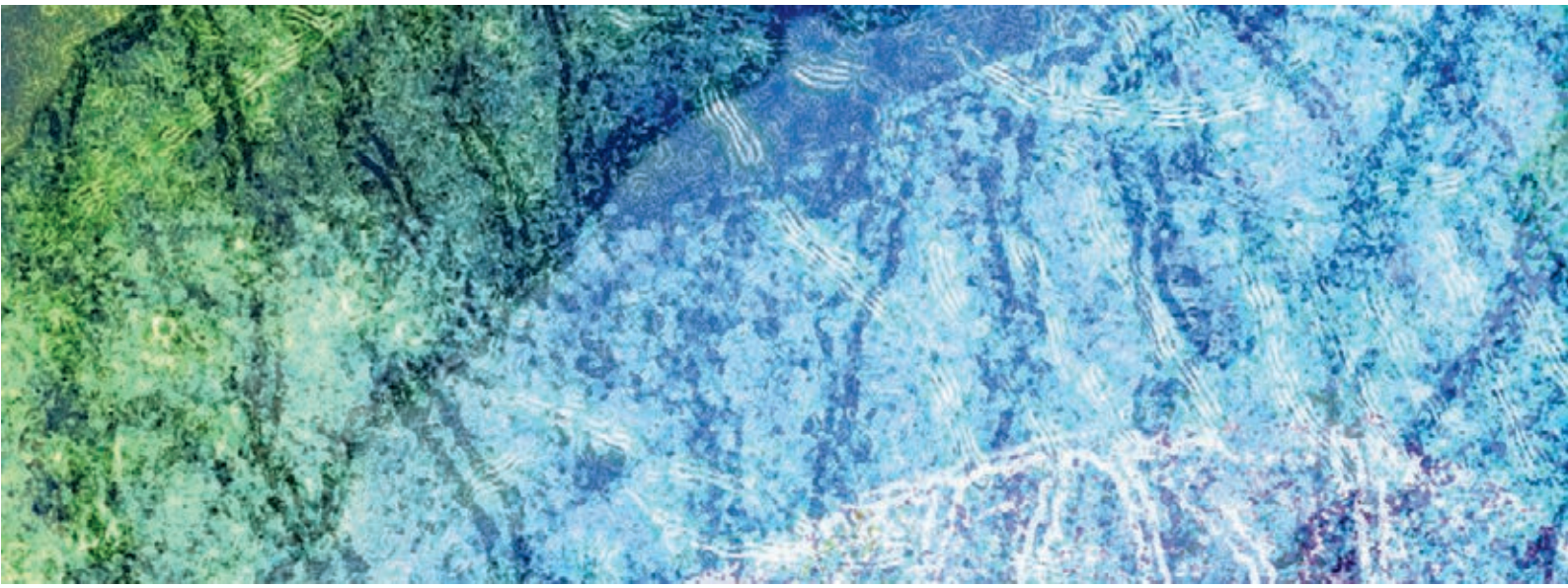
European Molecular Biology Laboratory

Research at a Glance 2012



EMBL

European Molecular Biology Laboratory

Research at a Glance 2012

EMBL

www.embl.org

Contents

"EMBL combines a critical mass of expertise and resources with organisational flexibility, enabling us to keep pace with today's biology." – Iain Mattaj

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 and related fields.

During the intervening 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. EMBL is also heavily engaged in discussions of science policy and strategy at the European level, where it coordinates multiple activities. 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, interdisciplinary and international environment. Freedom, flexibility and a regular turnover of staff allow 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 that 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

About EMBL

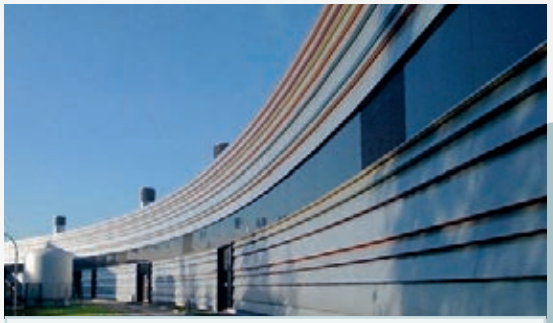
The European Molecular Biology Laboratory (EMBL) is a world-class international research organisation, with some 85 independent groups covering the spectrum of molecular biology. Scientists represent disciplines including biology, chemistry, physics, and computer science across EMBL's five sites: the main laboratory in Heidelberg, Germany, and four outstations: Grenoble, France; Hamburg, Germany; Monterotondo, Italy; and the European Bioinformatics Institute (EMBL-EBI), Hinxton, United Kingdom.

Europe's flagship laboratory for the life sciences

EMBL was founded in 1974 to create a central European laboratory in the emerging field of molecular biology. It remains the only intergovernmental research organisation in Europe that performs research in the molecular life sciences, and is directly supported by 20 member states and one associate member, Australia. EMBL's goals are:

- Outstanding life science research: setting trends and pushing the limits of technology.
- Providing world-class research infrastructure and services to the member states.
- Training and inspiring the next generation of scientific leaders.
- Driving research, innovation and progress through technology development, interactions with industry and technology transfer.
- Taking a leading role in the integration of life science research in Europe.

The European Bioinformatics Institute (EMBL-EBI) is located on the Wellcome Trust Genome Campus in Hinxton, near Cambridge. As a European hub for biomolecular data, the EBI offers the scientific community access to a variety of bioinformatics services, alongside which, a number of active research groups work in areas that complement and extend these services.



EMBL Hamburg develops novel, innovative technologies in structural biology, such as high-throughput crystallisation and data interpretation software as well as operating cutting-edge synchrotron radiation beamlines and offering world-leading facilities and expertise to the research community. It also has an ambitious research programme for structures of multifunctional proteins and protein complexes of biomedical relevance.

