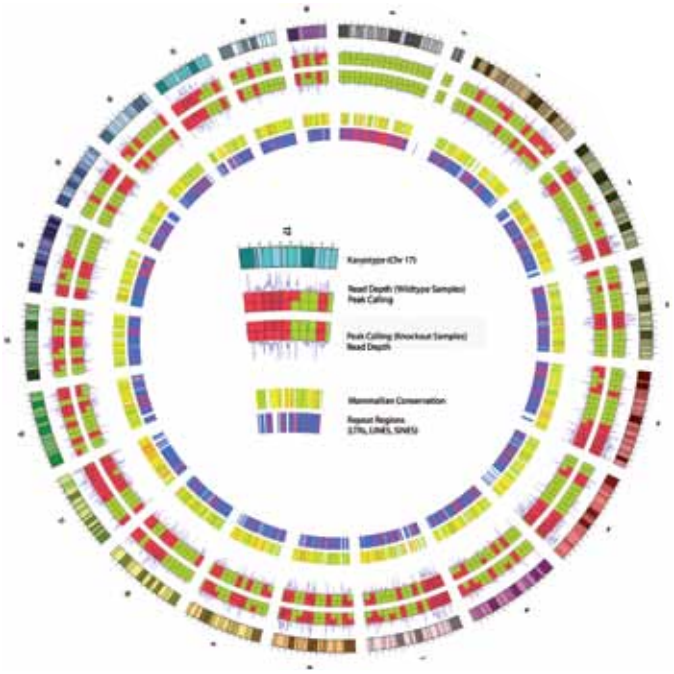
Previous and current research

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. Our group aims to predict and describe the functions of genes, proteins and regulatory RNAs as well as 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, protocols and datasets for functional genomics. Our research currently focuses on determining the functions of regulatory RNAs. 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 and perform relevant experiments to 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. By building these integrated networks we hope to place miRNAs into a functional context that will help us better understand the function and importance of these regulatory molecules. We will also cross-compare miRNA alignments with likely target alignments to identify possible cases of correlated evolution where changes in miRNA sequence are compensated by changes to the regulatory target. We will continue our work on piRNAs, to build an accurate database of piRNA loci in animals and to explore the importance and evolution of these 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. 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.



piRNA sequencing data from two wild-type and two mutant mice. The mouse karyotype is shown around the outside of the plot. Sequencing reads from both sets of samples are shown as line plots. Hotspots for piRNA biogenesis are illustrated as red squares on the middle tracks. The two inside tracks show mammalian conservation and repetitive genomic elements (innermost).

Selected references

Bartonicek, N. and Enright, A.J. (2010) SylArray: A web-server for automated detection of miRNA effects from expression data. *Bioinformatics,* (in press)

Guerra-Assuncao, J.A. and Enright, A.J. (2010) MapMi: automated mapping of microRNA loci. *BMC Bioinformatics,* 11, 133

Rasmussen KD., *et al*. (2010) The miR-144/451 locus is required for erythroid homeostasis. *J. Exp. Med.,* 207, 1351-1358

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

Nick Goldman

PhD University of Cambridge, 1992.
Postdoctoral work at National Institute for Medical Research, London, and University of Cambridge.
Wellcome Trust Senior Fellow, 1995-2006.
At EMBL-EBI since 2002.

Evolutionary tools for sequence analysis

Previous and current research

Our research concentrates on the mathematics and statistics of data analyses 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; the provision of these methods to other scientists via stand-alone software and web services; and applications of such techniques, focusing on comparative genomics and the bulk analysis of biological sequence data. Collaborations with 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 2009–2010, the group has continued to gain expertise with next-generation sequencing (NGS) data. This vast source of new data promises great gains in understanding genomes but brings with it many new challenges. Our aim is to increase our understanding of the process of evolution and to provide new tools to elucidate the function of biological molecules as they change over evolutionary timescales.

Future projects and goals

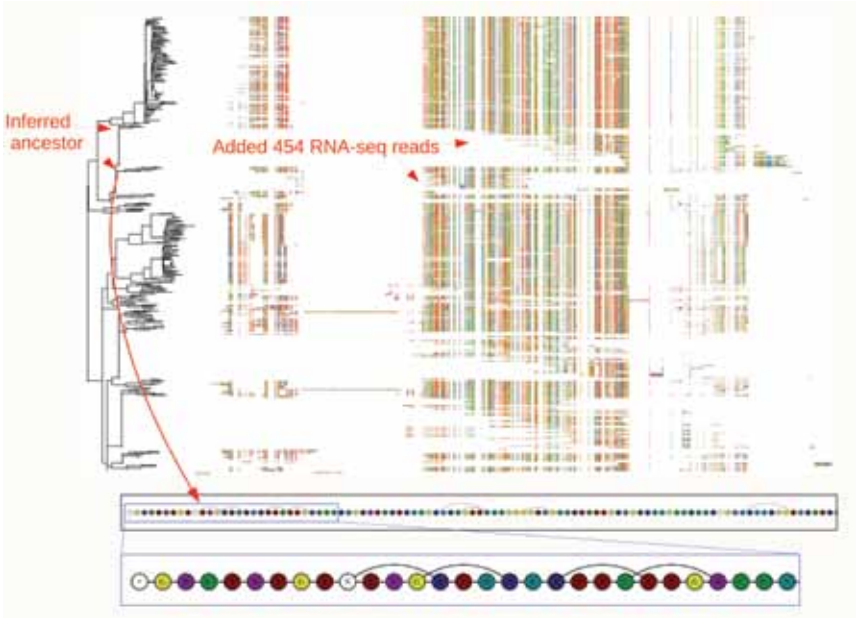
The study of genome evolution continues to inspire us with novel problems in phylogenetic methodology. The complex nature of the non-independence of sequence data due to their evolutionary relatedness continues to generate statistically challenging problems and we will continue to contribute to this theoretical field. We remain dedicated to practical applications of these methods in order to promote best practice in computational evolutionary and genomic biology, to keep in touch with the evolving needs of laboratory scientists and to continue to benefit from a supply of motivational biological questions where computational methods can help. In 2011 we will continue to expand our efforts in NGS topics; it is increasingly clear that before we can gain the full benefit of inexpensive and extensive genome sequencing we will have to devise suitable sequencing strategies, understand and allow for the significant error rates of even the most advanced sequencers and take account of the evolutionary origins of the genomes we study.

Selected references

Talavera, D., Taylor, M.S. and Thornton, J.M. (2010) The (non)malignancy of cancerous amino acidic substitutions. *Proteins,* 78, 518-529

Chor, B., *et al*. (2009) Genomic DNA K-Mer spectra: models and modalities. *Genome Biol.,* 10, R108-R108

San Mauro, D., *et al*. (2009) Experimental design in caecilian systematics: phylogenetic information of mitochondrial genomes and nuclear rag1. *Syst. Biol.,* 58, 425-438



The PAGAN graph aligner can reconstruct sequence history for an existing alignment, and align NGS RNA-seq reads to reconstructed ancestors (top). The input reads can be in FASTQ format and paired-ended; homopolymer errors can be modelled and low-quality bases trimmed. Reads can be assigned to a node or a set of nodes, or the best node can be searched for either among all nodes or within a predefined subset. PAGAN is included in the Pachinko analysis pipeline (Albert Vilella, EBI/Compara) that allows the assembly of RNA transcripts from relatively low-coverage data and without a closely related genome sequence. Alternative possible ancestral sequences are recorded as graphs (bottom) that permit the transfer of both inferred alignments and uncertainty of true insertion-deletion history throughout the phylogeny.

Nicolas Le Novère

PhD, Pasteur Institute, Paris,1998.

Postdoctoral research at the University of Cambridge, UK, 1999-2001.

Research fellow, CNRS, Paris, 2001-2003.

At EMBL-EBI since 2003.

Computational systems neurobiology

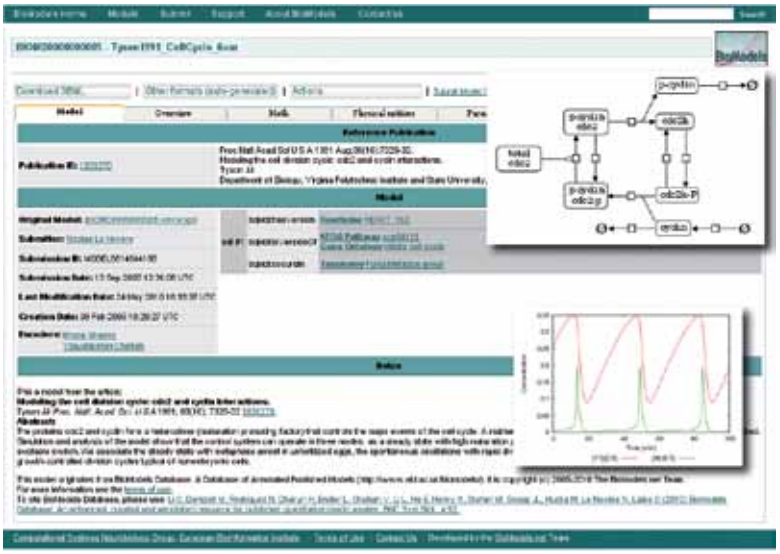
Previous and current projects

Our 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. 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 bi-stability.

The group provides community services that facilitate research in computational systems biology. For example, we lead the development of standard representations, encoding and annotating schemes, tools and resources for kinetic models in chemistry and cellular biology. The Systems Biology Markup Language (SBML) is designed to facilitate the exchange of biological models between different software. The Systems Biology Graphical Notation (SBGN) is an effort to develop a common visual notation for biochemists and modellers. We also develop standards for model curation (e.g. MIRIAM, MIASE) and controlled vocabularies (e.g. SBO, the Systems Biology Ontology) to improve model semantics. In order to manage perennial cross-references, we are developing MIRIAM Resources and its associated URN scheme. Finally, the BioModels Database is the reference resource where scientists can store, search and retrieve published mathematical models of biological interest, launch online simulations or generate sub-models.

Future projects and goals

The activity of the neurobiology side of the group will expand to cover the signalling pathways involved in synaptic plasticity more comprehensively. While the emphasis will remain on biochemistry, whole-neuron behaviours will be incorporated, in particular electrophysiology. On the technology side, the software infrastructure running the BioModels Database will be rewritten to cope with new challenges (e.g. size and type of models, authentication and security, easy deployment). Concerning the content, we will extend the support to other types of models (e.g. PK/PD models) and new formats. MIRIAM Resources will be updated in order to support more data types and to provide improved resolution services.



The BioModels Database: main page of a model of the cell cycle, together with the reaction graph and the reproduction of published results by our curators.

Selected references

Li, C., *et al*. (2010) BioModels Database, enhanced curated and annotated resource of published quantitative kinetic models. *BMC Systems Biol.,* 4, 92

Tolle, D. and Le Novère, N. (2010) Brownian diffusion of AMPA receptors is Sufficient to explain fast onset of LTP. *BMC Systems Biol.,* 4, 25

Edelstein, S.J., Stefan, M.I. and 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

Nicholas Luscombe

PhD University College London, 2000.

Postdoc at the Department of Molecular Biophysics & Biochemistry, Yale University, USA.

At EMBL-EBI since 2005.

Joint appointment with Gene Expression Unit, EMBL-Heidelberg.

Genomics and regulatory systems

Previous and current research

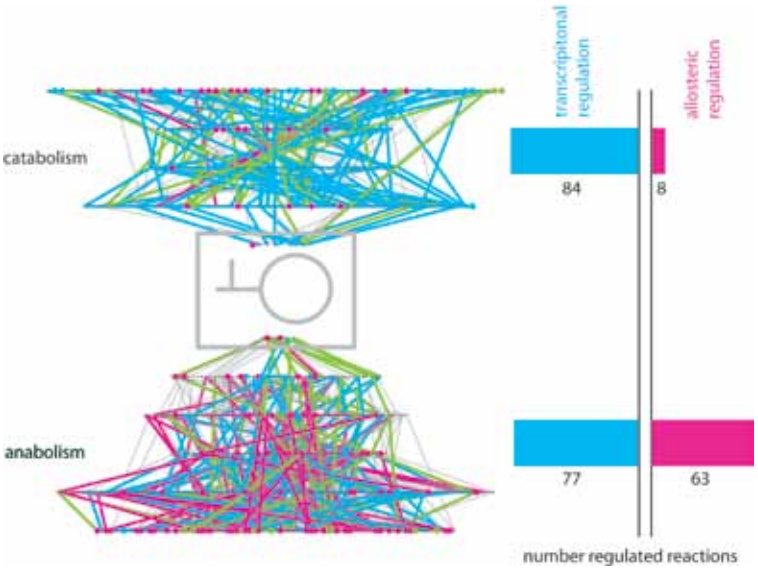
Cellular life must recognise and respond appropriately to diverse internal and external stimuli. By ensuring the correct expression of specific genes at the appropriate times, the transcriptional regulatory system plays a central role in controlling many biological processes: these range from cell-cycle 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.

Much of our basic knowledge of transcriptional 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 – and continue to – generate information describing the function and organisation of regulatory systems on an unprecedented scale. Genomic studies now allow us to examine the regulatory system from a whole-organism perspective. However, observations made with these data are often unexpected and appear to complicate our view of gene expression control.

The rising flood of biological data demands the application of computational methods to answer many interesting questions. The combined strength of bioinformatics and genomics gives us the ability to uncover general principles, providing global descriptions of entire systems. Research in the Luscombe Group is dedicated to understanding how transcription is regulated and how this regulatory system is used to control biologically interesting phenomena. We work on two major groups of organisms in parallel: higher eukaryotes and bacteria.

Future projects and goals

We will continue our analysis of genome-scale data to understand how transcription is regulated and how it is used to control interesting systems. A major focus continues to be our close interactions with research groups performing functional genomic experiments.



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

Selected references

Konig, J., *et al.* (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.*, 17, 909-915

Reimand, J., *et al*. (2010) Comprehensive reanalysis of transcription factor knockout expression data in *Saccharomyces cerevisiae* reveals many new targets. *Nucleic Acids Res.*, 38, 4768-4777

Seshasayee, A.S.N., Fraser, G.M. and Luscombe, N.M. (2010) Comparative genomics of cyclic-di-GMP signaling in bacteria: post-translational regulation and catalytic activity. *Nucleic Acids Res.*, 38, 5970-5981

Vaquerizas, J.M., *et al*. (2010) Nuclear pore proteins NUP153 and Megator define transcriptionally active regions in the *Drosophila* genome. *PLoS Genet.*, 6, e1000846

PhD in Applied Mathematics,
University of Cambridge, 2008.

Postdoctoral research in the Department
of Human Genetics, University of Chicago.

At EMBL since September 2010.

John Marioni

Computational and evolutionary genomics

Previous and current projects

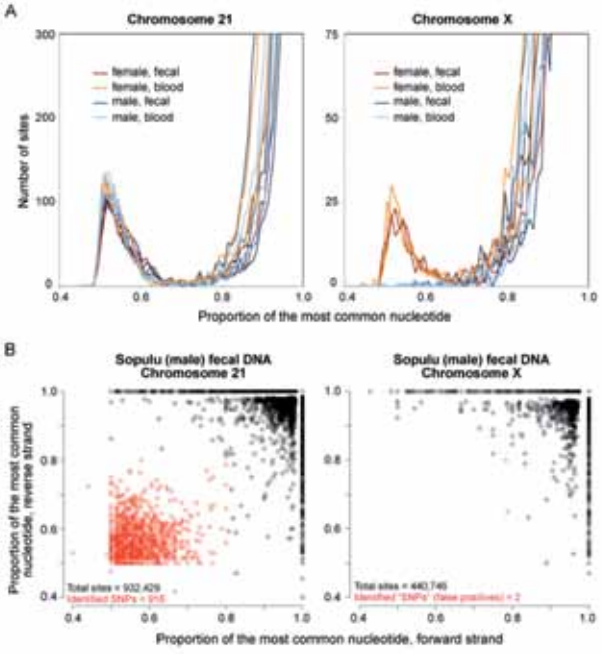
Our work focuses on the development of computational and statistical methods to answer pertinent questions in evolutionary biology. We apply the approaches we develop to better understand how changes in genetic architecture (i.e., the relationship between DNA sequencing and gene expression) provide insights into the adaptive evolution of complex organisms and developmental processes.

Next-generation sequencing has revolutionised genetics, facilitating the generation of high-resolution maps of entire genomes, transcriptomes, and regulatory features (e.g., transcription factor binding sites) - to make the most of the opportunities afforded by these technologies, it is essential to develop effective statistical techniques.

Our work, while focused on the development of computational methods, depends upon close collaborations with outstanding empirical labs such as Detlev Arendt's group (brain evolution, see page 19), Duncan Odom's laboratory at the CRUK-CRI (modelling gene regulation in mammals), and Azim Surani's group at the Gurdon Institute (understanding mouse embryogenesis): we work together to frame biological questions of interest, to design appropriate studies, and to analyze and interpret the data generated.

Future projects and goals

As sequencing technology continues to develop, it is becoming possible to move from studying genetic variation at the tissue level to studying genome-wide differences at the single-cell level. Data generated using single-cell approaches will be critical for understanding the etiology of cancer, early developmental processes, and the functioning of the brain amongst other biological processes where heterogeneous cell populations are an important feature. To make the most of these opportunities, developing methods for storing, visualising, interpreting, and analysing the data generated is critical – this will be a significant focus of research in the forthcoming years. In addition, we will continue to work on methods for analysing conventional next generation sequencing data, building upon work that we have performed previously. This work will focus on methods for normalising the data to ensure samples from different individuals are comparable, as well as the development of statistical approaches for downstream analyses of the data..



Identifying SNPs in chimpanzees from DNA extracted from faecal samples using capture and sequencing DNA technology.

(A) Frequency distributions of the proportion of the most common nucleotide at each targeted site, separately by chromosome. For the overwhelming majority of sites, the most common nucleotide proportion equals 1 (the Y axis is cut off). There is a dearth of sites with intermediate-proportion nucleotides on the X chromosome in male samples.

(B) Plots of the most common nucleotide proportion by mapped strand, for each site with filtered read coverage ≥ 10 on each strand for one selected sample, separately by chromosome. Heterozygous sites were identified as those with most common nucleotide proportion ≤ 0.8 on both strands (red circles).

Dietrich Rebholz-Schuhmann

PhD in Immunology, University of Düsseldorf, 1989.

Senior scientist at GSF, Munich, 1995.

Director Healthcare IT, LION Bioscience AG, Heidelberg, 1998.

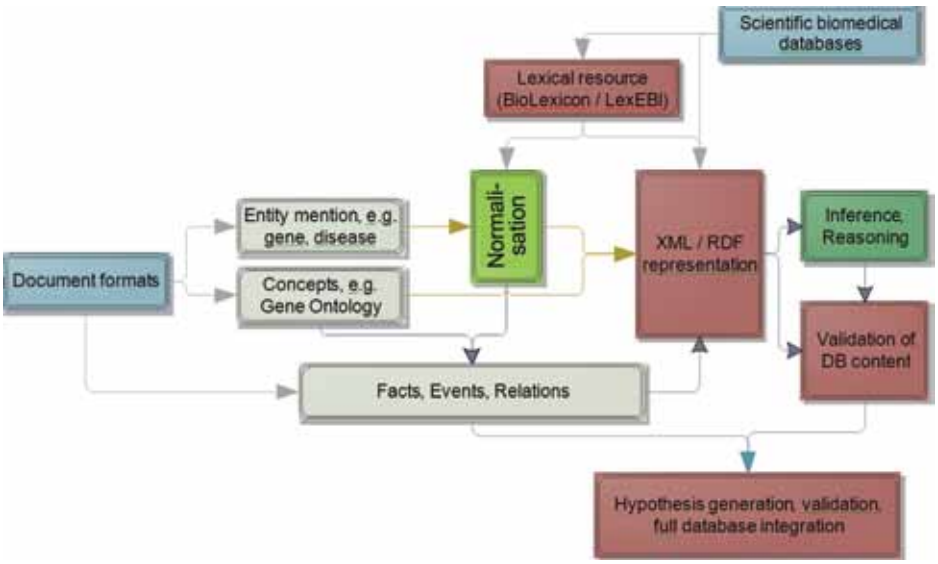
At EMBL-EBI since 2003.