EMBL Heidelberg is home to five research units, central scientific services, the administration, and the laboratory's technology transfer arm, EMBL Enterprise Management (EMBLEM). Heidelberg is the largest centre for biomedical research in Germany and there are many bilateral links between EMBL scientists and local research institutions.

Hinxton

Hamburg

Heidelberg

Grenoble

Monterotondo

EMBL Grenoble builds and operates beamlines for macromolecular crystallography, develops instrumentation and techniques, and provides facilities and expertise to visitors in collaboration with its campus partners, the European Synchrotron Radiation Facility (ERSF) and the Institut Laue-Langevin (ILL). The outstation is also part of the Unit of Virus Host Cell Interactions (UVHCI).

EMBL Monterotondo, near Rome, focuses on mouse genetics and functional genomics, and offers expertise in mammalian physiology and production of mouse models of human diseases. Researchers form dynamic partnerships with other international research and clinical centres. The outstation shares a campus with Italian national research groups (IBC-CNR) and the headquarters of the the European Mouse Mutant Archive.

Career Opportunities

At EMBL's five sites, there are opportunites across the spectrum of life science research for PhD students, postdoctoral fellows, group leaders, and many other professionals – from software development to chemistry and engineering. EMBL's staff comprises more than 1500 people from over 70 different countries – this internationality creates an atmosphere that is creative, interdisciplinary and collaborative, with an unparalleled breadth of expertise and complementary skills.

PhD programme

Training is one of EMBL's core missions and our International PhD Programme is renowned for offering excellent education to prospective scientists.

Research independence, dedicated mentoring and an international environment are the cornerstones of the programme, in which close to 200 students from all over the world are currently enrolled.

EMBL students obtain their degree from a national university or jointly with EMBL and one of its 29 prestigious partner universities in the member states.

EMBL recruits PhD students twice a year. For more details please check **www.embl.de/training/eipp** or contact **predocs@embl.de**

Postdoctoral fellows

Postdoctoral fellows at EMBL benefit from the expertise of world class scientists, state-of-the-art scientific equipment, training in career development and an excellent seminar programme.

Our research groups encourage a balance between senior and young scientists, creating the ideal environment to share and discuss research endeavours while supporting junior colleagues to develop and grow into new positions.

The EMBL Interdisciplinary Postdocs (EIPOD) programme builds on highly interactive research between units and is aimed at candidates whose research crosses scientific boundaries.

Please contact group leaders directly to find out if a position is available, or check **www.embl.de/jobs**

Why work here?

EMBL is an equal opportunity employer offering attractive conditions and benefits appropriate to an international research organisation. All employees benefit from excellent working conditions, a young and international atmosphere and a high-quality infrastructure of social services.

On-site childcare is available at some of EMBL's locations, helping staff to combine professional and family life.

Group and team leaders

EMBL fosters the pursuit of ambitious and longterm research projects at the highest level. Group and team leaders have the freedom to set their own scientific directions and are encouraged to explore the most challenging research areas.

Support for team and group leaders includes funding for a number of staff and laboratory space with equipment. Research collaborations between groups are an integral part of EMBL's scientific culture.

In addition to advanced scientific development, EMBL offers vocational training to improve skills in areas such as coaching, team management and communication. Establishing a good work-life balance is emphasised at every career stage.

Other careers

EMBL has ongoing opportunities for physicists, computer scientists and electronic engineers, especially early in their careers: Ever-more sophisticated analysis of very large data sets at the European Bioinformatics Institute (EMBL-EBI) draws on a skilled workforce from many disciplines: from scientific expertise in the life sciences to technical know-how in software development. Similarly, qualified technical staff are highly sought after to operate beamlines at EMBL's outstations in Hamburg and Grenoble.

Other positions include interface development, communications, user support, industry liason and training. We offer advanced scientific development and vocational training to improve skills in areas such as coaching and communication.


DxxG
RxYYxR
GxxY
KxxA
AxxK
KxxD
KxxR
RxxY
GxxN
YG
SxY
DxY
GxxI
YRRY
YxxR
DY
KA
GxxM
DxR
YxxQ
YS
RxxG
AxK
GxxR

Directors' Research

Directors' Research covers three thematically distinct research groups, headed by the Director General and Associate Director of EMBL, and the Director of the European Molecular Biology Organization (EMBO).

The Mattaj group investigates how, where, and at what points in the cell cycle the Ran GTPase controls cell division events like spindle or nuclear envelope (NE) assembly. Studies in the group have demonstrated that Ran's mitotic functions occur by the same mechanism as nucleo-cytoplasmic transport; investigated how RAN controls nuclear pore complex (NPC) assembly; and identified other protein and lipid kinases and phosphatases involved in regulating NE assembly. The group plans to build on these results to answer questions like why events such as NE assembly occur in telophase, rather than other times during mitosis, or how the different steps of spindle formation are regulated.

The Hentze group combines biochemical- and systems-level approaches to investigate the connections between gene expression, cell metabolism, and their role in human disease. Key goals of the group include collaborative efforts to: understand the basic mechanisms underlying protein synthesis and its regulation by miRNAs and RNA-binding proteins; explore, define and understand REM (RNA, Enzyme, Metabolite) networks; study the mechanisms and circuits that maintain iron homeostasis and its connection to the immune system; and understand more about the molecular pathways of iron overload, iron deficiency, and iron management disorders.

In investigating the mechanisms and forces that determine cell shape in *Drosophila*, the Leptin group studies two cell types. They look at how the cells at the tips of the fruit fly's tracheal system re-arrange their components as they grow rapidly and branch out to carry air to the animal's tissues. And at the tissue level, the group investigates how forces generated by single cells give the embryo's ventral furrow its final shape. The group also studies medaka and zebrafish to understand how signals from damaged cells are recognised by the innate immune system. They are developing methods to assay immune and stress responses in real time as the fish's cells encounter pathogens and stress signals.

The RanGTPase as a spatial regulator

Iain Mattaj

PhD 1979, University of Leeds.

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

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

The Mattaj group looks to understand how, where and at what points in the cell cycle the Ran GTPase controls cell division events like spindle or nuclear envelope assembly.

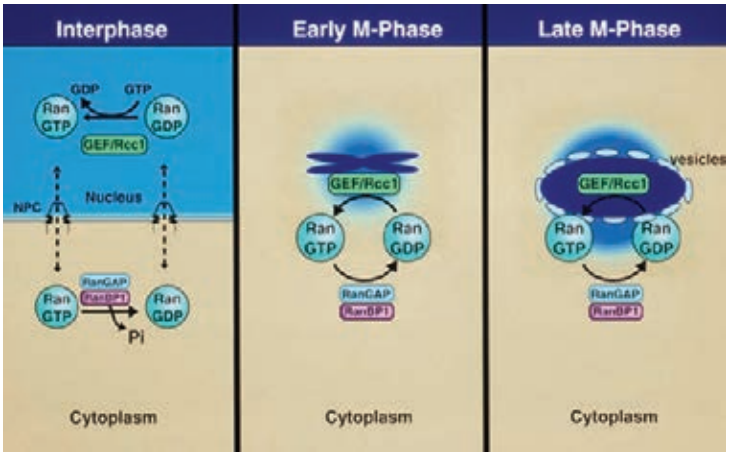
Previous and current research

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, and so on. 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

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

Gorjanacz, M. & Mattaj, I.W. (2009). Lipin is required for efficient breakdown of the nuclear envelope in *Caenorhabditis elegans*. *J. Cell Sci.*, 122, 1963-69

Yokoyama, H., Gruss, O.J., Rybina, S., Caudron, M., Schelder, M., Wilm, M., Mattaj, I.W. & Karsenti, E. (2008). Cdk11 is a RanGTP-dependent microtubule stabilization factor that regulates spindle assembly rate. *J. Cell Biol.*, 180, 867-75

Gorjanacz, M., Jaedicke, A. & Mattaj, I.W. (2007). What can *Caenorhabditis elegans* tell us about the nuclear envelope? *FEBS Lett.*, 581, 2794-801

Cytoplasmic gene regulation and molecular medicine

The Hentze group combines biochemical and systems level approaches to investigate the connections between gene expression, cell metabolism, and their role in human disease.

Matthias 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 of EMBL since 2005.

Selected references

Medenbach, J., Seiler, M. & Hentze, M.W. (2011). Translational Control via Protein-Regulated Upstream Open Reading Frames. *Cell*, 145, 902-913

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

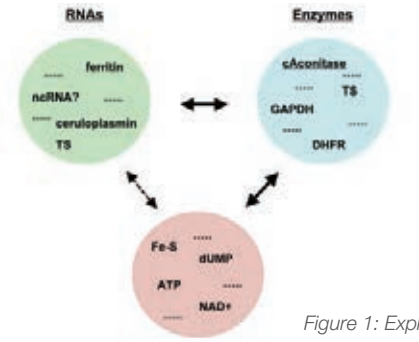


Figure 1: Exploring REM networks

Previous and current research

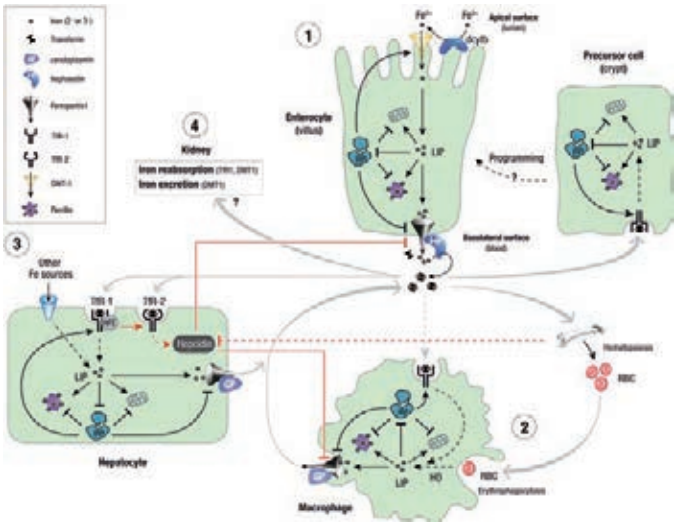
Important steps in the control of gene expression are executed in the cytoplasm by regulation of mRNAs via RNA-binding proteins (RBPs) and non-coding regulatory RNAs (e.g. miRNAs). We are elucidating these regulatory mechanisms, combining 'reductionist' biochemical and system's level approaches in mammalian, yeast and *Drosophila* model systems.

We recently developed 'mRNA interactome capture' to define 'all' RBPs associated with mRNAs *in vivo* (Castello et al., 2012). This work offers an ideal starting point for exploration 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 post-transcriptional processes of nonsense-mediated decay (NMD) and 3' end processing and their importance in genetic diseases (with Andreas Kulozik, Heidelberg University).

Our second major interest is the biology of mammalian iron metabolism (figure 2). This work includes the definition of the functions of the IRE/IRP regulatory network and its crosstalk with the iron hormone hepcidin. Within the MMPU (with Martina Muckenthaler, Heidelberg University), 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.