Genome-scale analysis of regulatory systems

Previous and current research

Text mining comprises the fast retrieval of relevant documents from the whole body of the scientific literature and the extraction of facts from these texts. Text-mining solutions are becoming mature enough to be automatically integrated into workflows for research and into services for the general public, for example delivery of annotated full text documents as part of UK Pubmed Central (UKPMC).

Research in our group focuses on building semantic resources, extracting facts from the literature and knowledge discovery. Our goal is to connect literature content automatically to biomedical databases and to evaluate new findings against existing knowledge. Our work is split into three tightly coupled parts: named entity recognition (genes, diseases) and its quality control (e.g. UKPMC project); fact extraction and knowledge discovery (e.g. identification of gene–disease associations); and further development of the IT infrastructure for information extraction, fact delivery and reasoning across semantic resources. All solutions contribute to the standardisation and semantic enrichment of the scientific literature.



Literature analysis analyses scientific documents, identifies entities, concepts and facts (grey boxes) and normalises the entities to database entries with the support of a lexical resource (BioLexicon / LexEBI). RDF representations of the facts in combination of ontological resources supports inference and reasoning across the data content.

Selected references

Hoehndorf, R., Oellrich, A. and Rebholz-Schuhmann, D. (2010) Interoperability between phenotype and anatomy ontologies. *Bioinformatics* (in press)

Rebholz-Schuhmann, D., *et al*. (2010) CALBC Silver Standard Corpus. *J. Bioinform. Comput. Biol*., 8, 163-179

Rebholz-Schuhmann, D. and Nenadic, G. (2010) Biomedical Semantics: the Hub for Biomedical Research 2.0. *J. Biomed. Semantics,* 1, 1

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

Postdoctoral research at Harvard Medical School and Massachusetts Institute of Technology, 2007-2010.

Group leader at EMBL-EBI since 2010; joint appointment with Genome Biology Unit.

Julio Saez-Rodriguez

Computational analysis of information transfer within signalling networks implicated in disease

Previous and current projects

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.

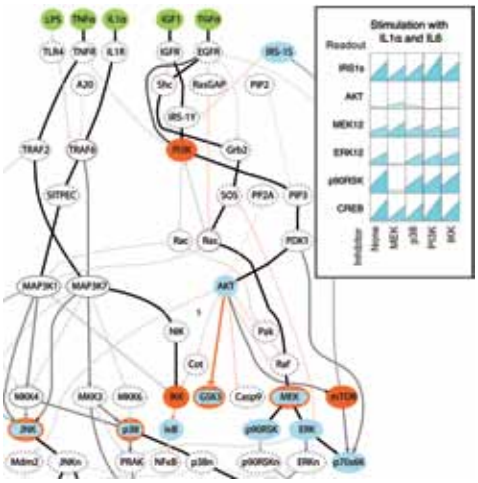
Our group develops computational methods and tools to analyse signalling networks, and we collaborate closely with experimentalists to tackle specific biological questions. We develop mathematical models that integrate high-throughput biochemical data with various sources of prior knowledge, with an emphasis on providing both predictive power of new experiments and insight on the functioning of the signalling network. To 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 as well as their application to yield insights of medical relevance. We plan to expand the palette of formalisms and levels of detail of the models, spanning very simple Boolean logic models that coarsely describe signalling networks to differential equation systems that describe 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 different sources, such as public databases, and data from proteomics, 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 signaling 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 (see Saez-Rodriguez et al. *2009 for details). 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 visualized with DataRail. Figure adapted from Saez-Rodriguez* et al. Mol. Syst. Biol. *2009 under the conditions of the Creative Commons Attribution-Non-Commercial-Share Alike 3.0 Licence.*

Selected references

Alexopoulos, L. G., Saez-Rodriguez, J., *et al*. (2010). Networks inferred from biochemical data reveal profound differences in TLR and inflammatory signaling between normal and transformed hepatocytes. *Mol. Cell., Prot,* 9(9), 1849

Saez-Rodriguez, J., Alexopoulos L. G. *et al*. (2009). Discrete logic modeling as a means to link protein signaling networks with functional analysis of mammalian signal transduction. *Molecular Systems Biology,* 5:331, 2009

Saez-Rodriguez, J, Goldsipe, A., *et al*. (2008). Flexible informatics for linking experimental data to mathematical models via DataRail. *Bioinformatics,* 24(6):840-7 2008

Kremling, A. & Saez-Rodriguez, J. (2007). Systems Biology - an Engineering Perspective. *Journal of Biotechnology*, 129, 2, 329-351

Rolf Apweiler

PhD University of Heidelberg, 1994.
At EMBL since 1987, at EMBL-EBI since 1994
Joint team leader, PANDA Group (Proteins).

Protein and nucleotide data: PANDA proteins

Previous and current research

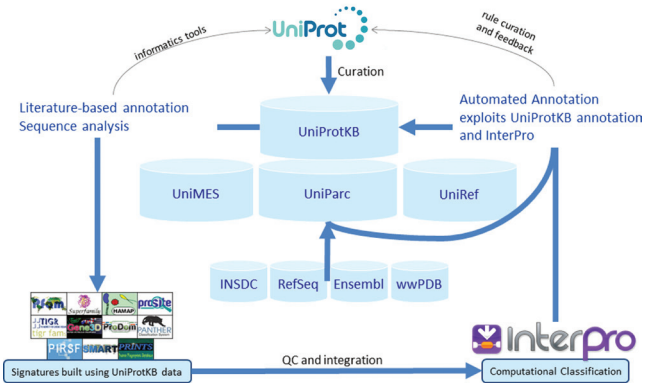
The Protein and Nucleotide Data (PANDA) group was created in June 2007 by merging the former Ensembl (Ewan Birney) and Sequence Database (Rolf Apweiler) groups. The PANDA group focuses on the production of protein-sequence, protein-family and nucleotide-sequence databases at EMBL-EBI. The Apweiler group maintains and hosts the UniProt protein resource, the InterPro domain resource and a range of other biomolecular databases. These efforts can be categorised into two major categories: proteins, and cheminformatics and metabolism. In addition to its PANDA activities, the group has a complementary research component.

The activities of the PANDA proteins teams centre on 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 ENA and Ensembl. All UniProt data undergoes annotation with Gene Ontology (GO) terms and uses the classification into protein families and domains provided by InterPro. 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 approaches to annotating all the sequence data without experimental functional data. Protein-interaction and -identification data is 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. Rolf Apweiler supervises one PhD student directly.

Future projects and goals

We intend to improve integration and synchronisation of all PANDA resources. In addition to major improvements of our current systems, we will add mining of high-throughput genomics and proteomics datasets to our automatic annotation toolset. Despite the abundance of data from large-scale experimentation on a genome-wide level (e.g. expression profiling, protein–protein interaction screens, protein localisation), the systematic and integrated use of this type of information for high-throughput annotation of proteins remains largely unexplored. We therefore intend to build on on-going 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: analysing different data types regarding their suitability for the approach; developing data structures that allow the efficient integration and mining of data of different types and quality; benchmarking the obtained results; and applying the new methodologies to UniProtKB/TrEMBL annotation.



UniProt sequence and annotation data flow.

Selected references

Apweiler, R. (2009) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.*, 38 (Database issue), D142-D148

Birney Research Group. PhD Sanger Institute, 2000.
At EMBL-EBI since 2000.
Joint team leader, PANDA Team (Nucleotides),

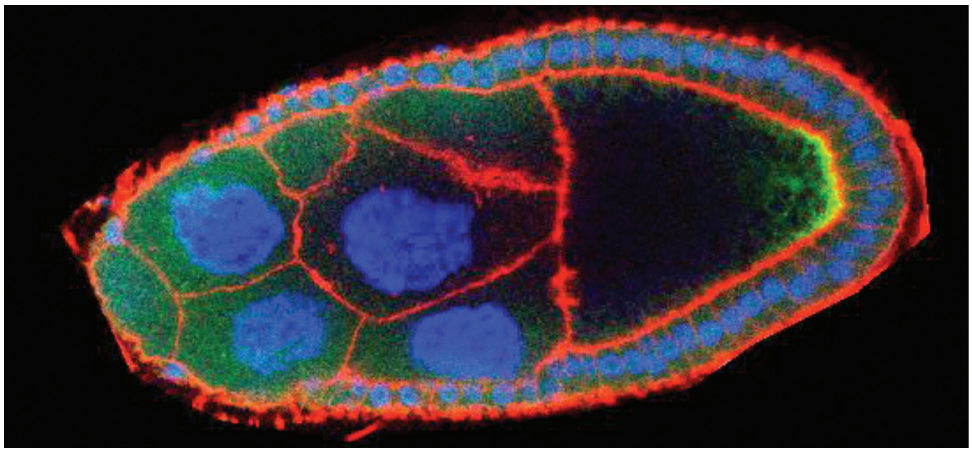
Ewan Birney

Protein and nucleotide data: PANDA nucleotides

Previous and current projects

DNA sequence remains at the heart of molecular biology and bioinformatics. Dr Birney provides strategic oversight to the DNA resources in the Panda Team, and also supervises a small number PhD students and postdocs in an independent research group. Dr Birney's research can be broken down into two main areas:

- **Development of sequence algorithms**: Over time students have developed multiple alignment, assembly and more recently sequence compression tools all involving DNA sequence. In the future Dr Birney hopes to look at segmental duplication (in particular in ways that do not rely on a reference genome), bacterial pan-genome graph behaviour and additional compression approaches.
- **Using intra-species variation to study basic biology:** For the past five years there has been a tremendous increase in the use of genome-wide association to study human diseases; however, this approach is a very general one that can be applied to nearly any measureable phenotype present on an animal with an accessible, outbred population. We are pursuing a number of both molecular (e.g. expression and chromatin levels) and basic biology (e.g. human skeleton and *Drosophila* egg chamber) measurements in different species. In the future we hope to expand this to a variety of other basic biological phenotypes in other species, ranging from marine worms through Japanese rice-fish to humans.



A Drosophila *egg chamber (oocyte) present in a near-isogenic wild* Drosophila *line. The upper large area is the egg. The cells below with large nuclei, stained blue, are the nurse cells. We are able to find genetic variations associated with the size and shape of the nurse cells.*

Selected references

Hsi-Yang Fritz, M., Leinonen, R., Cochrane, G., Birney, E. (2011) Efficient storage of high throughput sequencing data using reference-based compression. *Genome Res.,* (in press). Published online 18 January

McDaniell, R., Lee, B.K., Song, L., *et al*. (2010) Heritable individual-specific and allele-specific chromatin signatures in humans. *Science,* 328, 235-239

Alvis Brazma

PhD in Computer Science, Moscow State University, 1987.

Postdoctoral research at New Mexico State University, USA.

At EMBL-EBI since 1997.

Functional genomics

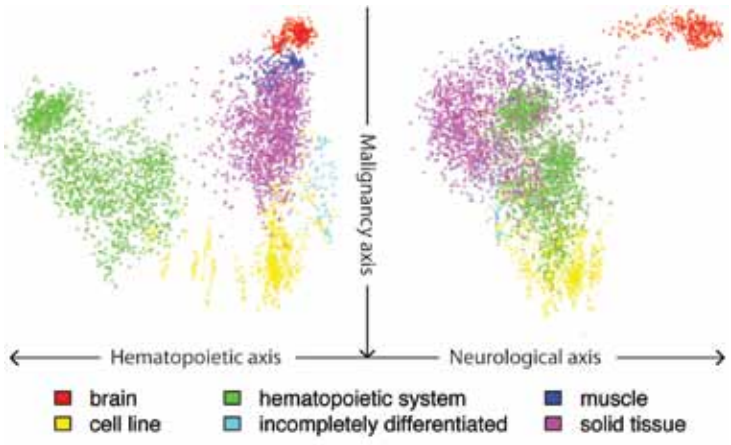
Previous and current research

The Functional Genomics Team includes the teams led by Misha Kapushesky, Helen Parkinson, Ugis Sarkans and a number of staff reporting directly to Alvis Brazma. The teams focus on functional genomics data services; research in data analysis, algorithms and methods; and research and development related to biomedical informatics. We run one of the EBI's core resources, ArrayExpress, which comprises the Archive of Functional Genomics Data and the Gene Expression Atlas. We have released a new core resource: the BioSample Database, which will hold information about all samples and phenotypes deposited in any of the core databases at EBI. Our PhD students focus on analysing functional genomics data, building models for systems biology and developing new methods and algorithms. Integration of data across multiple platforms, including genotypes, is another important area of activity. We also contribute substantially to training in transcriptomics and the general use of EBI tools.

Our on-going research projects focus on regulation of gene expression through analysis of large-scale integrated functional-genomics data sets. In particular we are focusing on high throughput sequencing based data analysis and on transcriptomic/genomic associations with human diseases. The goal is to understand how gene expression depends on molecular regulatory mechanisms as well as on genetic and experimental factors. We use integrative approaches that draw on the vast amounts of public data collected in ArrayExpress and other EBI resources, as well as participating in many international collaborations..

Future projects and goals

In 2011 we will work on populating the BioSample Database with all data from EBI assay databases, accepting reference layer datasets and securing data exchange with NCBI. We also plan to increase the throughput of the sequencing-based functional-genomics data, including RNAseq datasets into the Expression Atlas. We will continue our work with medically relevant collaborative projects. Further research into integrative data analysis will concentrate on using next-generation sequencing data, integrating genotype and gene expression data and building systems biology models.



The 5372 samples are shown as dots, colour-coded for the six major clusters identified by comparing gene-expression profiles. The left and right panels are projections of the same 3D shape viewed from two different perspectives.

Selected references

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

Lukk, M., *et al*. (2010) A global map of human gene expression. *Nature Biotechnol*., 28, 322-324

McKinney, E.F., *et al*. (2010) A Cd8+ T cell transcription signature predicts prognosis in autoimmune disease. *Nat. Med.,* 16, 586-591

Zheng-Bradley, X., *et al*. (2010) Large-scale comparison of global gene expression patterns in human and mouse. *Genome Biol.,* (in press)

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

Guy Cochrane

The European Nucleotide Archive

Previous and current projects

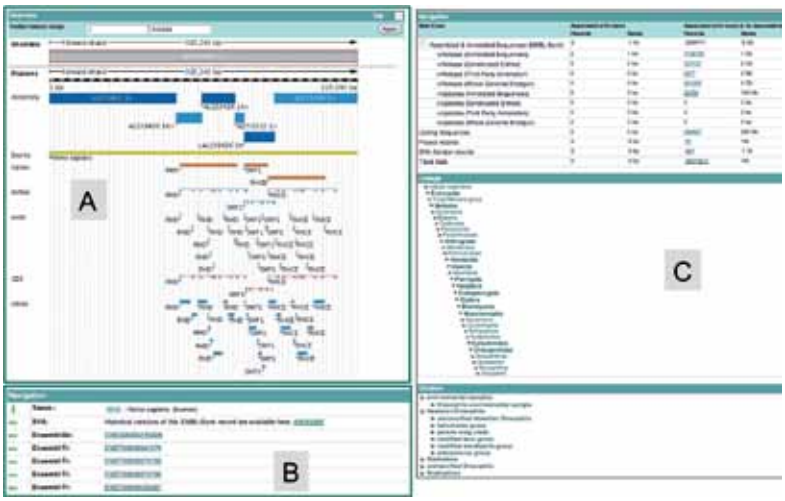
The European Nucleotide Archive (ENA) provides globally comprehensive primary data repositories for nucleotide sequencing information. ENA content spans the spectrum of data, from raw sequence reads through assembly and alignment information to functional annotation of assembled sequences and genomes. Services for data providers include interactive and programmatic submission tools and curation support. Data consumers are offered a palette of services provided both over the web and through an increasingly sophisticated programmatic interface.

Reflecting the centrality of nucleotide sequencing in the life sciences and the emerging importance of the technologies in applied areas such as healthcare, environmental and food sciences, ENA data and services form a core foundation upon which scientific understanding of biological systems has been assembled and our exploitation of these systems will develop. With an on-going concentration on data presentation, integration (within ENA and with resources beyond it), tools and services development, the team's commitment is to the utility of ENA content and the broadest reach of sequencing applications.

Future projects and goals

We will work to develop extensively the ENA's data presentation services, including the browser, our programmatic services and the internal data indexing structures that feed these services. This will require work at the level of the underlying technologies (e.g. sequence similarity search) and the provision of new services and data portals. We will take strong direction from the growing user base as to where we should focus these improvements. Our submissions services will be enhanced in a number of ways. We will develop an interactive submissions tool for smaller-scale next generation sequencing studies tailored to the needs of the smaller biology laboratory submitting community. We will develop support for third-party validation of incoming next generation sequence data to provide better integration with such submission services as that operated by ArrayExpress. Finally, we will work with our international partners to provide better coverage of genome assembly information to support the genome browser community.

The many challenges and opportunities brought by next generation sequencing technologies will continue to influence our strategic thinking. Technical developments will certainly continue to be required to support the aggressive growth of data from these platforms. In addition, we will heighten our response to the ever-increasing penetrance of sequencing as a general assay platform that has arisen from rapidly falling costs; here, we will deepen our use of our model in which domain-specific services (e.g. submission and presentation tools) are delivered collaboratively on top of more generic core repository services that are maintained solely by the ENA team.



The ENA browser. (A) assembly and functional annotation; (B) navigation box taken from an mRNA sequence record, showing links to a taxonomic record for the source organism, links to historical versions of the sequence record and external links to gene and transcript records in Ensembl; (C) taxon-centric view for a taxon, showing links to sequence data and associated information and lineage information.

Selected references

Leinonen, R., *et al*. (2010) Improvements to services at the European Nucleotide Archive. *Nucleic Acids Res.*, 38, D39-D45

Shumway, M., Cochrane, G. and Sugawara, H. (2010) Archiving next generation sequencing data. *Nucleic Acids Res.*, 38, D870-D871

Paul Flicek

DSc Washington University, 2004.
Honorary Faculty Member,
Wellcome Trust Sanger Institute since 2008.
At EMBL-EBI since 2005. Team leader since 2008.