Figure 2: Systems biology of mammalian iron metabolism



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 underlying physiological iron homeostasis.
- 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: **www.embl.de/research/partnerships/mmpu**

Visualising complex cell shapes and signalling pathways

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 Max Planck Institute for Developmental Biology, Tübingen.

Full Professor, Univ. of Cologne, Institute of Genetics.

Director of EMBO & group leader at EMBL since 2010.

The Leptin group studies the mechanisms and forces that determine cell shape in Drosophila *and uses the zebrafish to analyse innate immune signalling.*

In vivo imaging of innate immune responses

The innate immune system provides rapid defence against pathogens and also deals with non-pathogenic stresses. Macrophages and dendritic cells, two key players in this system, patrol the body and respond to stimuli from damaged cells via extra- and intracellular sensors. We aim to understand how such signals are recognised and how the appropriate subcellular and intercellular responses are triggered. We have discovered that one family of sensors, the cytoplasmic NOD-like receptors (NLRs), are particularly abundant in fish.

The zebrafish and medaka model systems allow *in vivo* observation of physiological processes. Specifically, we can watch pathogens and the cells that attack them. By genetic and chemical engineering we will generate *in vivo* fluorescent reporters for immune signalling events. These will be used to assay immune and stress responses in real time and at high spatial and temporal resolution as the cells of the fish encounter pathogens and stress signals.

Cell shape determination during development

The shape of a developing organism is generated by the activities of its constituent cells: growth and proliferation, movements and shape changes. We are particularly interested in shape changes.

One study concerns an extremely complex single cell, the terminal cell of the *Drosophila* tracheal system. It is highly branched and carries air to target tissues through an intracellular tube bounded by plasma membrane. During its rapid growth, the cell faces the task of synthesising large amounts of membrane and sorting it correctly to the outer and inner membrane domains. Extensive re-organisation of the secretory organelles precedes membrane growth. The cytoskeleton, small GTPases, and polarity determinants direct the process.

In another project, we try to understand how the forces generated by individual cells are integrated within the supracellular organisation of a tissue to give the tissue its final shape. We study the formation of the ventral furrow in the early embryo, which is well understood in regard to its genetics and cell biology. The cells that form the furrow are the major force generators driving invagination, but to allow furrow formation, the neighbouring cells must respond and they may contribute.

New genes we are currently discovering in a genetic screen for mRNAs localised in the branches of tracheal cells will be used for two purposes: a bioinformatic study of the signals that guide mRNAs to their specific subcellular localisation and genetic and cell biological studies on how they contribute to branching and tube formation at that location. *In vivo* imaging with multi-colour probes will be used to analyse the cellular mechanisms. To understand force integration across many cell populations, we will use quasi-simultaneous time-lapse imaging of multiple-angle views of the gastrulating embryo. We will measure the specific shape changes in all the cells of the embryo. We use genetic and mechanical manipulations to reveal the underlying control circuits. These studies are complemented by computational modelling.

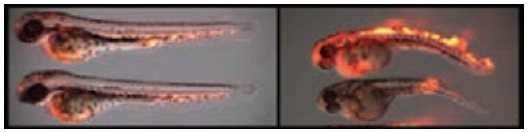
Selected references

Mathew, S.J., Rembold, M. & Leptin, M. (2011). A role for Traf4 in polarizing adherens junctions as a prerequisite for efficient cell shape changes. *Mol. Cell Biol.*, 31, 4978-93

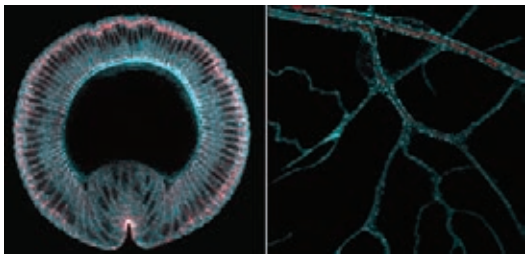
Jayanandanan, N., Gavis, E.R., Riechmann, V. & Leptin, M. (2011). A genetic *in vivo* system to detect asymmetrically distributed RNA. *EMBO Rep.*, 12, 1167-74.

Aggad, D., Stein, C., Sieger, D., Mazel, M., Boudinot, P., Herbomel, P., Levraud, J.P., Lutfalla, G. & Leptin, M. (2010). *In Vivo* Analysis of Ifn-Y1 and Ifn-Y2 Signaling in Zebrafish. *J. Immunol.*, 185, 6774-82

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, 384-6

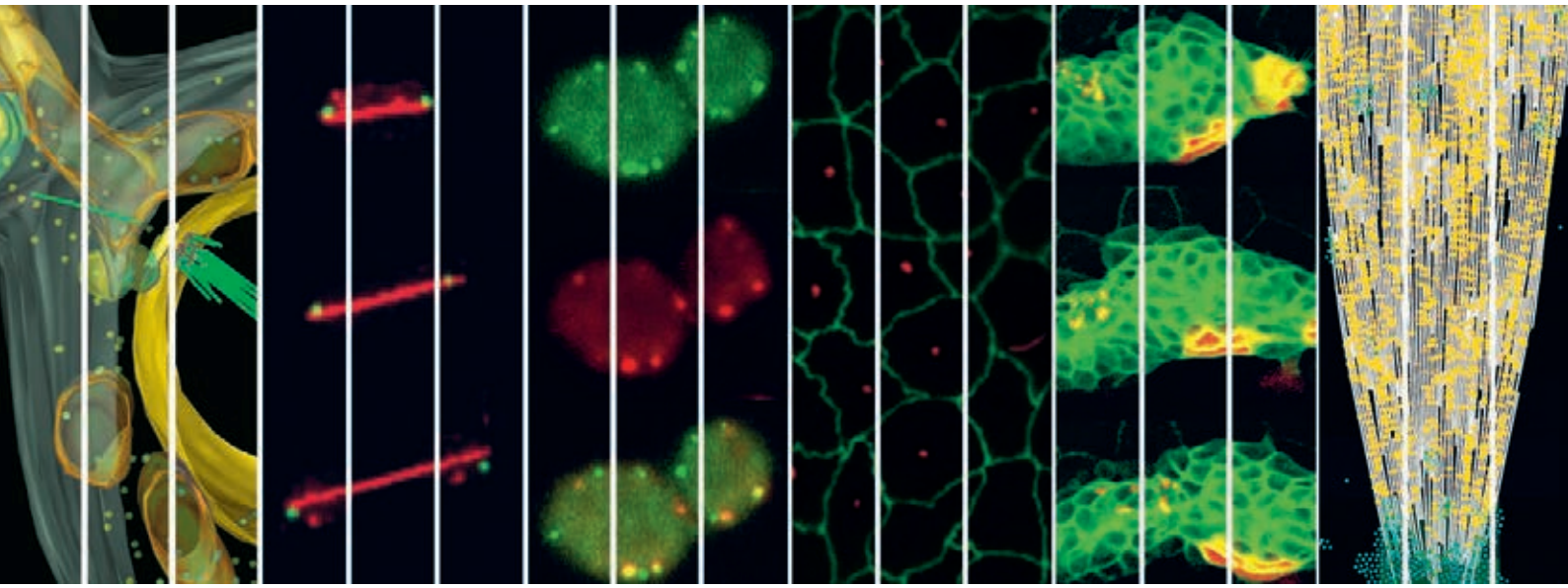


Zebrafish larvae 24 hours after infection with fluorescent bacteria. Normal fish survive and eventually clear the bacteria (left), but if the interferon signalling pathway is compromised (right) the bacteria proliferate and the fish die



Two systems to study cell shape. Left: cross section of a 3-hour old embryo in which the ventral cells are beginning to invaginate; about 100 cells. Right: In the single, highly branched terminal tracheal cell the ER (blue) is involved in delivering membrane to the cells outer and inner (red) plasma membrane


DxxG
RxYYxR
GxxY
KxxA
AxxK
RxxY
GXXN
YG
GxxM
DxY
YxxR
Gxxl
YRRY



Cell Biology and Biophysics

In the Unit, physicists and chemists work closely together with biologists to elucidate the fundamental rules that govern dynamic cell organisation and function. At the same time, groups are developing new instruments and technologies in order to reach this ambitious goal.

Cells are the smallest autonomous units of life and occupy the midpoint between the molecular and macroscopic scales. In order to understand how living systems are built and function, we need to understand the physical principles that underlie cellular organisation and function.

It is in the cell where we will first understand the basic processes of life at the molecular level in a physiological context. The cell provides the natural coordinate system in space and time onto which we have to map and integrate genomic, transcriptomic, proteomic, structural and biophysical information about the molecules that make up living systems. In one sentence, cell biology has become an integrative hub of much of modern biological research.

This is a time of tremendous opportunity for cell biology, but realising it also represents a formidable challenge and requires new concepts and approaches. Individual cellular processes, such as signalling, membrane trafficking, cytoskeletal dynamics, gene expression or cell division, can no longer be studied in isolation but need to be considered as integrated events. The default situation is that the molecular machinery that performs these functions is complex and combinatorial at the single protein, protein complex, and pathway level. This requires new ways of thinking about cellular functions that use network biology as standard tools in our repertoire and employ quantitative theoretical methods to generate mechanistic and predictive models that rely on realistic physical principles at the cellular, subcellular and molecular scale. Therefore, cell biology needs to integrate traditionally separate disciplines to realise its potential.

Novel developments in microscopy, computer simulations and chemical biology-based probes are a particular strength of the Unit. We constantly explore new directions to integrate new approaches and disciplines to answer cell biological questions. New correlative light/electron and superresolution imaging methods, as well as mechanistic biochemistry allow us to directly interface between cell and structural biology to understand molecular mechanisms. On the other hand, advances in live microscopy methods, now allow us to carry out cell biology in developing organisms to understand how cell organisation and collective cell behaviour leads to organ formation.

Mechanisms of cellular functions are often best understood when the organisation of the cell changes dramatically to carry out new functions. This is the case when cells divide, or when they change their fate. 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, when progenitor cells differentiate into new cell types the genetic programme is changed and a reorganisation of cellular architecture takes place, guided by rules that we begin to unravel. Understanding these rules and principles is our challenge for the years to come.

Jan Ellenberg
Head of the Cell Biology and Biophysics Unit

Systems biology of cell division and nuclear organisation

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.
Head of Cell Biology and Biophysics Unit since 2010.

The Ellenberg group studies how cells divide and organise in mitosis and meiosis, where errors can lead to problems such as cancer and infertility.

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 endoplasmic reticulum and perforated by nuclear pore complexes (NPCs). In M-phase, most metazoan cells reversibly disassemble the nucleus. Chromosomes are condensed, attached to cytoplasmic spindle microtubules, faithfully segregated and decondensed and the nucleus rapidly reassembles. Errors in this beautiful cycle of cell division can lead to severe consequences, such as cancer in somatic cells and infertility in gametes.

The overall aim of our research is therefore to systematically elucidate the mechanisms underlying cell division and nuclear organisation in mitosis and meiosis. To achieve this, we are developing advanced fluorescence microscopy-based methods to systematically identify the required proteins and study their function in living cells. 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 human cells for mitosis and cancer, and mouse oocytes for meiosis and infertility.