Vertebrate genomics

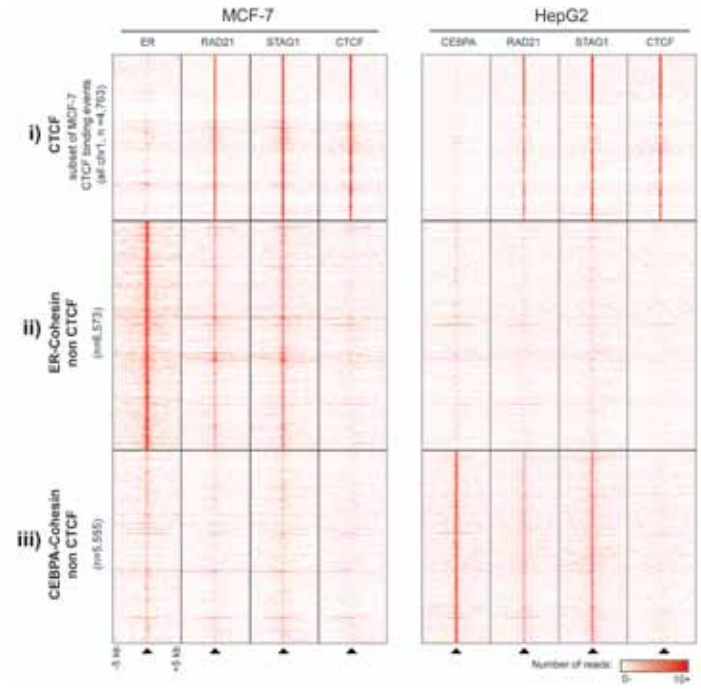
Previous and current research

We are a combined service and research team focusing on: genome annotation, archiving and distributing variation data, creating and deploying large-scale bioinformatics infrastructure, computational epigenomics and comparative regulatory genomics. Our major service projects include the Ensembl project, the European Genome-phenome Archive (EGA), EBI's mouse informatics team, the DGVa database of structural and copy-number variation and the data-management activities of the 1000 Genomes Project.

Recent research projects have used comparative functional genomics methods to investigate the evolution of transcriptional regulation across the vertebrate lineage as well as regulatory aspects of cell-type specificity. In the most significant of these projects, we directly interrogate the evolutionary mechanisms of transcription factor binding using matched experiments in liver tissues across 300 million years of evolution. Contrary to expectation there is little conservation in the location of transcription factor binding and what conservation that is observed conservation appears associated with embryonic development. Despite this significant rewiring of transcriptional regulation, these factors still manage to maintain the largely conserved gene expression and function of liver tissue. The results shed light on the apparently contradictory findings of the ENCODE project which observed that many sites of transcription factor binding are apparently not under evolutionary constraint at the sequence level.

Future projects and goals

Our research projects will continue to explore the evolution of transcriptional regulation, with experiments focused both on other DNA-binding proteins as well as on relatively short evolutionary time scales for both human and other model organisms. We are working to development algorithmic approaches to understand and compare epigenomic data between and within existing and newly sequenced genomes. The service aspects of the group over the next year will expand our support for primary analysis and assembly in addition to integrative analysis of high-throughput DNA sequence data being generated in a number of international projects. The major efforts will be devoted to the full 1000 Genomes Project, the US-led ENCODE project, the International Human Epigenome Consortium and the International Cancer Genome Consortium.



Cohesin binding events associated with CTCF are cell-type invariant (A), while cohesin locations not associated with not associated with CTCF binding are associated with tissue-specific master regulators (B, MCF7 breast cancer cells and oestroegen receptor; C, HepG2 liver cells and CEBPA). (From Schmidt, D., Schwalie, P.C., Ross-Innes, C.S., et al. *(2010) A CTCF-independent role for cohesin in tissue-specific transcription. Genome Res. 20:578-588.)*

Selected references

1000 Genomes Project Consortium. (2010) A map of human genome variation from population-scale sequencing. *Nature,* 467:7319, 1061-1073

Schmidt, D., Wilson, M.D., Ballester, B. *et al*. (2010) Five-Vertebrate ChIP-seq Reveals the Evolutionary Dynamics of Transcription Factor Binding. *Science,* 328:5981, 1036-1040

Atanur, S.S., *et al*. (2010) The genome sequence of the spontaneously hypertensive rat: Analysis and functional significance. *Genome Res.,* 20:6, 791-803

Henning Hermjakob

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

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

At EMBL-EBI since 1997.

Proteomics services

Previous and current research

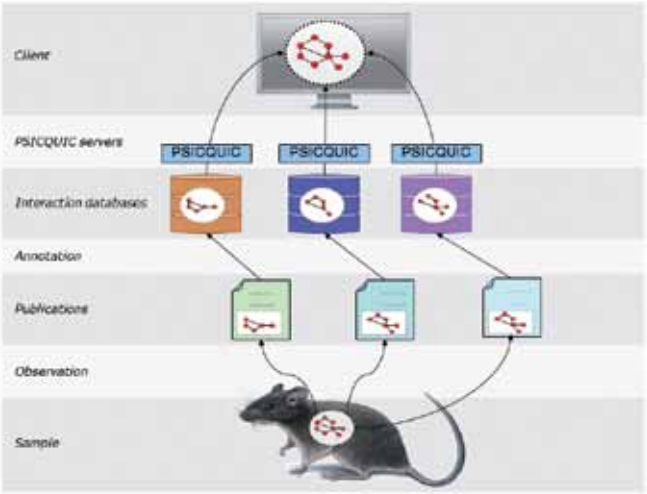
The Proteomics Services team develops tools and resources for the representation, deposition, distribution and analysis of proteomics and related data. We follow an open-source, open-data approach: all resources we develop are freely available. The team is a major contributor to the Proteomics Standards Initiative (PSI) of the international Human Proteome Organization (HUPO). We provide reference implementations for the PSI community standards, in particular the PRIDE proteomics identifications database and the IntAct molecular interaction database. We provide the Reactome pathway database in collaboration with New York University and the Ontario Institute for Cancer Research. In the context of the EU Virtual Physiological Human project, we contribute to the development of an interoperability framework that bridges physiology and molecular biology.

Future project and goals

After rapid development and achievement of major milestones in the molecular interaction domain, we need to consolidate achievements, selectively open the IMEX collaboration to new partners and develop advanced tools to take advantage of detailed IMEX curation and the integrative PSICQUIC interface. A major challenge is the complete redevelopment of the PRIDE database, necessary to cope with the rapid increase in data content but also to turn PRIDE from a publication-centric repository to a key source for protein expression information.

Beyond the technical challenges of data quantity, the two major conceptual challenges are to capture the very diverse quantitative proteomics data and to develop quality criteria to enable the selective export of high confidence PRIDE data to other resources like UniProt, Reactome or integrative data analysis tools. We plan to intensify data integration within and beyond the projects of the Proteomics Services team, in particular using web services and the DAS. We will also continue to integrate Reactome pathways and IntAct molecular interactions, as well as integrating PRIDE and IntAct, to enable efficient data deposition and navigation between molecular interactions and underlying mass spectrometry data.

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.



PSICQUIC and IMEX support a truly distributed provision of molecular interaction data: an experimental system is observed in multiple independent studies, resulting in multiple publications. Based on the collaboration in the International Molecular Exchange Consortium (IMEX), independent interaction databases curated these publications in work-sharing mode. Data is then released in PSI format through the PSICQUIC interface. A web client queries all PSICQUIC services and integrates the data on the client side.

Selected references

Martens, L., *et al.* (2010) mzML - a community standard for mass spectrometry data. *Mol. Cell. Proteomics,* (in press)

Perreau, V.M., *et al.* (2010) A domain level interaction network of amyloid precursor protein and Abeta of Alzheimer's disease. *Proteomics,* 10, 2377-2395

Lee, K., *et al.* (2010) Mapping plant interactomes using literature curated and predicted protein-protein interaction data sets. *Plant Cell,* 22, 997-1005

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

Sarah Hunter

MSc. University of Manchester, 1999.
At EMBL-EBI since 2005.

InterPro: an integrated database

Previous and current research

The InterPro team coordinates the InterPro, CluSTr and Metagenomics projects at EMBL-EBI. InterPro is used to classify proteins into families and predict the presence of domains and functionally important sites. The project integrates signatures from the major protein signature databases into a single resource, and currently includes data from Pfam, PRINTS, PROSITE, ProDom, SMART, TIGRFAMs, PIRSF, SUPERFAMILY, CATH-Gene3D, PANTHER and HAMAP. During the integration process, InterPro rationalises instances where more than one protein signature describes the same protein family/domain, uniting these into single InterPro entries and noting relationships between them where applicable. Additional biological annotation is included, together with links to external databases such as GO, PDB, SCOP and CATH. InterPro precomputes all matches of its signatures to UniProt Archive proteins using the InterProScan software, and displays the matches to the UniProt KnowledgeBase 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. 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. The new metagenomics portal is intended to provide metagenomics researchers with access to EBI's functional analysis pipelines, links to data archives (e.g. the ENA) and a web interface to manage and visualise these data.

A redeveloped InterPro website, released in early 2011, is more intuitive for users. The new InterProScan has an entirely different Java- and database-based architecture at its core, and features improved algorithms for transmembrane helix and signal peptide prediction.

Future projects and goals

The focus now for the metagenomics portal is the design and implementation of a web interface to allow users who have submitted their data for analysis and archiving to be able to manage and interact with their data. The basic interface was rolled out in early 2011, and further improvements and enhancements – including better submission and visualisation tools – are scheduled throughout the year.

Evolution of the new designs for the InterPro web interface, from the original site (a) through to user-guided prototyping (b) and an initial implementation (c). Further user testing and refinement of the design are scheduled throughout 2011.

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

BA in Mathematics and Comparative Literature,
Cornell University, NY, USA, 2000.

PhD in Genetics,
University of Cambridge, UK, 2010

At EMBL-EBI since 2001

Misha Kapushesky

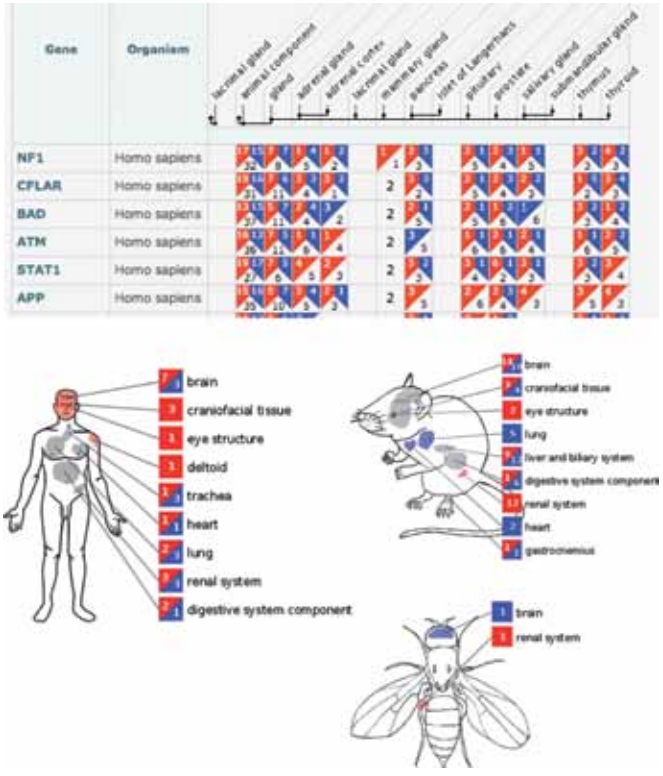
Functional genomics atlas

Previous and current research

The Functional Genomics Atlas Team develops and runs the Expression Atlas database and the R Cloud service. The Expression Atlas is a value-added database of transcriptomics datasets, providing semantically rich searches and visualisations of gene activity in curated public data from the ArrayExpress Archive. The R Cloud is the cloud-computing infrastructure used by the Atlas and offered as a remotely accessible R statistical analysis environment service to external users. We provide the Expression Atlas as stand-alone software capable of storing various types of -omics data. Through collaborations with the European Nucleotide Archive (ENA) group (page 74), we have developed support for storing next generation sequencing studies in the Atlas. The Atlas Team conducts research in the area of functional genomics data analysis and integration with our collaborators in the EU-funded SYBARIS project on biomarkers of antifungal drug resistance and disease susceptibility.

Future project and goals

Stand-alone Atlas deployment leads naturally towards the development of a distributed, federated query model. We have built a first prototype of the Distributed Atlas and plan to expand on this project, integrating diverse multiomics data types starting with transcriptomics, proteomics and eventually adding metabolomics datasets. Building on the analytics back-end of the Atlas, we plan to extend it to support complex, multifactorial experiment designs, gene set enrichment analysis with published gene signatures and a sample-based similarity search across Atlas experiments. R Cloud will be further integrated with the Atlas to promote easier online data sharing, processing and analysis. Next generation sequencing data support features prominently in Atlas plans, as well as integration with other EBI resources (e.g. Ensembl, ENA, EGA). Together with the Rebholz group (page 69), we are also developing semantic web features (e.g. RDF export) for the Atlas.



One year after the launch of the Gene Expression Atlas service at the EBI, we open-sourced the stand-alone Atlas software used to run it. Using monthly downloadable public software, data and ontology releases, users can install the complete Atlas system locally and use it to load and view private datasets together with the public data. The Atlas includes the global map of human gene expression as well as a pipeline for processing RNA-seq studies. The R Cloud service, launched in summer 2010, provides remote access to Atlas data and the R cloud-computing environment on EBI servers.

Selected references

Goncalves, A., *et al*. (2010) ArrayExpressHTS: distributed computing for RNA-seq data processing and quality assessment. *Bioinformatics*, (in press)

Kapushesky, M., *et al*. (2010) Gene Expression Atlas at the European Bioinformatics Institute. *Nucleic Acids Res.*, 38 (Suppl 1), D690-698

Lukk, M., *et al.* (2010) A global map of human gene expression. *Nat. Biotechnol.*, 28, 322-324

Malone, J., *et al.* (2010). Modeling sample variables with an Experimental Factor Ontology. *Bioinformatics*, 26, 1112-1118

Paul Kersey

PhD University of Edinburgh 1995.
At EMBL since 1999. Team leader since 2008.

Ensembl Genomes

Previous and current research

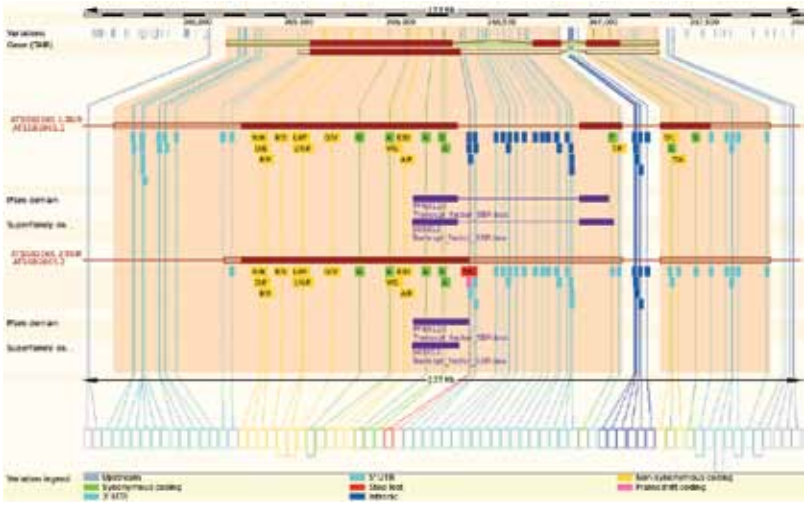
The Ensembl Genomes team is responsible for providing services based on the genomes of non-vertebrate species. The falling costs of DNA sequencing (for deciphering unknown sequences and assays of known sequences) have led to an explosion of reference genome sequences and genome-wide measurements and interpretation. Ensembl Genomes provides five portals (for bacteria, protists, fungi, plants and invertebrate metazoa) offering access to these data through a set of programmatic and interactive interfaces, which were originally developed in the context of the (vertebrate-focused) Ensembl project. Collectively, the two projects span the taxonomic space.

The development of next generation sequencing technologies has led to the performance of complex and highly data-generative experiments, now performed even in species studied only by small communities with little informatics infrastructure. Through collaborating with the EBI and re-using our established toolset, such small communities can store, analyse and disseminate data more cheaply and powerfully. Our leading collaborators include VectorBase (Lawson et al. 2009), a resource focused on the annotation of invertebrate vectors (the EBI is a direct participant); WormBase; and Gramene. Our major areas of interest include broad-range comparative genomics and the visualisation and interpretation of genomic variation, which is being increasingly studied in species throughout the taxonomy.

Our interest in bacteria has led us to become involved in the development of Microme, a new resource for bacterial metabolic pathways.

Future projects and goals

In the next year, we should further increase the number of genomes included in Ensembl Genomes and plug the remaining gaps in our taxonomic coverage. In particular, we have been awarded funding to establish PomBase, a new resource for the fission yeast genome, and PhytoPath, a resource focused on phytopathogen genomes; we have also received funding to work on the genome and variome of the wheat rust pathogen. In each case, we have partnered with leading research groups who bring their knowledge of the biological domain to our collaborations. The genomes of bacteria are less well served by our current models of data organisation than those of eukaryotes and a major restructuring of our services for these will occur soon; this will result in significantly increased coverage of this kingdom within our resources. The public launch of Microme is also expected before the end of 2011.



Ensembl variation image from the SPB8 gene of Arabidopsis thaliana *(encoding a squamosa promoter binding protein-box domain protein). The data set is constructed from three programs to measure population wide variation, two of which are part of a program to completely sequence the genomes of 1001 individuals of this species (108 of which are currently available); the third is a chip-based assay of 214 000 known single-nucleotide polymorphisms (SNPs) in a further 931 genomes. These efforts have so far identified over 7 million SNPs, over 500 000 insertion/deletion loci and over 200 million individual genotypes. The Ensembl variation infrastructure includes support for visualising variation loci (and their effects on genes) in their genomic contexts (illustrated), individual sequence and links to phenotype. A data-mining tool for variation data, using the BioMart data warehousing tool, is also available.*

Selected references

Kersey, P.J., *et al.* (2010) Ensembl Genomes: extending Ensembl across the taxonomic space. *Nucleic Acids Res.*, 38 (Database issue), D563-D569

Kirkness, E.F., Haas, B.J., Sun, W., *et al.* (2010) Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc. Nat. Acad. Sci. USA*, 107, 12168-12173

Lawson, D., *et al.* (2009) VectorBase: a data resource for invertebrate vector genomics. *Nucleic Acids Res.*, 37 (Database issue), D583-D587

Gerard Kleywegt

PhD University of Utrecht, 1991. Postdoctoral research, University of Uppsala.

Coordinator, then programme director of the Swedish Structural Biology Network, 1996-2009.

Professor of Structural Molecular Biology, University of Uppsala, 2009.

At EMBL-EBI since 2009.

The Protein Data Bank in Europe (PDBe)

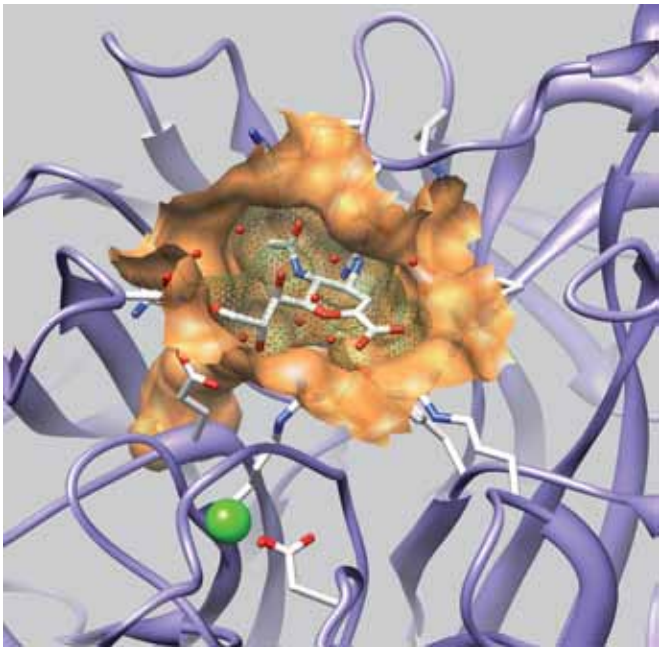
Previous and current research

The Protein Data Bank in Europe is one of the six core molecular databases hosted by the EMBL-EBI. PDBe is also the European partner in the Worldwide Protein Data Bank organisation (wwPDB), which maintains the single international archive for biomacromolecular structure data. The other wwPDB partners are the RCSB and BMRB in the United States and PDBj in Japan. PDBe is a deposition and annotation site for the two major databases containing biomacromolecular structure data: the Protein Data Bank (PDB) and the Electron Microscopy Data Bank (EMDB). Whereas the PDB is maintained by the wwPDB partners, EMDB is a joint venture between PDBe, RCSB and Baylor College of Medicine. The work of PDBe, wwPDB and EMDB is guided by scientific advisory committees (one for each organisation) that meet annually.