In the recent past we showed that mitotic nuclear breakdown and reformation is initiated by the ordered dis- and reassembly of NPCs, which interestingly use a different pathway to assemble during interphase. Furthermore, we demonstrated that meiotic spindle assembly and asymmetric positioning rely on novel mechanisms and that meiotic chromosome biorientation is highly error prone (figure 1). We have also identified hundreds of new cell division genes by screening the entire human genome using time-lapse microscopy (figure 2).

Future projects and goals

The objective of our future work is to gain comprehensive mechanistic insight into the division of human mitotic cells, to establish methods for systems analysis of the meiotic division of mammalian oocytes and to provide a biophysical basis to understand nuclear organisation.

For mitosis, we are aiming to achieve a systems level understanding of all proteins identified in our phenotypic profiling of the human genome. To this end we are continuously automating advanced fluorescence imaging techniques to study protein function in live cells in high throughput.

For oocyte meiosis, we are pursuing the molecular mechanism of homologous chromosome segregation and are developing new gene silencing methods and imaging systems to make this physiological model for infertility accessible for systems biology.

For nuclear organisation, we are currently focusing on the structure and assembly of the nuclear pore, the targeting and function of inner nuclear membrane proteins and chromosome architecture and dynamics.

Selected references

Kitajima, T.S., Ohsugi, M., and Ellenberg, J. (2011). Complete Kinetochore Tracking Reveals Error-Prone Homologous Chromosome Biorientation in Mammalian Oocytes. *Cell*, 146, 568-81

Conrad, C., Wunsche, A., Tan, T.H., Bulkescher, J., Sieckmann, F., Verissimo, F., Edelstein, A., Walter, T., Liebel, U., Pepperkok, R., and Ellenberg. J. (2011). Micropilot: automation of fluorescence microscopy-based imaging for systems biology. *Nat. Methods*, 8, 246-51

Dultz, E., and 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., Walter, T., Heriche J.-K., *et al*. (2010). Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature*, 464, 721-7

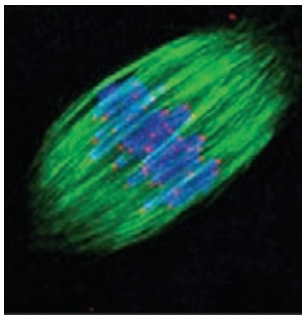


Figure 1: Meiotic spindle of a mouse oocyte. Chromosomes (blue) are biooriented by the meiotic spindle microtubules (green) at their kinetochores (red)

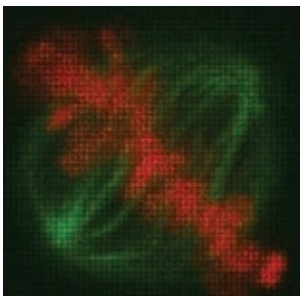


Figure 2: Mitotic spindle composed of thousands of microscopy images of human cells in which individual genes were silenced. Chromosomes (red) are made of images from genes that affect their segregation, while the mitotic spindle (green) is composed of images from genes affecting its assembly

Cellular electron tomography of cells

The Antony team uses electron tomography to investigate how the cell's microtubule scaffolding is organised.

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 Heidelberg since 2003.

Selected references

Roque H., Ward J.J., Murrells L., Brunner D. and Antony C. (2010). The Fission Yeast XMAP215 Homolog Dis1p Is Involved in Microtubule Bundle Organization. *PLoS One*, 5, e14201

Hoog, J.L., Huisman, S.M., Sebo-Lemke, Z., Sandblad, L., McIntosh, J.R., Antony, C. & Brunner, D. (2011). Electron tomography reveals a flared morphology on growing microtubule ends. *J. Cell Sci.*, 124, 693-698

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 Sci.*, 124, 978-87

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

Previous and current research

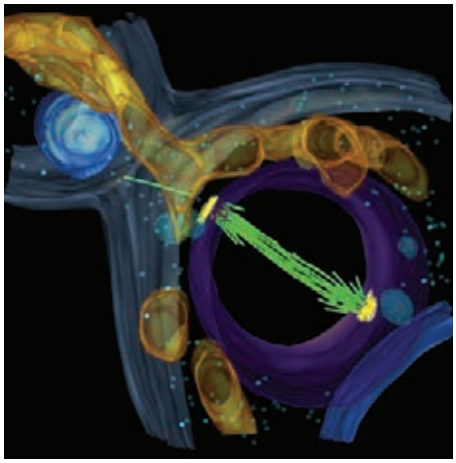
The focus of interest in the team is the organisation of microtubular cytoskeleton arrays in both fission yeast and budding yeast, as well as in the meiotic spindle. For this purpose we use electron tomography (ET), which allows the reconstruction, modelling and quantification of subcellular elements,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

Previously, we have studied microtubule (MT) arrays in fission yeast and tip mutants (called +TIPs, Mal3p (EB1) and tip1p (Clip-170) deletion mutants) and characterised the resulting MT organisational aberrant patterns, as well as the MT lattice structural defects (in collaboration with the Brunner group, formerly EMBL).

Currently, one major project involves studying the role and organisation of microtubules in the budding yeast mating pathway (in collaboration with the Knop (formerly EMBL) and Nédélec (page 27) groups). The project aims at deciphering the morphological and molecular characteristics of the karyogamy process in budding yeast. We focus on the nuclear congression process by resolving detailed microtubule organisation and their connection with the cellular organelles using an electron tomography approach. In parallel we have also been identifying the molecular basis responsible for the force generation that brings the two nuclei together. For this purpose we generated mutants that we analysed using live cell fluorescence microscopy approaches combined with the generation of 3D models derived for electron tomograms acquisition – this work is well advanced and will be submitted for publication in 2012.

Future projects and goals

In conjunction with internal and external research groups (Nédélec group (EMBL), Rebecca Heald (UCB) and Steffen Prohaska (Zuse Institute Berlin)) we run a major project aiming to reconstruct at EM resolution the *Xenopus laevis* meiotic 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 by joining 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 mid-zone of the spindle and at the poles. This research is supported by the Electron Microscopy Core Facility (page 70).



Mitosis in the budding yeast – 3D reconstruction of budding yeast cells during metaphase: The mitotic spindle is established between microtubules (green) emanating from each spindle pole body (yellow) (Picture by EMBL Predocs; Bar, 300 nm)

Multicellular morphogenesis

Darren Gilmour

PhD 1996, Cambridge University.

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

Group leader at EMBL Heidelberg since 2004.

Using the zebrafish as a model, the Gilmour group takes an integrative, multiscale approach to study how cells collectively migrate and assemble into functional organs.

Previous and current research

Collective behaviour lies at the heart of all biological design. Whether it is the assembly of proteins into complexes or the organisation of animal societies, collective interaction creates something much greater than the sum of the parts. What about cells? Cell biology has told us a great deal about how individual cells are organised but very little about how they form complex tissues and organs. This is because the standard culture systems that are the cell biologists' workhorse have been selected for uniformity, but they lack interesting collective behaviour.

By contrast, studies on developing embryos are revealing a picture where every decision a cell makes – from the genes it expresses to the shape it adopts – depends on dynamic interactions with other cells. And while studies on embryos provide a more complex view of cell and tissue morphogenesis – where heterogeneous cells shape each other through dynamic interactions – this approach provides a number of opportunities for cell biology. For example, using embryonic systems we can hopefully understand how cellular organisation feedbacks on the genome to drive differentiation. Moreover, a precise understanding of how cells organise each other could accelerate the use of tissue engineering approaches in human healthcare.

We are taking an integrative, multiscale approach to study how cells collectively migrate and assemble into functional organs, using the zebrafish lateral line organ as a model. Here, a migrating epithelial primordium comprising of 100 cells assembles and deposits a series of rosette-like mechanosensory organs. We chose it for a number of reasons: It is a complete organogenesis process that takes place on a remarkably small spatiotemporal scale; its superficial migration route, beneath a single transparent cell layer, makes it the dream *in vivo* sample for quantitative imaging approaches; genetic screens have identified regulators of its behaviour that are of great interest due to their role in human disease – for example, it is guided by Cxcr4/SDF1 signalling, a chemokine-receptor pair known to control many human cancers.

Future projects and goals

We have developed *in vivo* imaging and analysis tools that allow an entire morphogenesis process to be addressed at different spatiotemporal scales. By integrating these data using statistical multiplexing methods, we are able to unequivocally define the relationship between different tissue behaviours (such as motility and shape) and explain these at the level of underlying machinery (such as actin dynamics and chemokine signalling). Such correlations are subsequently validated using acute perturbation experiments and the data are combined using computational modelling approaches. As much of what we find is likely to be applicable in other contexts, we hope to move towards a systems-level understanding of the interplay between gene activity, cell organisation and tissue mechanics during tissue morphogenesis.

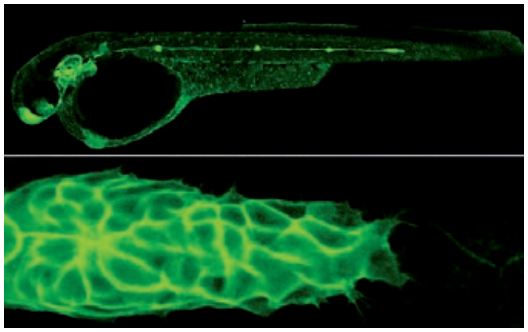
Selected references

Streichan, S.J., Valentin, G., Gilmour, D., Hufnagel, L. (2011). Collective cell migration guided by dynamically maintained gradients. *Phys. Biol*. 8, 045004

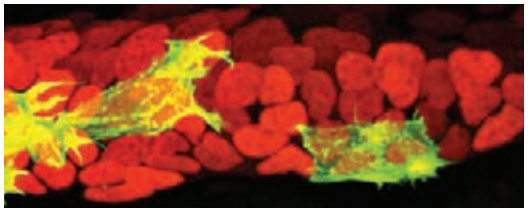
Friedl, P., Gilmour, D. (2009). Collective cell migration in morphogenesis, regeneration and cancer. *Nat. Rev. Mol. Cell Biol.,* 10, 445-57.

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

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



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



Visualising actin dynamics (LifeAct-GFP) within migrating primordium

Chromosome structure and dynamics

The Häring group aims to understand the molecular machinery that organises chromosomes to allow their correct distribution among daughter cells.

Christian Häring

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

Selected references

Cuylen, S., Metz, J. & Haering, C.H. (2011). Condensin structures chromosomal DNA through topological links. *Nat. Struct. Mol. Biol.*,18, 894-901

Rutkowska, A., Haering, C.H. & Schultz, C. (2011). A FlAsH-based cross-linker to study protein interactions in living cells. *Angew Chem. Int. Ed.*, 50, 12655-8

Cuylen S, Haering CH. (2011). Deciphering condensin action during chromosome segregation. *Trends Cell Biol*. 21, 552-9

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

Previous and current research

Eukaryotic chromosomes undergo enormous changes in structure and organisation over the course of a cell cycle. One of the most fascinating is the transformation of interphase chromatin into individualised rod-shaped mitotic chromosomes in preparation for cell division. This process, known as chromosome condensation, is a key step for successful chromosome segregation, yet the underlying mechanisms are still poorly understood.