The major goal of PDBe is to provide integrated resources of structural data that evolve with the needs of biologists. To that end, PDBe endeavours to: handle deposition and annotation of structural data expertly as a wwPDB and EMDB deposition site; provide an integrated resource of high-quality macromolecular structures and related data; and maintain in-house expertise in all the major structure-determination techniques (i.e. X-ray crystallography, nuclear magnetic resonance spectroscopy, electron microscopy).

Future projects and goals

We have several ambitious goals for the coming years: improving awareness of PDBe in the biomedical community; becoming the logical first stop on any quest for structural information; and transforming the structural archive into a truly useful resource for biomedical and related disciplines. We will focus on the development of new services, tools and resources in our five strongest or most promising areas: refinement of our advanced services such as PDBePISA, PDBeFold, PDBeMotif and the new browsers; annotation, validation and visualisation of ligand data; integration with other resources; validation and presentation of information about the quality and reliability of structural data; and exposing experimental data in ways that help experts and non-experts alike understand the extent to which it supports the structural models produced.



Details of the drug-binding site in the X-ray crystal structure of the complex of the 1918 influenza virus H1N1 neuraminidase and Zanamivir (PDB entry 3B7E). The development of Zanamivir was aided substantially by knowledge of the 3D structure of its target, neuraminidase. The compound is currently marketed by GlaxoSmithKline under the trade name Relenza.

Selected references

Velankar, S., Best, C., Beuth, B., *et al.* (2010) PDBe: Protein Data Bank in Europe. *Nucleic Acids Res.,* 38, D308-D317

Velankar, S., Alhroub, Y., Alili, A., *et al.* (2010) PDBe: Protein Data Bank in Europe. *Nucleic Acids Res.,* 39 (Suppl. 1), D402-D410

Berman, H.M., *et al.* (2010) Safeguarding the integrity of protein archive. *Nature,* 462, 425

Jane Lomax

PhD in parasite population genetics, University of Cambridge, 2002.

At EMBL-EBI since 2002.

The Gene Ontology editorial office

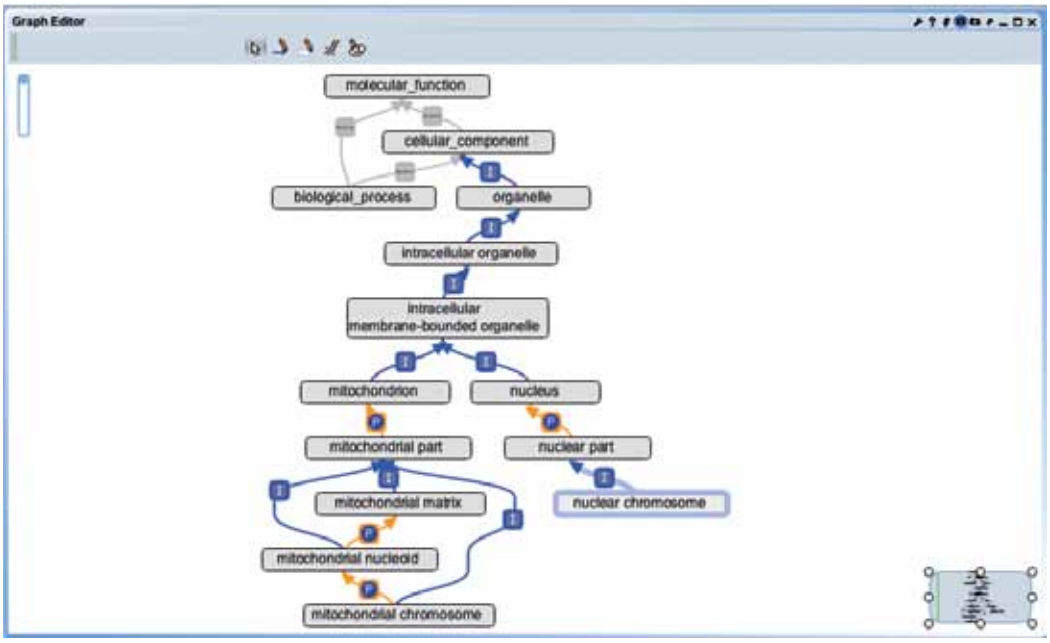
Previous and current research

The Gene Ontology (GO) is a major bioinformatics initiative to unify the representation of gene and gene-product attributes across all species. The aims of the Gene Ontology project are threefold: to maintain and further develop its ontologies of gene and gene product attributes; to annotate genes and gene products, and assimilate and disseminate annotation data; and to provide tools to facilitate access to all aspects of the data provided by the Gene Ontology project. The GO ontologies cover three key biological domains that are shared by all species: the cellular component (the parts of a cell or its extracellular environment); molecular function (the elemental activities of a gene product at the molecular level, e.g. binding or catalysis); and biological process (operations or sets of molecular events with a defined beginning and end, pertinent to the functioning of integrated living units, e.g. cells, tissues, organs, and organisms).

Groups participating in the GO Consortium include major model organism databases and other bioinformatics resource centres. At EMBL-EBI, the GO Editorial Office plays a key role in managing the distributed task of developing and maintaining the GO vocabularies. We contribute to a number of other GO project efforts, including web presence, software testing, user support and education.

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 2010 on signalling, kidney development and other topics will therefore continue, and we intend to start developing terms in the area of neurobiology. We will continue adding further sets of cross-products to GO, allowing us to improve TermGenie so that more routine term addition can be done automatically. This will free up editing time for more complex, biologically detailed work. These cross-product sets will include links to ChEBI, the first set of external cross-products to be added to GO. We also hope to start making cross-products to the Cell Ontology on 2011. Additional links between the biological process and molecular function ontologies will be created using new process-specific function terms.



Structure of the Gene Ontology, shown in OBO Edit.

Selected references

Mungall, C.J., *et al*. (2010) Cross-product extensions of the Gene Ontology. *J. Biomed. Inform.*, (in press). Published online 10 February

Smith, B., *et al*. (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.*, 25, 1251-1255

Johanna McEntyre

PhD in Plant Biotechnology,
Manchester Metropolitan University, 1990.

Editor, Trends in Biochemical Sciences, Elsevier,
Cambridge, UK, 1990-1997.

Staff scientist, NCBI, National Library of Medicine,
NIH, USA, 1997-2009.

At EMBL-EBI since May 2009.

Literature resources

Previous and current research

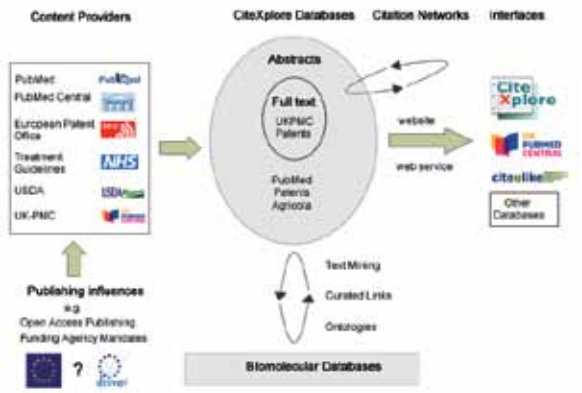
The scientific literature is a key component of the biomedical research life cycle. Providing new ways to access the literature, within the context of other biomedical data resources, will be essential as the scientific community endeavours to organise and make the best use of the flood of data promised by emerging sequencing technologies. The literature holds great promise as a force for integrating information, as it contains the formal record of the community's collective understanding of the biomedical and related sciences. Meaningful links between data resources and the literature will equip researchers better for data analysis, navigation and discovery. With several thousand new research articles published every day, linking articles to each other – and to the broader scientific literature such as textbooks, theses and patents – will become a necessity if we are to leverage the investment in scientific research to greater potential.

One approach to deeper integration is to identify terms of interest within research articles and use these to link similar papers and related data resources. Text mining is a high-throughput approach to identifying biological terms in large volumes of text data, providing a basis for the development of new search and browse applications. The continued refinement of text-mining techniques, along with the growing portion of articles that are published as open access, will stimulate precise, deep linking in the future.

The goal of the literature services at EMBL-EBI is to build text-based resources for the life sciences, integrated with other public-domain data resources hosted at EMBL-EBI. To this end, we run the citations database CiteXplore, which contains around 25 million biomedical abstracts from sources such as PubMed, Agricola, and Patents. The database is updated daily and links to a number of EBI data resources. We also calculate citation information for the records we hold: about 10 million of these articles have been cited at least once; as such, ours represents one of the largest public-domain citation networks in the world. Considerable recent effort has been focussed on building UK PubMed Central, a full text database of life science research articles, in collaboration with the British Library, University of Manchester, and the NCBI. This website was released into beta in January 2010, and has been improved significantly since then.

Future projects and goals

Our group will build on the current programme of work, evolving UKPMC into a European-based resource that represents European science, promoting open-access publishing and providing an alternative to PubMed Central (USA). Key to realising this vision is the development of a fast and reliable search, and the integration of the research articles with related data resources used in scientific and clinical workflows. Furthermore, engagement with the European scientific community, both directly and via existing publishing mechanisms, will help us to move towards building a public-domain content network across Europe.



Overview of the activities of the Literature Services group.

Maria J. Martin

BSc in Veterinary Medicine, Universidad Complutense, Madrid, 1990.
PhD in Molecular Biology (Bioinformatics), Universidad Autonoma, Madrid, 2003.
At EMBL-EBI since 1996.
Technical team leader since 2009.

UniProt development

Previous and current research

The Universal Protein Resource (UniProt) Development Team provides the bioinformatics infrastructure for this resource. It is also responsible for maintaining and developing tools for the UniProt Curation Team. UniProt provides the scientific community with a central repository of protein sequences with comprehensive coverage and a systematic approach to protein annotation. It comprises four focused database layers, which incorporate, interpret, integrate and standardise data from large and diverse sources; this makes it the most comprehensive catalogue of protein sequence and functional annotation.

UniProt database layers

- UniProt Knowledgebase (UniProtKB) provides the central database of protein sequences with accurate, consistent and rich sequence and functional annotation;
- UniProt Metagenomic and Environmental Sequences (UniMES) database is a repository specifically developed for the newly expanding area of metagenomic and environmental data;
- UniProt Archive (UniParc) provides stable, comprehensive, non-redundant sequence collection by storing the complete body of publicly available protein sequence data;
- UniProt Reference Clusters (UniRef) provide non-redundant data collection based on the UniProt Knowledgebase and UniParc in order to obtain complete coverage of sequence space at several resolutions.

Future projects and goals

We will finish integrating the automatic annotation systems developed by the three consortium members under a single, unified database, rule-annotation tool and pipeline infrastructure. Once this has been established we will explore data exchange mechanisms to provide the annotation communities with both annotation rules and the means to annotate in our system. We will develop automatic systems to organise and visualise complete proteomes, which will allow users to have a global genome/proteome and gene-product-centric view of the sequence space from which they can drill down to the variations and annotations specific to each protein.

The availability of large functional-genomics and proteomics datasets requires improved data integration approaches. The team will cooperate with diverse data providers to develop and assess new protocols for the exchange and integration of information in UniProt. We will also explore novel approaches to increasing community participation, for example by providing easy-to-use mechanisms for making data contributions.

Selected references

The UniProt Consortium (2010). The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.*, 38, D142-D148

Jain, E., *et al*. (2009) Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics*, 10, 136

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

Cochrane, G., Martin, M.J. and Apweiler, R. (2010) Public data resources as a foundation for a worldwide metagenomics data infrastructure. In: *Metagenomics: theory, methods, and applications*, Marco, D., Editor. Norwich, UK: Caister Academic Press, 212 pp

Claire O'Donovan

BSc (Hons) in Biochemistry, 1992, University College Cork, Ireland.

Diploma in Computer Science, 1993, University College Cork, Ireland.

At EMBL since 1993. At EMBL-EBI since 1994. Technical team leader since 2009.

UniProt content

Previous and current research

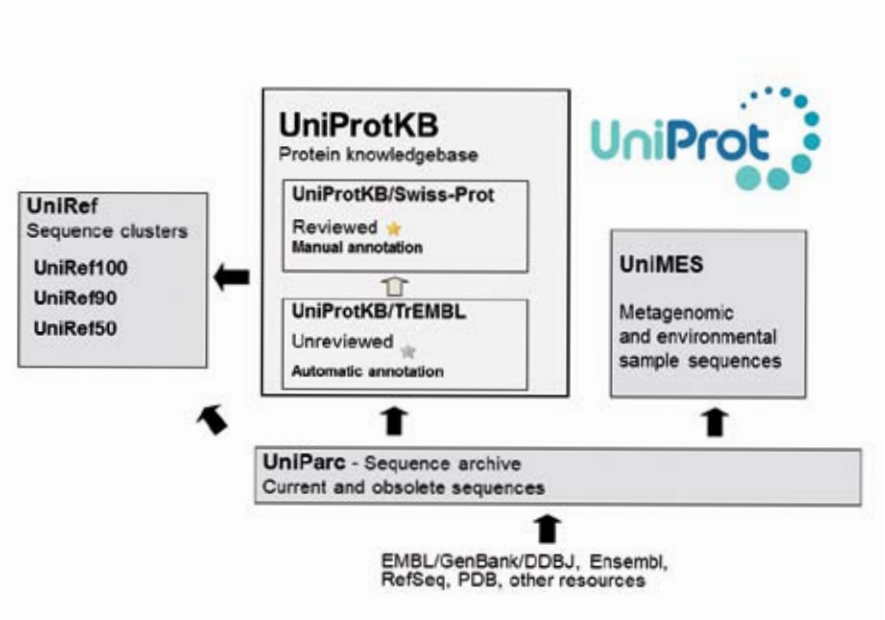
The Universal Protein Resource (UniProt) is a collaboration of the EMBL-EBI, the Swiss Institute of Bioinformatics (SIB) and the Protein Information Resource group at Georgetown University Medical Center and the University of Delaware. Its purpose is to provide the scientific community with a single, centralised, authoritative resource for protein sequences and functional annotation. The primary mission of the consortium is to support biological research by maintaining a freely accessible, high-quality database that serves as a stable, comprehensive, fully classified, richly and accurately annotated protein sequence knowledge base with extensive cross-references and querying interfaces. The UniProt databases consist of four database layers optimised for different purposes (see page 83).

Gene Ontology (GO) is a well-established, structured vocabulary that has been successfully used in gene product functional annotation. The UniProt–Gene Ontology Annotation (UniProtKB–GOA) database was created at the EMBL-EBI in 2001. The aim of the UniProtKB-GOA project is to provide high-quality manual and electronic annotations to the proteins stored in UniProtKB using GO vocabulary.

Future projects and goals

Our main goal is to test 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.



UniProt database organisation.

Selected references

The UniProt Consortium. (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.*, 38, D142-D148

Binns, D., *et al.* (2009) QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics, 25,* 3045-3046

Alam-Faruque, Y., *et al.* (2010) The Renal Gene Ontology Annotation Initiative. *Organogenesis, 6,* 71-75

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

John Overington

PhD in Crystallography, Birkbeck College, London, 1991.
Postdoctoral research, ICRF, 1990-1992.
Pfizer 1992-2000. Inpharmatica 2000-2008.
At EMBL-EBI since 2008.

ChEMBL: a database of bioactive drug-like small molecules

Previous and current research

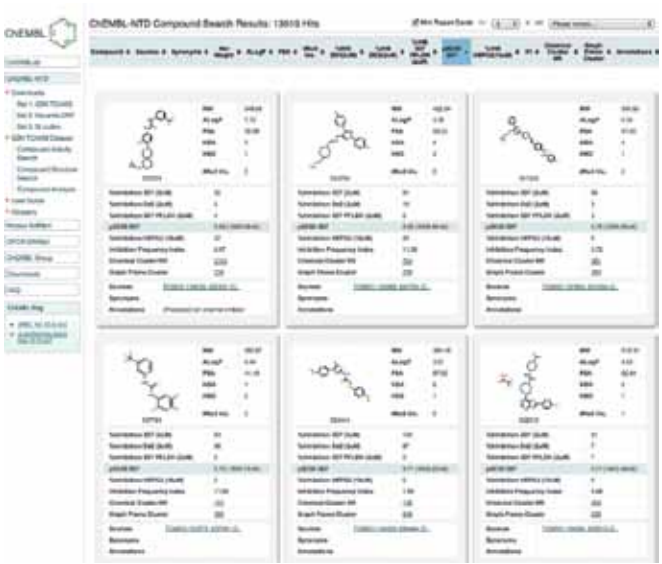
The ChEMBL group develops and manages the EBI's database of bioactive, drug-like small molecules, which contains two-dimensional structures, calculated properties and abstracted bioactivities such as binding constants, pharmacology and ADMET data. ChEMBL data are abstracted and curated from the primary scientific literature, and cover a significant fraction of the structure–activity relationship and discovery of modern drugs. 2010 marked the first full year of staffing for the group, and has seen a number of milestones for the ChEMBL resource. Its first public release – in January 2010 – achieved broad coverage in the press and was extremely well received by the scientific community. The data is widely accessed via the web interface and via download of the entire database for local searching, and advanced tools developed by the ChEMBL team for interactive filtering and data selection provide added value to users.

We have two active research areas. The first is the building of a computational system to analyse functional and binding data for peptides, and then to propose their optimisation in order to improve pharmaceutical properties, stability, affinity and selectivity. We published a paper on the analysis of ligand efficiency measures for the content of ChEMBL as well as a series of similarity maps for natural and unnatural amino acids. This work is funded under the EIPOD scheme, with the designed peptides planned for synthesis and bioassay in the lab of Maja Köhn (page 30).

The second area of research is a comprehensive analysis of 'tool compounds' or 'chemical probes'. We have assembled a number of sets of compounds that are generally considered to be chemical tools, that is, small molecules that are used to probe the function of specific proteins in either a cell or an in vivo model system. These compounds have been characterised for various properties (e.g. affinity, molecular size); approaches to predict the affinity variances across model organism species have been developed (i.e. across rat, mouse, and human orthologues).

Future projects and goals

We will release the drugability prioritisation and analysis tools, and also populate the database with biotherapeutic and clinical candidate development data. Also of high priority will be completing integration with core EMBL-EBI resources such as Ensembl, UniProt, PDBe and ArrayExpress.



Representative screen-shot of the ChEMBL database showing flexible querying and powerful anlaysis routines for bioactivity data.

Selected references

Abad-Zapatero, C., *et al*. (2010) Ligand efficiency indices for an effective mapping of chemico-biological space: the concept of an atlas-like representation. *Drug Discov. Today,* 15, 804-811

Gaulton, A. and Overington, J.P. (2010) Role of open chemical data in aiding drug discovery and design. *Future Med. Chem., 2*, 903-907

Helen Parkinson

PhD in Genetics, 1997.

Research associate in Genetics, University of Leicester, 1997-2000.

At EMBL since 2000.

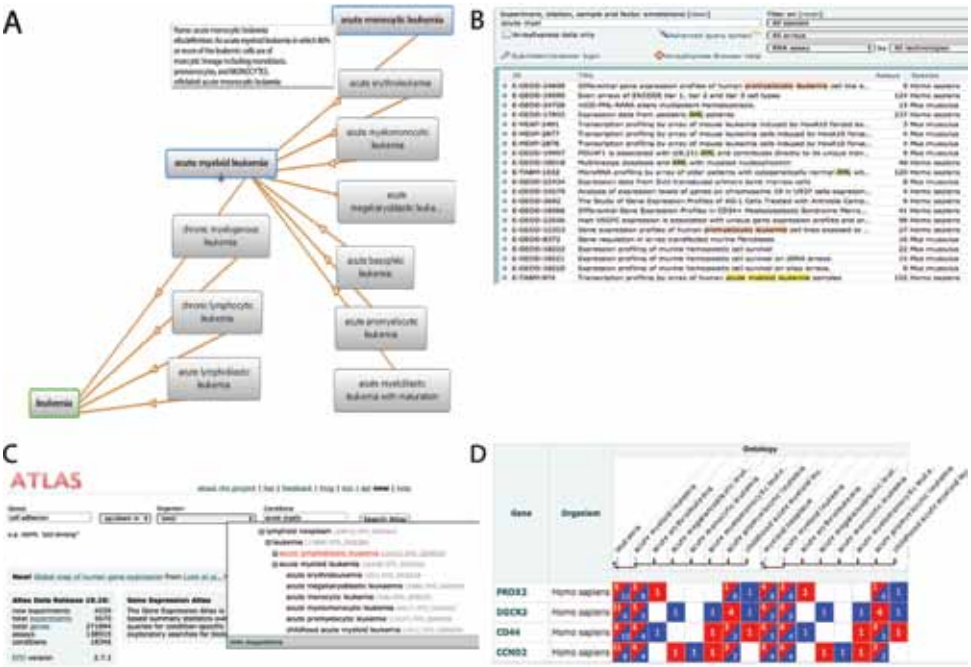
Functional genomics production

Previous and current research

The Functional Genomics Production Team manages data content and user interaction for the core EBI databases: the ArrayExpress Archive, the Gene Expression Atlas and the new Biosamples Database. All three resources have complex metadata representing experimental types, variables and sample attributes for which we require semantic markup in the form of ontologies. We develop ontologies and software for the annotation of complex biological data, including the Experimental Factor Ontology (EFO) for functional genomics annotation (Malone, 2010), the Software Ontology, the Ontology for Biomedical Investigation and the Vertebrate Anatomy Ontology (VBO). We collaborate with international partners to develop MAGE-TAB based data management infrastructure and annotation tools for gene expression data. The team has expanded its remit to deal with the change in technology from arrays to RNA sequencing experiments; this has resulted in collaboration with the EBI databases ENA and EGA to provide data flow and integration between these sequence databases and ArrayExpress. The curation, annotation, quality control and ontology development work of the production team supports meta-analyses of gene expression data which provide content for the Gene Expression Atlas and research into global gene expression analysis.

Future projects and goals

We will work to improve the volume and quality of annotation for RNA-Seq data by working with data generating centres such as the Wellcome Trust Sanger Institute to automate RNA-Seq data submissions. EFO will be extended to support annotation of these data, for example for single cell sequencing studies, and also for data integration in the sample database, where we will develop new terms for cell lines and samples used in genome-wide association studies (GWAS) studies. Finally, we are working to use EFO for RDF export of data from the GXA jointly with the Rebholz-Schuhmann group (page 69) with support from the EBI Industry Programme.



EFO is a data-driven application ontology that can be visualised as a node edge diagram showing terms placement and definitions in the BioPortal terminology browser (A), used to query ArrayExpress Archive Data (B) and used for query and visualisation for variables in the Gene Expression Atlas (C-D) in the heatmap view.

Selected references

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

Malone, J., *et al.* (2010) Modeling sample variables with an experimental *factor ontology Bioinformatics, 26,* 1112-1118

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-D872

Shankar, R., *et al.* (2010) Annotare: a tool for annotating high-throughput biomedical investigations and resulting data. *Bioinformatics, 26,* 2470-2471

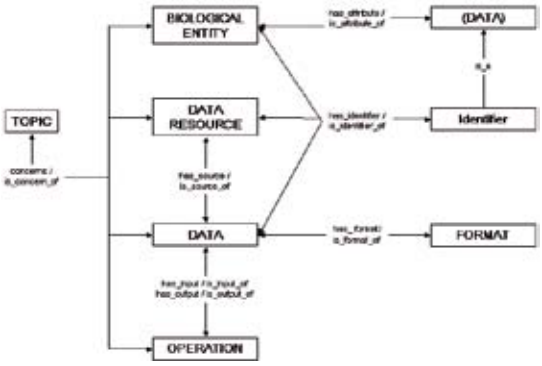
Peter Rice

BSc University of Liverpool, 1976.
EMBL Heidelberg, 1987-1994.
Sanger Centre 1994-2000.
LION Bioscience 2000-2002.
At EMBL-EBI since 2003.

Developing and integrating tools for biologists

Previous and current research

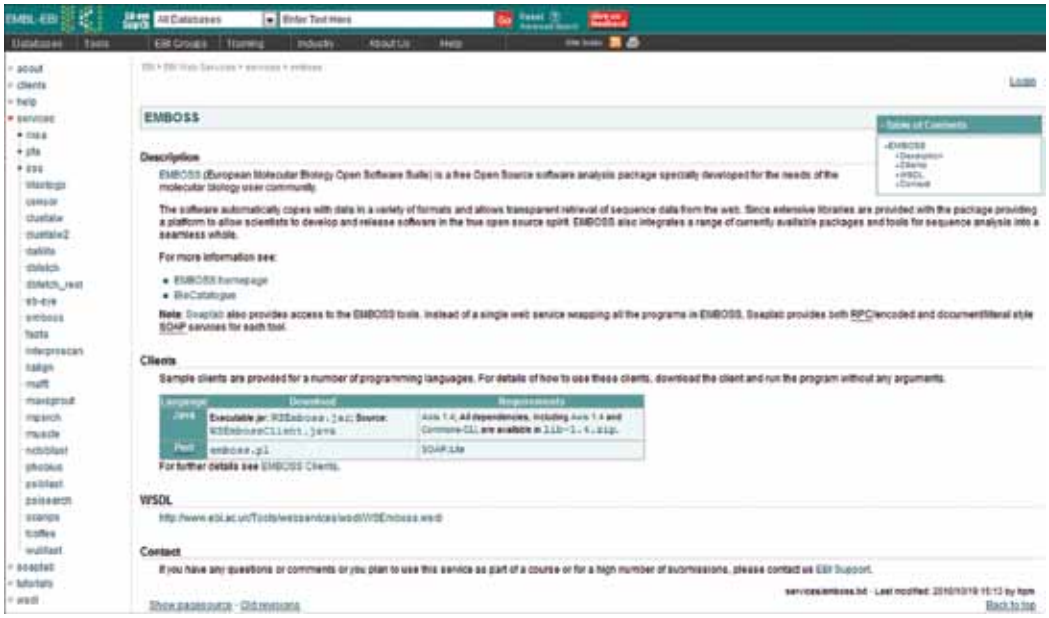
The team focuses on the integration of bioinformatics tools and data resources. We also investigate and advise on the e-Science and Grid technology requirements of the EMBL-EBI through application development, training exercises and participation in international projects and standards development. Our group is responsible for the development of the EMBOSS open-source sequence analysis package and for the EMBRACE project, which integrates access to bioinformatics tools and data content through standard-compliant web services.



EDAM terms and relations

Future projects and goals

The EMBOSS package is developing rapidly to cover new data types, new sources of data and new data access methods. In the next year we will add generic access to all public bioinformatics data resources with database types for feature annotation, bio-ontologies, taxonomic data, data resource descriptions (from our catalogue of public data resources) and general text or URL-based access to other sources of information. The EMBRACE consortium ends in July 2011. Our work on EMBRACE-compliant services will continue, with the further development of the EDAM ontology as part of EMBOSS and the maintenance of fully annotated SOAP services within SoapLab, also through EMBOSS.



Ugis Sarkans

PhD in Computer Science, University of Latvia, 1998.

Postdoctoral research at the University of Wales, Aberystwyth, 2000.

At EMBL-EBI since 2000..

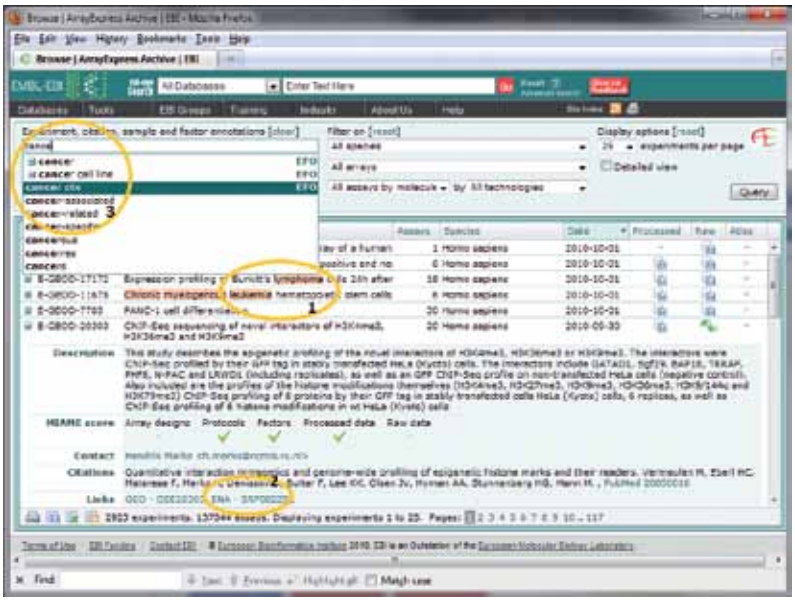
Functional genomics software development

Previous and current research

Our team has been developing software for ArrayExpress since 2001. By October 2010, ArrayExpress held data from almost 500 000 microarray hybridisations and is clearly one of the major data resources of EMBL-EBI. The software development team is building and maintaining several components of the ArrayExpress infrastructure, including data management tools for ArrayExpress Archive (the MIAME-compliant database for the data that support publications); the ArrayExpress Archive user interface; MIAMExpress (a data annotation and submission system); and an array design re-annotation system for aligning user-provided annotation to a uniform reference system. In addition, our team has participated in building the BioSamples database, a new EBI resource, since early 2010.

Future projects and goals

After migration to the new software infrastructure we intend to concentrate on improving our data submission tools, taking into account the growing popularity of sequencing-based functional genomics experiments. The ArrayExpress Archive user interface will undergo further improvements, including simplifying and rationalising information presentation. We plan to bring the BioSamples database to full production status, integrating this valuable new resource with other databases at EMBL-EBI and NCBI.



Some of the new features of the ArrayExpress Archive user interface. Highlight 1, new ontology-aware functionality; Highlight 2, data jointly managed by ArrayExpress and ENA; Highlight 3, search-field autocompletion.

Selected references

Malone, J., *et al.* (2010) Modeling sample variables with an Experimental Factor Ontology. *Bioinformatics, 26*, 1112-1118

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

Christoph Steinbeck

PhD Rheinische Friedrich-Wilhelm-Universität, Bonn, 1995.Postdoc at Tufts University, Boston, 1996-1997.

Group leader, Max Planck Institute of Chemical Ecology, Jena, 1997-2002.

Group leader, Cologne University 2002-2007.

Lecturer in Cheminformatics, University of Tübingen, 2007.

At EMBL-EBI since 2008.

Cheminformatics and metabolism

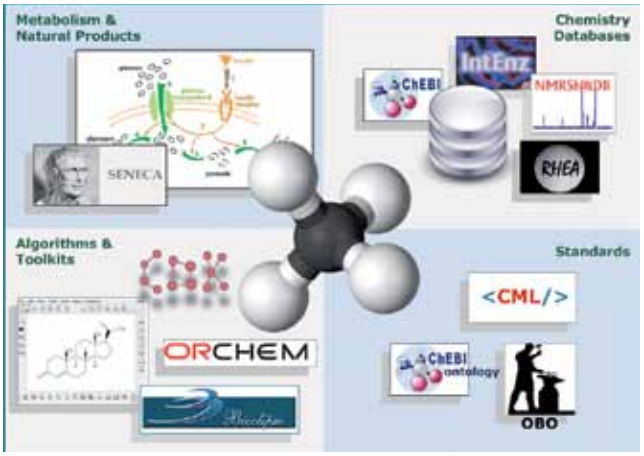
Previous and current research

The Cheminformatics and Metabolism team provides the biomedical community with information on small molecules and their interplay with biological systems. Our database portfolio includes ChEBI, the EBI's database and ontology of chemical entities of biological interest, as well as Rhea and IntEnz, our enzyme-related resources. The group develops methods to decipher, organise and publish the small-molecule metabolic content of organisms. We work on methods to reconstruct genome-scale metabolic models of microbes, develop algorithms to predict metabolomes based on genomic and other information, to determine quickly the structure of metabolites by stochastic screening of large candidate spaces and to enable the identification of molecules with desired properties. This requires algorithms for the prediction of spectroscopic and other physicochemical properties of chemical graphs based on machine learning and other statistical methods.

We are further investigating the extraction of chemical knowledge from the scientific literature by text- and graph-mining methods. This, as well as our work on chemical database technology and curation, is supported by research into chemical ontologies. Together with an international group of collaborators we have developed a number of widely known and used open-source cheminformatics software packages. The Chemistry Development Kit (CDK), which originated in our lab, is the leading open-source Java library for structural cheminformatics. Based on this, we have developed the cheminformatics workflow/pipelining system CDK-Taverna, which allows researchers to build executable data-processing workflows in a LegoTM-like manner, as well as OrChem, our structure-registration and -searching system for the OracleTM-database. In collaboration with partners in Uppsala we initiated Bioclipse, an award-winning, rich client for chemo- and bioinformatics.

Future projects and goals

We have received a grant from the BBSRC to build the missing community resource for metabolomics at the EBI. The database will be cross-species and cross-application and will cover metabolite structures and their reference spectra as well as their biological roles, locations and concentrations. The project will be fully compliant with open standards in metabolomics, including existing minimum reporting standards, or will actively contribute to their creation where these have not been developed. It will further provide the community with a repository for metabolomics experiments reference in scientific publications, matching the functionality of other 'omics' repositories at the EBI, such as the proteomics resource PRIDE. Our team is also leading an effort to integrate all enzyme-related information resources at the EBI into what we call the Enzyme Portal. The Enzyme Portal will provide unified access to resources like IntEnz and Rhea developed in our group as well as to some of the resources of the Thornton group, including their Cofactor Database, EC2PDB, the Catalytic Site Atlas and others.



Scope of work in the cheminformatics and metabolism group.

Selected references

Griffin, J.L. and Steinbeck, C. (2010) So what have data standards ever done for us? The View from Metabolomics. *Genome Med., 2,* 38

Kuhn, T., *et al*. (2010) Cdk-Taverna: an open workflow environment for cheminformatics. *BMC Bioinform.,* 11, 159

Rijnbeek, M., and Steinbeck, C. (2009) Orchem - an open source chemistry search engine for Oracle. *J. Cheminformatics, 1,* 17

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 highly automated ESRF crystallography beamlines are all are equipped with EMBL-designed high-precision microdiffractometers and frozen crystal sample changers. One beamline is run by the outstation and the ESRF in collaboration with India. 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 integrated projects. These include a very successful robotic system for nanovolume crystallisation and a novel, high-throughput screening method, ESPRIT, which enables soluble protein domains to be indentified in otherwise badly expressed or insoluble proteins.

More recently, a Eukaryotic Expression Facility (EEF) has been established specialising in the expression of multi-subunit complexes in insect cells, building on and developing further the well known MultiBac method. 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. A confocal microscope with facilities for cross-correlation spectroscopy is available for the study of complex formation in cells, as well as a new top-end Polara electron microscope with cryo-tomography capability.

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 mechanisms of transcriptional regulation, including at the epigenetic level, is another important topic. Structural analysis of eukaryotic transcription factor 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 understand the biogenesis and role of piRNAs, which are critical for silencing transposons in the germ line.

Another major focus is the study of segmented RNA viruses, particularly influenza and bunyaviruses, with the aim of understanding how they replicate and also as targets 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 and Head of Outstation since 1989.
Director of CNRS-UJF-EMBL International Unit for Virus Host Cell Interactions (UVHCI) since 2007.

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

Previous and current research

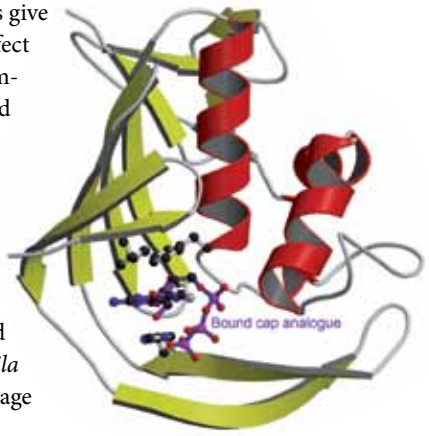
We use X-ray crystallography as a central technique to study the structural biology of protein-RNA complexes involved in RNA metabolism, translation, virus replication and epigenetics. Current major focuses are on negative strand RNA virus polymerases, innate immune system receptors and the dosage compensation complex.

The nuclear cap-binding complex (CBC) binds to m7Gppp cap at the 5' end of Pol II transcripts 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 obtained the first structural information on the interacting domains of these three proteins whose ternary complex formation triggers decay.

Aminoacyl-tRNA synthetases play an essential role in protein synthesis by charging specifically their cognate tRNA(s) with the correct amino acid. We aim to obtain structural information to help understand the catalytic mechanism of the enzymes and their substrate specificity for ATP, cognate amino acid and tRNA. Most recently we solved the structures of a class I enzyme, leucyl-tRNA synthetase, and a class II enzyme prolyl-tRNA synthetase, each with their cognate tRNAs bound. Both these enzymes contain a large inserted editing domain able to recognise and hydrolyse mischarged amino acids. This proof-reading activity is essential for maintaining translational fidelity. We have collaborated in the elucidation of the mechnisms of a new anti-fungal compound that targets the editing site of leucyl-tRNA synthetase and have now extended this using structure-based approaches to design new anti-bacterials that are active against multi-drug resistant strains.

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 published the structure of the UPF1-UPF2 complex and are now determining the architecture of the entire trimeric UPF1-UPF2-UPF3 complex. A major ongoing project is structure determination of the trimeric influenza virus RNA-dependent polymerase, the viral replication machine. We have determined the structure of four distinct domains from the polymerase, including the two key domains 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). This work is now being extended to the polymerases of other segmented RNA viruses such as bunyaviruses which also perform cap-snatching. In collaboration with the Ellenberg (page 7) and Briggs (page 40) groups we are now engaged in confocal and cross-correlation fluorescence studies as well as correlative EM microscopy of the transport and assembly of the influenza polymerase and RNPs in living, infected cells. We also work on the structure and mechanism of activation of Rig-I, an intracellular pattern recognition receptor of the innate immune system which signals interferon production upon detection of viral RNA. Finally we have two new projects concerning ncRNAs: the structure and mechanism of the X-chromosome dosage compensation (MSL complex), which in *Drosophila* contains an essential non-coding RNA and secondly, in collaboration with the Pillai group (page 98), structural studies of proteins involved in the piRNA pathway.



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., Kulozik, A., Kadlec, J., Sattler, M. & Cusack, S. (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, 914-918

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

Rock, F.L., Mao, W., Yaremchuk, A., *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.

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

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.

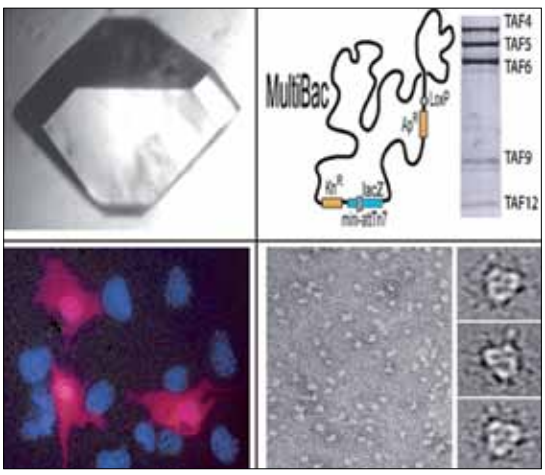
Our major methodologies include molecular biology and X-ray crystallography. However, we readily apply techniques from other fields that are required for our research, both by ourselves and through collaborative efforts. A prerequisite for successful structural study of many complexes, both by electron 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. Furthermore, after expression and biochemical analysis do not provide the flexibility needed 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 multi-component membrane protein complexes and a 1 MDa core assembly of human TFIID general transcription factor.

Future projects and goals

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 99) 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). We are also 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.



Florent Cipriani

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

Diffraction instrumentation team

Previous and current research

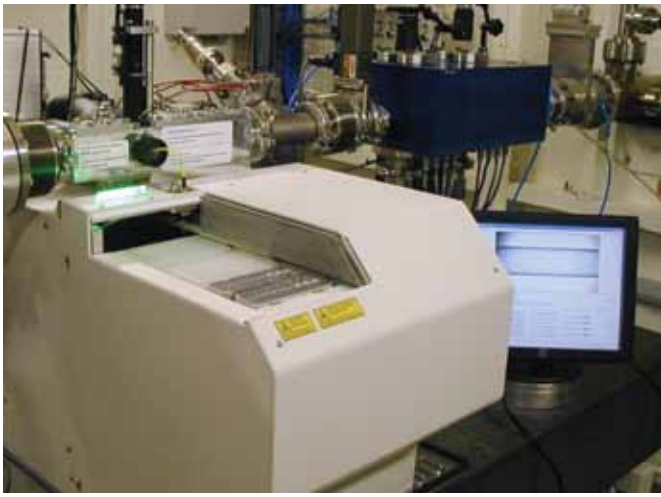
Our team develops instruments and methods for X-ray scattering experiments in collaboration with the synchrotron crystallography team (page 96) and the ESRF structural biology group, as well as contributing to the development of the EMBL@PETRA3 beamlines (see page 100). Part of our work is also to support and improve equipment at the ESRF beamlines. Most of our technologies are transferred to industry and available to other synchrotrons worldwide via industrial companies.

ESRF's macromolecular crystallography (MX) beamlines still rely on our MD2/MD2M diffractometers, SC3 sample changers and C3D crystal centring software to process several hundred crystals per day. In Hamburg, the EMBL@PETRA3 beamline MX1 will soon be equipped with a MD2 diffractometer. To fully benefit from the micron-sized beam of MX2, we are currently developing a vertical Kappa diffractometer (MD3) based on a prototype spindle tested on the ESRF ID23-2 micro-focus beamline. The sub-micron stability obtained at crystal position, showed that it will be possible to re-orient and process crystals down to 2-3 micrometers in size.

The diffraction quality of large unit cell crystals, such as macromolecular complexes, is often a limiting factor to structure determination. The newly developed HC1 crystal dehydration device is now routinely used to help enhance crystal quality at beamlines by reducing the amount of solvent contained in the crystals matrix. It is also useful for testing crystals at room temperature. To satisfy growing user demand, a second HC1 machine has been put into operation. Our aim is to integrate dehydration control in MxCube and eDNA. A dehydration experiment database and associated tools have been developed to collect statistics on the results obtained from dehydration protocols in relation with crystallisation parameters. Enriched by HC1 users at the ESRF, Max-LAB and Diamond, this tool should help on new dehydration experiments in predicting initial dehydration levels and in proposing protocols successfully used with similar crystal parameters.

The automated BioSAXS sample environment project developed in collaboration with EMBL Hamburg and the ESRF is almost complete. At the ESRF ID14-3 beamline, several hundred samples and buffers stored in SBS plates can be exposed automatically to X-rays in a glass capillary at controlled temperature (see figure). Sample volumes down to 5 µl of solution can be transferred reliably to the exposure cell. In-line sample concentration measurement and pipetting functions are available for *in situ* sample dilution and additions. A second system will be installed early next year at EMBL@PETRA3. In collaboration with the HTX team (see page 95), we have developed the proof-of-concept on a technology to automatically harvest crystals grown in specifically designed crystallisation supports, on which we have filed a patent. The affordable number of samples processed at automated MX beamlines is partly limited by the design of the Spine sample holders. To overcome this problem, we have started studies for a compact and precise sample holder that should allow high-density crystal storage and reduced crystal alignment time. On this basis, a future sample holder standard should be developed in collaboration with European synchrotrons.

BioSAXS sample changer at the ESRF/EMBL beamline ID14-3.



Future projects and goals

Next year, EMBL@PETRA3 should receive a BioSAXS automated sample environment, and we expect the MD3 diffractometer introduced above to be operational at MX2. In the context of the ESRF MASSIF upgrade program, our main focus will be on the development of an automated crystal harvester and new sample holder standard, with the ambition to build a pilot automated integrates MX screening unit jointly with the HTX and synchrotron crystallography teams.

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, 1237-1246

Cipriani, F., Felisaz, F., Launer, L., Aksoy, J.S., Caserotto, H., Cusack, S., *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.

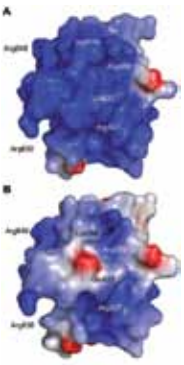
Influenza RNA polymerase: There is worldwide concern that currently circulating avian influenza viruses will cross the species barrier and become highly pathogenic, human transmissible strains with pandemic potential. This could result from residue changes in several influenza proteins, either by point 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.

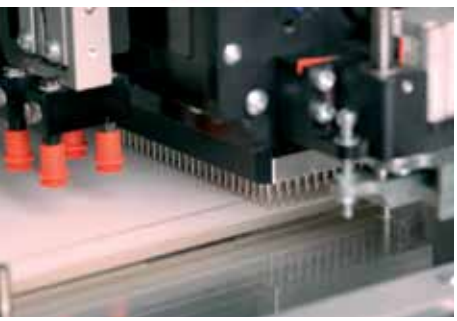
Human kinases: Cells have intricate mechanisms of sensing and responding to environmental changes. Upon a stimulus detected by a cellular receptor the complex system of signal transduction is activated that results in changes in gene expression. Protein kinases play a crucial role in cellular stress 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.





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.

Left: 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.

Selected references

Boivin, S., Cusack, S., Ruigrok, R.W. & Hart, D.J. (2010). Influenza Virus Polymerase: Structural insights into replication and host adaptation mechanisms. *J. Biol. Chem.*, 285(37), 28411-7

Yumerefendi, H., Tarendeau, F., Mas, P.J. & Hart, D.J. (2010). ESPRIT: An automated, library-based method for mapping and soluble expression of protein domains from challenging targets. *J. Struct. Biol.*, 172(1), 66-74

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., *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

José A. Márquez

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

The high-throughput crystallisation lab

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, which processes large numbers of samples using very low sample volumes. Since its opening, we have offered services to hundreds of scientists performing several million 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 P-CUBE.

One of the major problems we face is accurately capturing the enormous flux of information generated by experimental parameters and results. For this purpose, 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 crystallisation facilities in Europe. The HTX lab not only increases the efficiency in the process of structure solution, but also represents a technological advantage critical for the success of challenging projects, such as those studying protein complexes and large macromolecular assemblies.

Molecular mechanisms in sensing and signalling: Our research focus is on understanding the mechanisms of sensing and signalling at a structural level. Recently, we have obtained the structure of the receptor for abscisic acid (ABA), 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 in plants. 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.

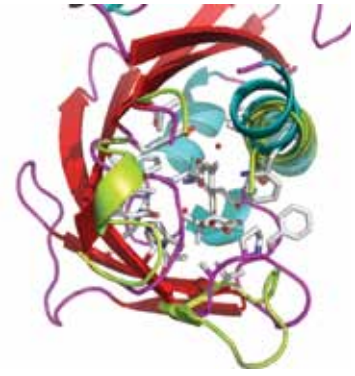
Future projects and goals

Developing new technologies in crystallisation: In collaboration with the diffraction instrumentation team (page 93) and based on solutions we have already explored and patented, we plan to develop a novel prototype system for automated crystal harvesting. This system is designed to fill the automation gap between crystallisation and data collection and will allow the development of integrated crystal screening stations.

In the laboratory, we will continue the study of signalling systems with a special focus on two major fields: TEC-family protein tyrosine kinases and the ABA signalling 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 synthesize 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.

High-throughput crystallisation robot at the HTX Lab.

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



Selected references

Santiago, J., Dupeux, F., Round, A., Antoni, R., Park, S.Y., Jamin, M., Cutler, S.R., Rodriguez, P.L. & Marquez, J.A. (2009). The abscisic acid receptor PYR1 in complex with abscisic acid. *Nature*, 462, 665-668

Santiago, J., Rodrigues, A., Saez, A., *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 of the ESRF in the design, construction and operation of synchrotron radiation beamlines. We are currently responsible for two macromolecular beamlines, ID14-4 (figure 1) and ID23-2, as well as the BioSAXS beamline at ID14-3. The team also manages the operation of BM14, which is run as a partnership with the ESRF and Indian government. BM14 is now welcoming users from India as well as EMBL member states. All the structural biology beamlines at the ESRF are performing well and a particular highlight is the continued use of the beamlines 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'. We also work in close collaboration with the diffraction instrumentation team (page 93) to develop hardware, software and novel methodologies for sample handling and data collection possibilities. Recent examples include the high throughput BioSAXS liquid handling robot recently installed on ID14-3 and software for optimised MX data collection using crystal reorientation strategies.

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

Future projects and goals

This year we will commence work on the UPBL10 project with our ESRF colleagues. This ambitious project is part of the ESRF upgrade program and will ensure that European users will have continued access to state-of-the-art structural biology beamlines for the next 10 years. This project will commence with the construction of a new bioSAXS beamline on BM29, which should welcome users later this year. A major upgrade of the X-ray optics for BM14 will also be completed this year, enabling a smaller and brighter X-ray beam for improved performance. The 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 continue 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 diffraction instrumentation team, the ESRF, and EMBL-Hamburg, will continue to be actively involved in the provision of a highly automated BioSAXS beamline. We hope that all our combined efforts will push the boundaries of structural biology to better understand the biological functions of more complex biological systems.

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

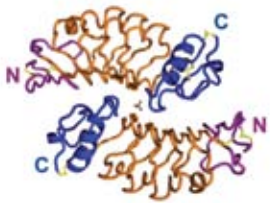


Figure 1 (top): View of the highly automated ID14-4 experimental hutch.
Figure 2 (bottom): Structure of the Slit2 dimerisation domain. Slit2 D4 N- and C-terminal caps are in purple and blue, respectively, the LRRs 1-5 are in orange and the disulphide bridges are in yellow.

Selected references

Gabadinho, J., Beteva, A., Guijarro, M., *et al*. (2010). MxCuBE: a synchrotron beamline control environment customized for macromolecular crystallography experiments. *J. Synchrotron Radiat.*, 17, 700-707

Hinz, A., Miguet, N., Natrajan, G., *et al*. (2010). Structural basis of HIV-1 tethering to membranes by the BST-2/tetherin ectodomain. *Cell Host Microbe*, 7, 314-323

Srivastava, A.,Gangnard, S., Round, A., Dechavanne, *et al*. (2010). Full-length extracellular region of the var2CSA variant of PfEMP1 is required for specific, high-affinity binding to CSA. *Proc. Natl. Acad. Sci. USA*, 107, 4884-4889

P. Pernot, A. Round, P. Theveneau, *et al*. (2010). New beamline dedicated to solution scattering from biological macromolecules at the ESRF. *J. Phys.: Conf. Ser*. 247

Daniel Panne

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

Integrating signals through complex assembly

Previous and current research

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

The systems we have been studying are involved in transcriptional regulation. Transcriptional regulation is mediated by transcription factors which bind to their cognate sites on DNA, and through their interaction with the general transcriptional machinery, and/or through modification of chromatin structure, activate or repress the expression of a nearby gene. The so-called 'cis-regulatory code', the array of transcription factor binding sites, is thought to allow read-out and signal processing of cellular signal transduction cascades. Transcriptional networks are central regulatory systems within cells and in establishing and maintaining specific patterns of gene expression. One of the best-characterised systems is 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 particularly 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.

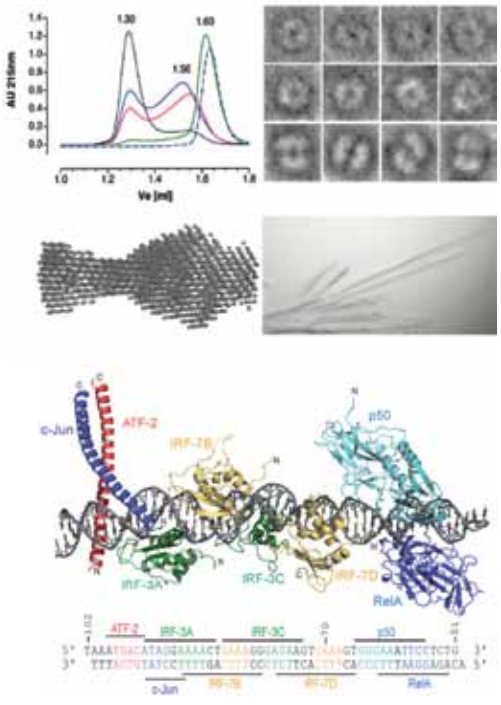


Figure 1 (top): We employ a number of different resolution techniques to visualise cellular structures. Figure 2 (bottom): Atomic model of the INF-β enhanceosome.

Selected references

Reynoird, N., Schwartz, B.E., Delvecchio, M., Sadoul, K., Meyers, D., Mukherjee, C., Caron, C., Kimura, H., Rousseaux, S., Cole, P.A., Panne, D., French, C.A. & Khochbin, S. (2010). Oncogenesis by sequestration of CBP/p300 in transcriptionally inactive hyperacetylated chromatin domains. *EMBO J.*, 29, 2943-2952

Shen A., Higgins D.E., and Panne D. (2009) (2009). Structure of the MogR repressor bound to its A-tract target operator. *Structure*, 13, 769-77

Panne, D. (2008). The enhanceosome. *Curr. Opin. Struct. Biol.*, 18, 236-242

Panne, D., Maniatis, T. & Harrison, S.C. (2007) An atomic model of the interferon-beta enhanceosome. *Cell*, 129(6), 1111-23

PhD 2002, University of Bern.

Postdoctoral research at the Friedrich Miescher Institute, Basel.

Group leader at EMBL Grenoble since 2006.

Ramesh Pillai

Regulation of gene expression by non-coding RNAs

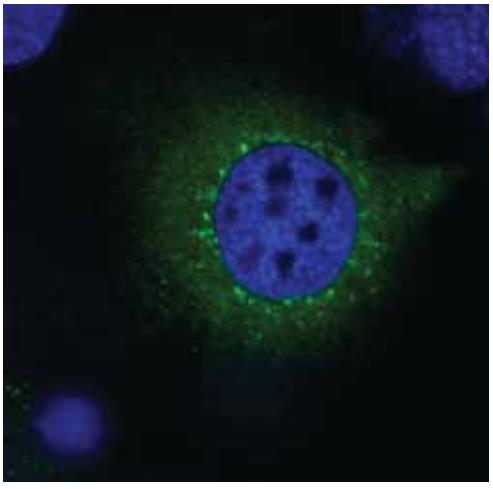
Previous and current research

Past invasion events from mobile genetic elements have left eukaryotic genomes littered with repeats and other transposon sequences. Much of these are inactive fossils, but some still retain the potential to get activated and cause genome instability. Protection from transposons is achieved by silencing them in the germline, which is then maintained throughout the life of the individual. Animal germ cells express a specialised class of ~30 nt small non-coding RNAs called piwi-interacting RNAs (piRNAs) which are implicated in guiding this silencing. Indeed, one universal feature of piRNAs in all animals is their origin from transposon-rich genomic regions. In mammals, they are believed to recruit DNA methyltransferases to transposon sequences. In *Drosophila*, maternally produced piRNAs are deposited in the egg and they contribute to protection from new transposons brought in by the paternal genome. Thus, piRNAs constitute an epigenetic component of the genome defense mechanism in animals.

Our lab is interested in understanding the molecular mechanisms involved in piRNA biogenesis and function. A striking feature of piRNAs is their clustered genomic origins. It is believed that a long single-stranded transcript arising from a cluster is processed into thousands of piRNAs. The mechanism of this processing and the identity of factors involved are unknown. We have taken a biochemical approach to identify these factors by isolating mouse Piwi-associated proteins. This led to the identification of Tudor domain-containing protein 1 (Tdrd1), which interacts by recognising symmetrical dimethyl arginine modification marks on Piwi proteins. Another factor is the putative helicase Mov10l, which is an essential piRNA biogenesis factor, as piRNAs fail to accumulate in mutant mice. In all these studies, we have used a variety of techniques ranging from protein biochemistry, cellular imaging, small RNA bioinformatics and mouse mutants. We are now setting up insect cell culture lines which have an active piRNA pathway, paving the way for potential mechanistic insight into the function of the identified factors. To deepen our understanding, we collaborate with structural biologists to obtain atomic resolution images of the identified pathway components. Recently, this effort resulted in a structure describing the recognition of the 2'-O-methyl mark on piRNAs by the PAZ domain of a Piwi protein.

Future projects and goals

We will continue to analyse additional factors identified in our complex purifications. Another goal would be to understand the features that define genomic regions as piRNA clusters, and whether there is a link between transcription from the clusters and piRNA biogenesis. We also hope to use live cell imaging techniques to study assembly of small RNPs *in vivo* and define the contribution of the individual constituents of the complex to this process. It is our desire to intensify the collaborative work on structural biology of Piwi complexes, adding another dimension to our understanding of germline small RNAs. In addition to small RNAs, our cells express longer non-coding RNAs (ncRNAs) which are implicated in a variety of gene regulatory functions, usually in epigenetic roles. We wish to apply biochemical methods to identify protein components of long ncRNPs to understand their contribution to the molecular function of the RNA.



Localisation of a tagged insect Piwi protein to perinuclear cytoplasmic granules in insect cell cultures. These are putative piRNA biogenesis sites, similar to the nuage in germ cells.

Selected references

Zheng, K., Xiol, J., Reuter, M., Eckardt, S., *et al* (2010). Mouse MOV10L1 associates with Piwi proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway. *Proc. Natl. Acad. Sci. USA*, 107(26), 11841-6

Simon, B., Kirkpatrick, JP., Eckhardt, S., Reuter, M., Rocha, EA., Andrade-Navarro, MA, Sehr, P., Pillai, R., Carlomagno, T. (2010). Recognition of 2'-O-methylated 3'-end of piRNA by the PAZ domain of a Piwi protein (in press)

Shoji, M., Tanaka, T., Hosokawa, M., *et al*. (2009). The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. *Dev. Cell,* 17, 775-787

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, 639-646

Christiane Schaffitzel

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

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 multi-component translation machinery at close to physiological conditions. By using state-of-the-art electron microscopes and by image processing of large data sets, EM structures of prokaryotic and eukaryotic ribosomes have been obtained at subnanomolar resolution, 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 the 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 or stalled at a defined step in translation. 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 cryo-EM structures of the complex of the ribosome with the translocation machinery SecYEG (figure 1), of the translating ribosome-signal recognition particle (SRP) complex and of the ribosome in complex with SRP and SRP receptor (figure 2). 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 targeting, membrane protein integration, folding and assembly. We analyse the membrane protein complexes biochemically, by nanoelectrospray mass spectrometry (collaboration with Carol Robinson, University of Cambridge) and cryo-electron microscopy.

In collaboration with the groups of Stephen Cusack, Matthias Hentze and Andreas Kulozik (MMPU, see page 50), we study mammalian ribosomal complexes involved in nonsense-mediated mRNA decay. We produce the eukaryotic factors involved by means of advanced recombinant eukaryotic technologies in collaboration with the Berger group (page 92).

Figure 1: 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.

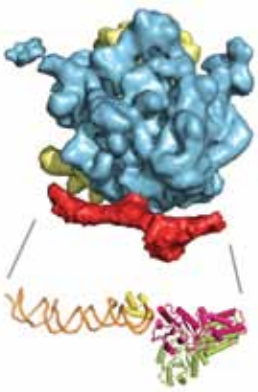


Figure 2: Cryo-EM structure of the ribosome (yellow/blue) bound to the signal recognition particle (SRP) and the SRP receptor (both in red). Below it is an atomic model of SRP (green-yellow/orange) and its receptor (pink).

Selected references

Estrozi, L.F., Boehringer, D., Shan, S.O., Ban, N. & Schaffitzel, C. (2011) Structure of the *E. coli* Co-translational Targeting Complex in the Stable Early Conformation. *Nature Structural & Molecular Biology* 18, 88-90

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

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

Mitra, K., Schaffitzel, C., Shaikh, T., Tama, F., Jenni, S., Brooks, C.L., Ban, N. & Frank, J. (2005) Structure of the *E. coli* protein-conducting channel bound to a translating ribosome. *Nature*, 438, 318-324

EMBL Hamburg Germany

EMBL Hamburg's laboratories are on the German Synchrotron Research Centre (DESY) campus, with synchrotron radiation (DORIS III) and laser (FLASH) facilities available. The PETRA storage ring is presently being converted into a world-leading, dedicated synchrotron radiation facility, PETRA III, 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. 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, for the first time, to unravel the mechanism involved in 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.

Structural systems biology in *M. tuberculosis*: During the last four 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 and 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.

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

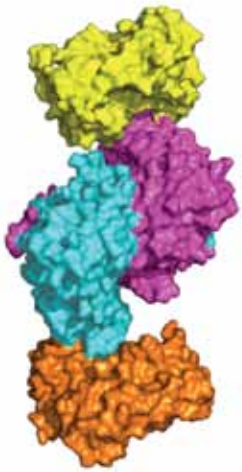
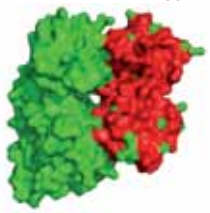


Figure 2: DAPK-CaM complex. Colour codes: DAPK, green; CaM, red.



Selected references

Schueller, N., Holton, S.J., Fodor, K., Milewski, M., Konarev, P., Stanley, W.A., Wolf, J., Erdmann, R., Schliebs, W., Song, Y.H. & Wilmanns, M. (2010). The peroxisomal receptor Pex19p forms a helical mPTS recognition domain. *EMBO J.*, 29, 2491-2500

Poulsen, C., Akhter, Y., Jeon, A.H., Schmitt-Ulms, G., Meyer, H.E., Stefanski, A., Stuhler, K., Wilmanns, M. & Song, Y.H. (2010). Proteome-wide identification of mycobacterial pupylation targets. *Mol. Syst. Biol.*, 6, 386

Sauer, F., Vahokoski, J., Song, Y.H. & Wilmanns, M. (2010). Molecular basis of the head-to-tail assembly of giant muscle proteins obscurin-like 1 and titin. *EMBO Rep.*, 11, 534-540

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

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

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 a test platform for developments for the PETRA III beamlines (see also Hermes group, page 103) 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 III.

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

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.

Robotic sample mounting system for protein crystals (MARVIN)

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. *J. Appl. Cryst.*, 40, S190-S194

Christoph Hermes

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

Instrumentation for structural research

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, detailed 3D picture can be obtained by protein X-ray crystallography (PX), which has become the dominant structural research tool in molecular biology. Each of these methods have 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 focusing 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.

As the intensity levels available in ML mode are identical to the values expected at PETRA III (although the beam is far less focused and monochromatic) this beamline has, for over a year, been operating as a test station for equipment to be potentially used at the new EMBL beamlines at PETRA III. These tests are providing extremely useful information on layout and design parameters of instruments built in-house, as well as in assessing the usefulness of commercial components. The newly developed beamline control system, based on economic industrial electronics and improved software, has proven both reliable and user-friendly.

We are continuously optimising the end-stations of the 'old' beamlines at the DORIS storage ring through the continual development of our automatic sample changer for protein crystals. This is the prototype of instruments foreseen to be installed at the PETRA III PX beamlines.

Future projects and goals

We will continue to improve the spectral quality of our existing beamlines, while at the same time we aim 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. Major components are currently being installed and will undergo extensive testing before being commissioned. In this context, very close collaboration between instrumentation and scientific groups will be required , to both rise to challenges and create opportunities in the fields of beamline instrumentation, sample handling, control electronics and software (see the Fiedler group, page 102). The opening of the European X-Ray Laser (XFEL) in Hamburg is scheduled for 2014, offering unprecedented research opportunities. Designing experiments which will exploit the potential of this unique facility requires us to overcome a large number of problems in various areas and at the same time presents great potential for significant advancements in structural research.

Our ultimate 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

Round, R., Franke, D., Moritz, S., *et al*. (2008). Automated sample-changing robot for solution scattering experiments at the EMBL Hamburg SAXS station X33. *J. Appl. Cryst*., 41, 913-917

Roessle, M.W., Klaering, R., Ristau, U., Robrahn, B., Jahn, D., *et al*. (2007). Upgrade of the small-angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg. *J. Appl. Cryst.*, 40,, S190-S194

Pohl, E., Ristau, U., Gehrmann, T., *et al*. (2004). Automation of the EMBL Hamburg protein crystallography beamline BW7B. *J. Synchrotron Radiat.*, 11, 372-377

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

Victor S. Lamzin

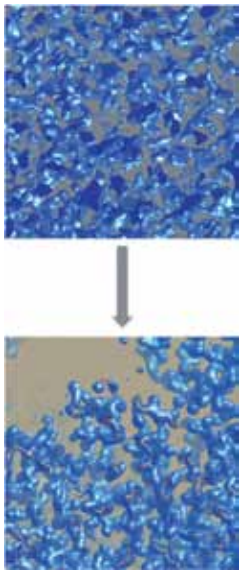
PhD 1987, Moscow State University.

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

We are fascinated by complex computational methods in information processing that can address data interpretation problems as we encounter them in structural biology. Recognising patterns in experimental data that describe macromolecules is 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, predominantly 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 novel approaches where a wider context of information, originating from complementary tools, is implemented in order to arrive at an integrated platform for a model of life.



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

Previous and current research

Pattern-recognition based methods are the foundation of one of the group's main foci, the ARP/wARP software project (Langer *et al.*, 2008) for protein/DNA/ligand crystal structure determination. The already comprehensive range of implemented methodologies and algorithms, complemented by an intuitive graphical user interface, is being continually improved and augmented with new procedures for dealing with more challenging problems of structural biology (Hattne & Lamzin, 2011). We can now recognise structural motifs in lower resolution maps (Heuser *et al.*, 2009), which should permit the combination of X-ray crystallographic data with that derived from electron microscopy. In a similar vein, sophisticated algorithms are being developed for the modelling of bound ligands and identification of novel binding sites.

The design of artificial intelligence to find the best way through the maze of structure determination protocols available is implemented in the group's AutoRickshaw software for validation of synchrotron beamline experiments and the building of macromolecular models. The group continues to enhance the range of state-of-the-art computational facilities that enable hundreds of researchers worldwide to perform remote structure determination.

The group uses its expertise to develop methods that aid the analysis of crystal growth. We have developed novel software XREC/FREC for the detection of crystals *in situ* and aim at downstream applications in the automation of experimental sample handling (Watts *et al.*, 2010). We also apply our techniques to the structural biology-driven characterisation of proteins from humans and their pathogens. Current targets include macromolecules from the malarial oxidative phosphorylation pathway, enzymes involved in drug and vitamin syntheses, and those related to the amyloidal fibril formation process (Wrenger *et al.*, 2011; Lapkouski *et al.*, 2009).

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. Coherent free-electron laser sources in Stanford and at DESY already allow experiments with biological samples that were previously unimaginable. Additionally, the state-of-the-art PETRA III synchrotron beamlines will soon be available. The field will play an increasing role in investigations of biological interactions, and we will remain at the forefront of developments.

Hattne J. & Lamzin V.S. (2011) A moment invariant for evaluating the chirality of three-dimensional objects. *J. R. Soc. Interface.* 8, 144-151

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.

Wrenger C., Müller I.B., Schifferdecker A.J., Jain R., Jordanova R. & Groves M.R. (2011) Specific Inhibition of the Aspartate Aminotransferase of Plasmodium falciparum. *J. Mol. Biol.* 405, 956-971.

Selected references

Watts, D., Muller-Dieckmann, J., Tsakanova, G., Lamzin, V.S. & Groves, M.R. (2010). Quantitive evaluation of macromolecular crystallization experiments using 1,8-ANS fluorescence. *Acta Crystallogr. D. Biol. Crystallogr.*, 66, 901-908

Schmidt, A. & Lamzin, V.S. (2010). Internal motion in protein crystal structures. *Protein Sci.*, 19, 944-953

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. Mol. Biol.*, 16, 94-95

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

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.

Structural biology of cell surface receptors

Previous and current research

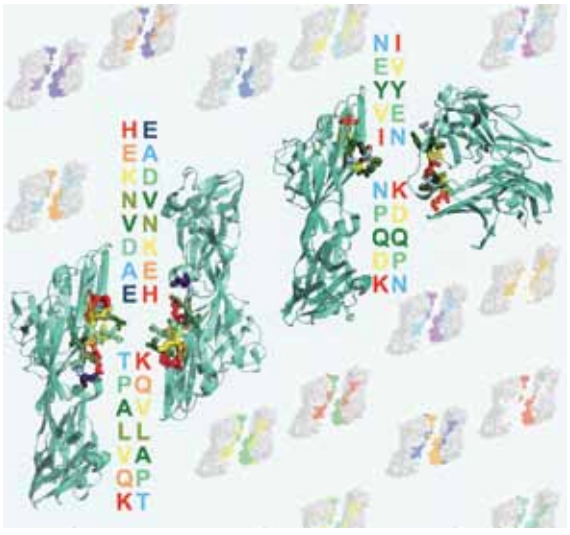
Intercellular interactions occur through supramolecular clusters that form asymmetric kinapses when a cell is scanning the environment, and symmetric synapses when cell–cell recognition is established. This ancient recognition process has been observed in such diverse systems as predatory amoeba, lymphocytes 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.

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.



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 19 008 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.*

Selected references

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, 487-491

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, 1099-1110

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, 10799-10804

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.

Jochen Müller-Dieckmann

X-ray crystallography, high-throughput crystallisation

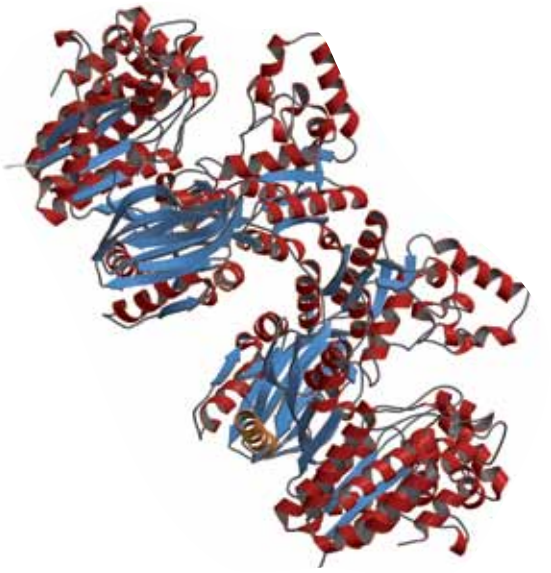
Previous and current research

Structural biology: Our goal is to understand the molecular principles of specific disease related proteins and those of signal transduction across membranes. For the latter, we have chosen a relatively simple yet well-characterised system from the model plant *Arabidopsis thaliana*. Here, we try to structurally characterise a group of functionally related, membrane bound receptors and their interaction with a common cytosolic downstream target, a protein kinase. The system regulates the response to the simple phytohormone ethylene.

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 *A. thaliana*. The functional unit of the receptors are disulfide linked dimers. All receptors share a small, highly conserved ER membrane embedded domain, which contains a single ethylene binding site. The cytosolic domain structure is related to bacterial two-component systems (TCS). In contrast to classical TCS, the immediate downstream target of the receptors, however, is CTR1, a Raf-like Ser/Thr protein kinase. Thus, this signalling pathway presents an interesting case, wherein a two component signalling system manipulates a MAP kinase kinase kinase (MAP3K) and possibly a MAP3K signalling 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 the goal to functionally characterise the early steps of ethylene signalling 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. Currently, users come from over 20 different nations. Since 2009 our platform is part of the European FP7 initiative P-Cube, which supports access to advanced infrastructures. More information on this initiative is available at: www.p-cube.eu.



Ribbon representation of Rv0066c (lcd2) of M. tuberculosis. *Shown is a dimer of lcd2 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.*

Future projects and goals

We have functionally and structurally characterised the active and inactive forms of the protein kinase domain of CTR1, as well as two sub-domains of the cytosolic ethylene receptor. Our next goals are the characterisation of larger parts of the receptor, including the full-length construct. We are also working on complexes of CTR1 with the receptors domains. The latter requires an interdisciplinary approach due to transient complex formation and unpredictable crystallisability. To this end, we will use a combination protein characterisation tools, X-ray crystallography and SAXS.

Selected references

Hatzopoulos, G.N. & Mueller-Dieckmann, J. (2010). Structure of translation initiation factor 1 from Mycobacterium tuberculosis and inferred binding to the 30S ribosomal subunit. *FEBS Lett.*, 584, 1011-1015

Watts, D., Müller-Dieckmann, J., Tsakanova, G., Lamzin, V.S. and Groves, M. R. (2010). Quantitive evaluation of macromolecular crystallization experiments using 1,8-ANS fluorescence *Acta Cryst.*, D66, 901-990

Mayerhofer, H., Mueller-Dieckmann, C. & Mueller-Dieckmann, J. (2010). Cloning, expression, purification, crystallization and preliminary X-ray crystallographic analysis of the protein kinase domain of constitutive tripe response 1 (CTR1) from *Arabidopsis thaliana*. *Acta Crystallogr.*, F67, 117-120

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.

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 102).

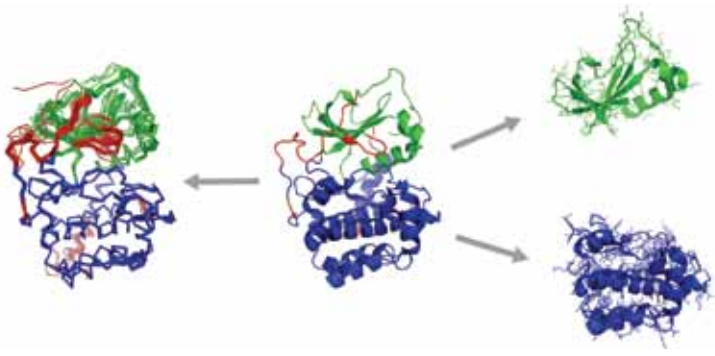
Partly due to the enormous progress in synchrotron radiation-based structural biology, structural data on biological macromolecules are produced at an ever-increasing rate, creating the need to develop tools for efficient mining of structural data. We are developing tools for which the central concept 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 as many as 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 investigate 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 2011. In small-angle X-ray scattering, the new beamlines will enable us to work with more complex and more dilute samples than presently possible. In macromolecular crystallography, the beamlines will provide features such as micro-focusing and energy tunability, allowing imaging of the content of small crystals containing large objects such as multi-component complexes.

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



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

Selected references

Schneider, T.R. (2008). Synchrotron radiation: micrometer-sized X-ray beams as fine tools for macromolecular crystallography. *HFSP J.*, 2, 302-306

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

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

Dmitri Svergun

PhD 1982, Institute of Crystallography, Moscow.

Dr. of Science 1997, Institute of Crystallography, Moscow.

At EMBL since 1991. Group leader since 2003.

Selected references

Prischi, F., Konarev, P.V., Iannuzzi, C., *et al* (2010). Structural bases for the interaction of frataxin with the central components of iron-sulphur cluster assembly. *Nat. Commun.*, 1, 95

Bertini, I., Giachetti, A., Luchinat, C., Parigi, G., Petoukhov, M.V., Pierattelli, R., Ravera, E. & Svergun, D.I. (2010). Conformational space of flexible biological macromolecules from average data. *J. Am. Chem. Soc.*, 132, 13553-13558

Albesa-Jove, D., Bertrand, T., Carpenter, E.P., *et al*. (2010). Four Distinct Structural Domains in Clostridium difficile Toxin B Visualized Using SAXS. *J. Mol. Biol.*, 396, 1260-1270

Jurk, M., Dorn, M., Kikhney, A., Svergun, D., Gartner, W., Schmieder, P. (2010). The switch that does not flip: the blue-light receptor YtvA from Bacillus subtilis adopts an elongated dimer conformation independent of the activation state as revealed by a combined AUC and SAXS study. *J. Mol. Biol.* 403: 78-87

Small-angle X-ray scattering from macromolecular solutions

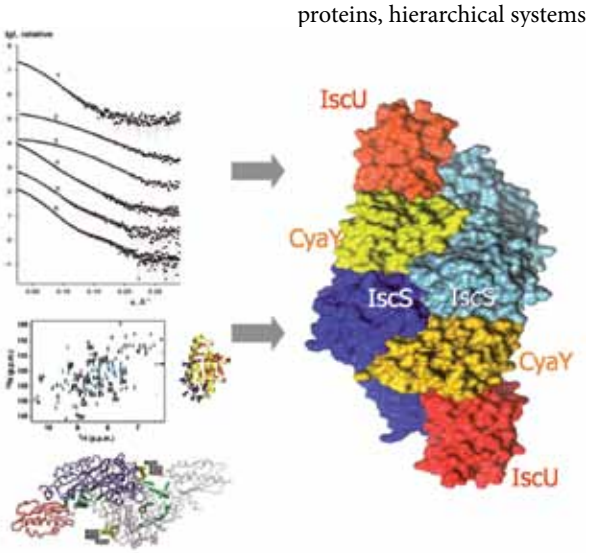
Previous and current research

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. 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 the SAXS-based structural models. This versatility and universality – and the fact that it does not need crystals to characterise the structure – make SAXS an ideal tool for systems structural biology, and the last decade saw a renaissance of biological SAXS worldwide.

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 macromolecular crystallography, NMR, electron microscopy, neutron scattering and bioinformatics. We developed the world's most used SAXS data analysis program package, ATSAS, which is employed in over 1300 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 more than six-fold increase in the user demand 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, with FedEx-style and remote data access options. All the X33 developments are being ported to the new high brilliance BioSAXS beamline, which is a presently under commissioning at the third-generation PETRA III storage ring at DESY.

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 analysis. In numerous exciting applications, SAXS is employed to study domain structure of individual macromolecules, conformational transitions (e.g. upon ligand binding), quaternary structure of complexes (see figure), but also oligomeric mixtures, intrinsically unfolded proteins, hierarchical systems and other objects of high biological and medical importance.



Quaternary structure of the functional complex of frataxin bacterial orthologue (CyaY) with the proteins IscS and IscU. The SAXS-derived rigid body model was validated by NMR and site-directed mutagenesis (Prischi et al *2010).*

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, electron microscopy and other methods;
- participation in collaborative projects at the X33 beamline, employing SAXS to study the structure of a wide range of biological systems in solution;
- complete automation of a biological SAXS experiment and data analysis at X33, and, in collaboration with the PETRA III group, at the new high-brilliance BioSAXS beamline.

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.

Scientific Head, EMBL Australia, 2010

Regenerative mechanisms in heart and skeletal muscle

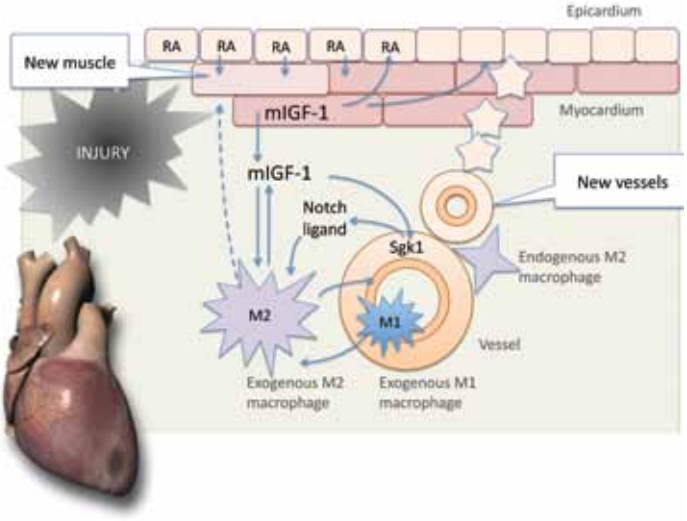
Previous and current research

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. Over the past decade we have used mouse genetics to enhance naturally acting signalling pathways that have proven highly effective in countering tissue decline. Local supplementation of Insulin-like Growth Factor-1 (IGF-1) propeptides orchestrates efficient repair of injured skeletal muscle tissues without scar formation, prevents age- and heart failure-related muscle atrophy and enhances bone marrow cell recruitment to the damaged tissue. Importantly, neither fully processed IGF-1 nor systemically administered IGF-1 counteracts muscle loss.

Since IGF-1 curtails the expression of inflammatory cytokines, we have investigated the role of the innate immune system in the regeneration process, using a genetic model in which prevention of macrophage polarisation blocks muscle regeneration. We have extended these findings to reveal the molecular and cellular mechanisms whereby localised IGF-1 propeptides induce effective tissue regeneration in the damaged heart. In our mouse model of enhanced cardiac repair, injury of the cardiomyocyte-specific mIGF-1 transgenic hearts activates a series of signalling intermediates, specifically in the epicardium - a layer of cells surrounding the heart that give rise to the coronary vasculature during embryogenesis. We are using mouse genetics to establish a role for exogenous or endogenous macrophages in promoting effective vascularisation and cardiomyocyte replacement in the uninjured and damaged heart. Supplemental mIGF-1 polarises macrophages to synthesise endogenous IGF-1 propeptides that target myocytes (improving survival by depressing NFkB protein turnover, activating SirT1 and inducing new muscle formation), endothelial cells (inducing cardiac regrowth by activating Sgk1 and Notch) and macrophages themselves (working indirectly to promote revascularisation, and directly as a stimulus to generate new muscle). We are pursuing the possibility that mIGF-1 potential progenitor cells for tissue regeneration may be disguised as components of the immune system, which can be coaxed into more productive engagement in the repair process with relatively simple, clinically applicable manipulations.

Future projects and goals

In our future research, we will exploit new conditional and inducible mouse genetic models and transgenic markers to characterise key cells and molecules governing the regeneration of mammalian tissues. We aim to define the signalling mechanisms whereby selected growth factors and their intracellular intermediates modulate immune cell lineages in control inflammation and in promoting tissue repair. We hope to use this knowledge for developing clinically relevant interventions in ageing, injury and degenerative disease.



Model of enhanced cardiac repair mediated by mIGF-1: Secreted locally by cardiomyocytes, transgenic mIGF-1 polarises invading macrophages to secrete their own mIGF-1, and stimulating RA synthesis in the epicardium, which induces new cardiomyocte formation in response to injury. In adjoining vasculature, mIGF-1 activates Sgk1, which induces Notch signalling in the vascular bed. This stimulates M2 macrophages to promote angiogenesis through endothelial cells and epicardial derivatives, with anastomosis by resident M2 macrophages.

Selected references

Bochmann L., Sarathchandra P., Mori F., Lara-Pezzi E., Lazzaro D., Rosenthal N. (2010) Revealing new mouse epicardial cell markers through transcriptomics. *PLoS One*, 5: (epub)

Temmerman L., Slonimsky E., Rosenthal N. (2010) Class 2 IGF-1 isoforms are dispensable for viability, growth and maintenance of IGF-1 serum levels. *Growth Hormone & IGF Research*, 20:255-263.

Catela C., Kratsios P., Hede M., Lang F., Rosenthal N. (2010) Serum and Glucocorticoid- inducible Kinase 1 (SGK1) is necessary for vascular remodeling during angiogenesis. *Dev. Dyn*, 239: 2149-60.

Vinciguerra M., Santini MP., Claycomb WC., Ladurner AG., Rosenthal N. (2010) Local IGF-1 isoform protects cardiomyocytes from hypertrophic and oxidative stresses via SirT1 activity. *Aging* 2, 43-62.

Cornelius Gross

PhD 1995, Yale University.
Postdoctoral research at Columbia University.
Group leader at EMBL Monterotondo since 2003.
Deputy Head of outstation and senior scientist since 2009.

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 humans.

Cellular substrates of anxiety: To help identify the cellular substrates of anxiety, we are using pharmaco-genetic transgenic tools for the rapid modulation of electrical activity in selected cell-types in the brain. We have used a pharmaco-genetic inhibition strategy to examine the contributions of hippocampal and amygdala cell-types to anxiety and fear 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

Gozzi, A., Jain, A., Giovanelli, A., Bertollini, C., Crestan, V., Schwarz, A.J., Tsetsenis, T., Ragozzino, D., Gross, C.T. & Bifone, A. (2010). A neural switch for active and passive fear. *Neuron*, 67, 656-666

Audero, E., Coppi, E., Mlinar, B., Rossetti, T., Caprioli, A., Banchaabouchi, M.A., Corradetti, R. & Gross, C. (2008). Sporadic autonomic dysregulation and death associated with excessive serotonin autoinhibition. *Science*, 321, 130-133

Lo Iacono, L. & Gross, C. (2008). Alpha-Ca2+/calmodulin-dependent protein kinase II contributes to the developmental programming of anxiety in serotonin receptor 1A knock-out mice. *J. Neurosci.*, 28, 6250-6257

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

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.

Paul Heppenstall

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 neurons;
- 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 Ca2+. *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

Martin Jechlinger

PhD 2002, Institute of Molecular Pathology (IMP) Vienna.

Postdoctoral research at Memorial Sloan-Kettering Cancer Center (MSKCC) New York.

Group leader at EMBL Monterotondo since 2010.

Mechanisms of oncogene dependence and tumour relapse

Previous and current research

Extensive evidence now supports the concept of oncogene addiction (the dependence of tumour cells on their initiating lesion for survival). In patients and mouse models interference with the activity of cancer-initiating oncogenes can result in tumor regression. However, novel therapies that target the products of mutant alleles in human cancers are only partly successful, since maintenance of remission requires long-term treatment and relapse often occurs in the presence of therapeutic agents. Hence, a better understanding of drug resistance and tumour recurrence is needed for the design of more successful anti-cancer strategies.

Transgenic mice carrying regulatable transgenes represent tractable systems for studying the mechanisms of oncogene dependence, the response and resistance to targeted drugs and tumour recurrence. In a complementary approach, we have developed a 3D culture system of primary mouse mammary epithelial cells to study detailed responses to the induction and de-induction of oncogenes (mimicking treatment with an ideally targeted drug). This 3D system produced phenotypic changes similar to those observed in the mammary glands of the transgenic mice from which the cultures were derived. In addition, this new approach identified and isolated cells that had survived oncogene withdrawal, exhibited characteristics of mammary gland progenitors and could efficiently re-populate the mammary fat pads of immundeficient mice. The successful isolation of a pure population of surviving cells after oncogene withdrawal will allow us to characterise these residual 'dormant' tumour cells in detail.

Future projects and goals

- Determine at which point during tumorigenesis cells acquire the ability to survive oncogene withdrawal.
- Identify the molecular properties that distinguish surviving-residual cells, from naïve cells.
- Interfere with the mechanisms important for survival of residual 'dormant' cells.

Selected references

Jechlinger, M., Podsypanina, K., Varmus, H. (2009). Regulation of transgenes in three-dimensional cultures of primary mouse mammary cells demonstrates oncogene dependence and identifies cells that survive deinduction. *Genes Dev.,* 23(14), 1677-88

Podsypanina, K., Du, Y.C., Jechlinger, M., Beverly, L.J., Hambardzumyan, D., Varmus, H. (2008) Seeding and propagation of untransformed mouse mammary cells in the lung. *Science,* 321, 1841-4

Sotillo, R., Schvartzman, J.M., Socci, N.D., Benezra, R. (2010). Mad2-induced chromosome instability leads to lung tumor recurrence after oncogene withdrawal. *Nature*, 464, 436-40

Sotillo, R., Hernando, E., Diaz-Rodriguez, E., Teruya-Feldstein, J., Cordon-Cardo, C., Lowe, S.W., Benezra, R. (2007) Mad2 overexpression promotes aneuploidy and tumorigenesis in mice. *Cancer Cell,* 11, 9-23

Christophe Lancrin

PhD 2003, Université Pierre et Marie Curie (Paris VI), Paris, France.

Postdoctoral research at the Paterson Institute for Cancer Research, Manchester, United Kingdom.

Group Leader at EMBL since January 2011.

The haemogenic endothelium: a key stage in the generation of the first blood cells

Previous and current research

The continuous generation of blood cells throughout life relies on the existence of haematopoietic stem cells (HSC) generated during embryogenesis. They have the ability to self-renew and to generate all types of blood cells. Any pathology affecting these cells could lead to development of serious diseases such as leukaemia and anaemia. That is why understanding how HSC and haematopoietic progenitors are produced during embryonic life is important.

The origin of blood cells has been the subject of an intense scientific debate during the last decade. It has been proposed that during embryonic development, haematopoietic cells arise from a mesodermal progenitor with smooth muscle, endothelial and haematopoietic potential called the haemangioblast. However, a conflicting theory instead associates the first haematopoietic cells with a phenotypically differentiated endothelial cell with haematopoietic potential, i.e. a haemogenic endothelium.

To investigate the cellular origin of blood cells, we used a model of early haematopoiesis based on the differentiation potential of the mouse embryonic stem cells (ESC) *in vitro*. These cells are derived from the inner cell mass of the blastocyst, an early-stage mouse embryo, and have the capacity to generate any cell types. Using this system coupled with time-lapse microscopy, clonogenic assays and flow cytometry analysis, we have demonstrated that the haemangioblast generates haematopoietic progenitors through the formation of a haemogenic endothelium stage, providing the first direct link between these two precursor populations. Together our results merge the two *a priori* conflicting theories on the origin of haematopoietic development into a single linear developmental process. This finding allowed us to identify the haemogenic endothelium as the immediate precursor of blood cells (figures 1 and 2).

Future projects and goals

Recently, the generation of the ESC-like induced pluripotent stem cells (iPSC) from fully differentiated cell type, such as skin fibroblast, provided a major breakthrough in the field of regenerative medicine. Indeed iPSC offer a great opportunity to implement replacement therapy by bypassing the use of human embryos to generate ESC, therefore decreasing ethical concerns. However, important work has to be done to differentiate efficiently iPSC or ESC toward specific cell type including blood cell progenitors such as HSC.

Consequently, in order to better understand the development of the haematopoietic system, the focus of our research is to unravel the mechanisms underlying the generation of haemogenic endothelium from its precursor, the haemangioblast, and its subsequent commitment to haematopoiesis. Combining genomics, time-lapse microscopy, and loss and gain of function experiments *in vitro* and *in vivo*, we plan to identify and study the genes responsible for the generation of the first blood progenitors during embryonic life. Our research will bring a further understanding of the mechanisms of cell fate decisions leading to the production of the first haematopoietic cells and enable the development of new strategies to improve methods of blood cells generation from ESC or iPSC for regenerative medicine.

Selected references

Sroczynska, P., Lancrin, C., Kouskoff, V. & Lacaud, G. (2009) The differential activities of Runx1 promoters define milestones during embryonic hematopoiesis. *Blood*, 114(26), 5279-5289.

Hoogenkamp, M., Lichtinger, M., Krysinska, H., Lancrin, C., Clarke, D., Williamson, A., Mazzarella, L., Ingram, R., Jorgensen, H., Fisher, A., Tenen, D.G., Kouskoff, V., Lacaud, G. & Bonifer, C. (2009) Early chromatin unfolding by RUNX1: a molecular explanation for differential requirements during specification versus maintenance of the hematopoietic gene expression program. *Blood,* 114(2), 299-309

Lancrin, C. Sroczynska, P., Stephenson, C., Allen, T., Kouskoff, V. & Lacaud, G. (2009) The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage. *Nature*, 457(7231), 892-895.

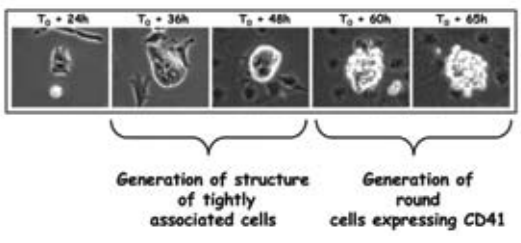


Figure 1: Time-lapse microscopy analysis of haemangioblast differentiation
The formation of a blast colony from the haemangioblast can be retrospectively divided in 2 consecutive phases: a generation of a structure of tightly associated endothelial cells and the production of round non adherent cells expressing the haematopoietic marker CD41.

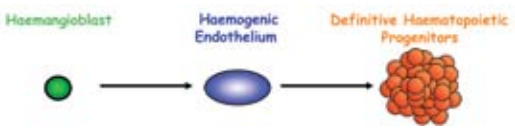


Figure 2: New model of blood cell origin: The haemangioblast and the haemogenic endothelium are part of the same developmental pathway to generate blood cell progenitors during embryonic life.

Dónal O'Carroll

PhD 1999, Research Institute of Molecular Pathology, Vienna.

Postdoctoral research at The Rockefeller University, New York.

Group leader at EMBL Monterotondo since 2007.

Adjunct member of faculty, The Rockefeller University, since 2007

Argonaute protein function in mammalian development and physiology

Previous and current research

The aim of the group is to understand the mechanisms by which Argonaute proteins contribute to mammalian development and physiology with an emphasis on mouse hematopoiesis, embryology and germ cell development. We address our challenges using state-of-the-art mouse genetic strategies coupled with high throughput sequencing approaches.

Members of the Argonaute family are highly conserved small-RNA-binding proteins with diverse functions ranging from the regulation of post-transcriptional gene expression, anti-viral defense, transposon silencing to the establishment of heterochromatin domains. The Argonaute family is defined by the presence of both the Paz and Piwi domains. Paz domains bind to the 3' end of small RNAs, whereas the Piwi domain adopts a classical RNase H fold with many Argonaute proteins being active small-RNA guided Slicer endonucleases. In mammals two subclades of Argonaute proteins exist, the Ago and Piwi sub-families.

MicroRNAs (miRNAs) and short-interfering RNAs (siRNA) are small non-coding RNA molecules that are potent negative regulators of gene expression. MiRNA/siRNA-mediated gene silencing is executed by the multi-protein RNA-induced silencing complex (RISC). At the core of RISC is an Ago protein that binds a small-RNA 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 among the Ago proteins Ago2 selectively controls early development of B lymphoid and erythroid cells. Our current RISC interests now focus on the physiological importance of Ago2 post-translational regulation *in vivo*. Having identified erythroid development as being sensitive to miRNA dosage, we now strive to mechanistically understand how key miRNA loci and Ago2 regulates terminal erythropoiesis.

Piwi proteins 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 ribonuclear particles. In the mammalian male germline, the members of the Piwi subclade of the Argonaute family, Mili and Miwi2, are essential for *de novo* DNA methylation of transposons and spermatogenesis. We currently address several basic questions on the intrinsic mechanism and function of mammalian Piwi proteins in and beyond transposon silencing.

Future projects and goals

- Identify miRNAs and their respective targets that control erythropoiesis.
- Determine the *in vivo* significance of post-translational modifications of Ago2.
- Understand the mechanism of Miwi2 and Mili in the process of epigenetic transposon silencing.
- Explore the function of Miwi2 and Mili during spermatogenesis beyond transposon silencing.

Selected references

Schaefer, A., Im, H.I., Veno, M.T., Fowler, C.D., Min, A., Intrator, A., Kjems, J., Kenny, P.J., O'Carroll, D. & Greengard, P. (2010). Argonaute 2 in dopamine 2 receptor-expressing neurons regulates cocaine addiction. *J. Exp. Med.*, 207, 1843-1851

Yang, J.S., Maurin, T., Robine, N., *et al*. (2010). Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. *Proc. Natl. Acad. Sci. USA*, 107, 15163-15168

Rasmussen, K.D., Simmini, S., Abreu-Goodger, C., Bartonicek, N., Di Giacomo, M., Bilbao-Cortes, D., Horos, R., Von Lindern, M., Enright, A.J. & O'Carroll, D. (2010). The miR-144/451 locus is required for erythroid homeostasis. *J. Exp. Med.*, 207(7):1351-8

O'Carroll, D., Mecklenbrauker, I., Das, P.P., *et al*. (2007). A Slicer-independent role for Argonaute 2 in hematopoiesis and the microRNA pathway. *Genes Dev.*, 21, 1999-2004

Index

Notes

Published by the Office of Information and Public Affairs

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

European Molecular Biology Laboratory

Heidelberg • Hinxton • Grenoble • Hamburg • Monterotondo

EMBL member states:
Austria, Belgium, Croatia, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom. Associate member state: Australia.

www.embl.org