The overall aim of our research is to unravel the action of molecular machines that organise chromosomes in order to allow their correct segregation. Insights into these mechanisms will be of great importance to our understanding of how cells inherit a complete set of their genomes every time they divide and prevent the occurrence of aneuploidies, which are hallmarks of most cancer cells and the leading cause of spontaneous miscarriages in humans.

One of the key players in the formation of mitotic chromosomes is a highly conserved multi-subunit protein complex, known as condensin. We have recently shown that condensin binds to mitotic chromosomes by encircling chromosomal DNA within a large ring structure formed by the structural maintenance of chromosomes (SMC) and kleisin subunits. Our working hypothesis is that condensin uses this topological principle to tie together loops of chromatin and thereby maintain mitotic chromosomes in their characteristic shape (figure 1).

Future projects and goals

We are using an interdisciplinary approach to advance our understanding of condensin function in yeast and mammalian cells by combining biochemical, molecular, and cell biological methods. In collaboration with other groups at EMBL, we are taking further advantage of structural and chemical biological techniques to discover how condensin loads onto chromosomes, how it interacts with other chromosomal components and how its activity is controlled.

In an independent project, we are exploring novel approaches to identify additional players that direct the formation of mitotic and meiotic chromosomes, using fission yeast genetics and advanced time-resolved light microscopy methods (figure 2).

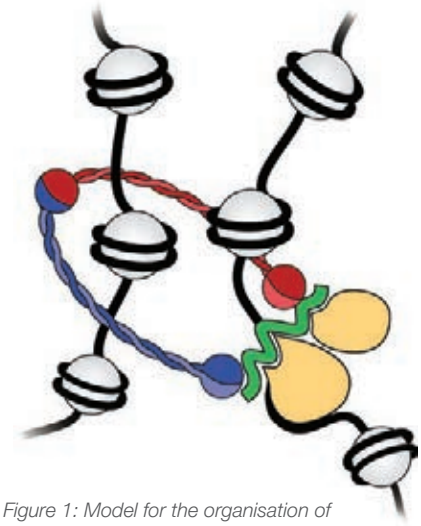


Figure 1: Model for the organisation of mitotic chromosomes by condensin rings

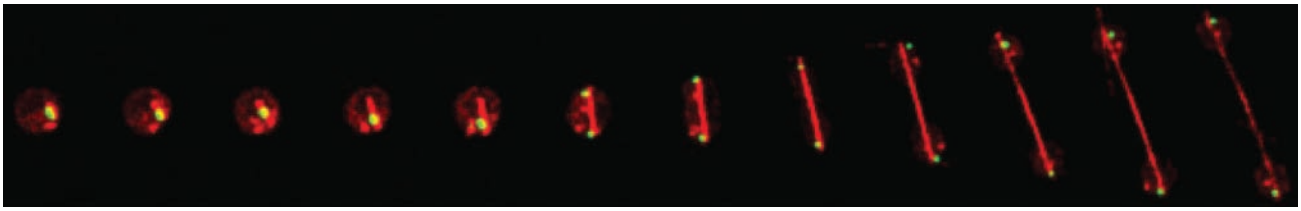


Figure 2: Monitoring chromosome structure and segregation in fission yeast cells

Dynamics of cell growth and tissue architecture

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.

The Hufnagel group studies the role of mechanical constraints on processes such as cell growth, programmed cell death, orientation of division, intra-tissue rearrangements and cell differentiation.

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.

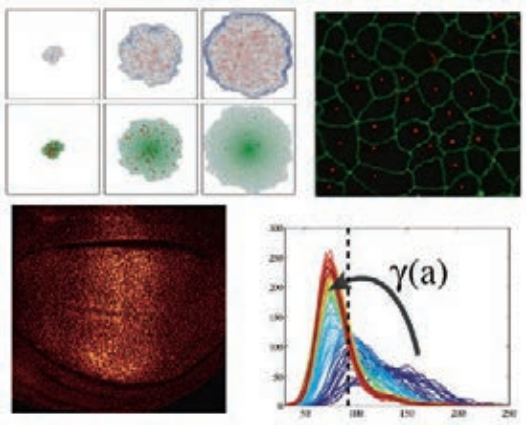
Selected references

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, 3835-40

Brockmann, D., Hufnagel, L. & Geisel, T. (2006). The scaling laws of human travel. *Nature*, 439, 462-65.

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, 512-22.



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

Dynamics of membrane trafficking

Using budding yeast as a model, the Kaksonen group wants to understand how the molecular machinery interacts to drive processes such as vesicle trafficking.

Marko Kaksonen

PhD 2002, University of Helsinki.

Postdoctoral research at the University of California, Berkeley.

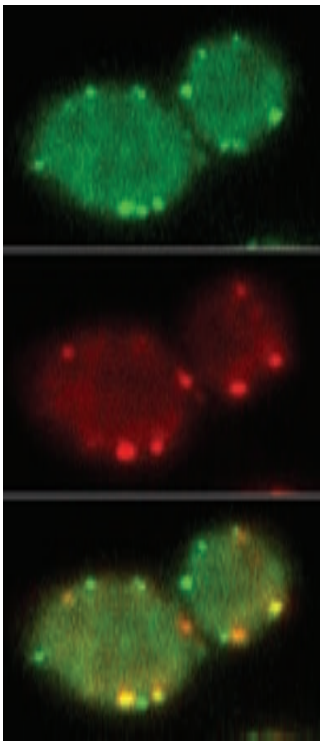
Group leader at EMBL Heidelberg since 2006.

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., Betts, M.J., Gvozdenovic-Jeremic, J., *et al.* (2010). A systematic screen for protein-lipid interactions in *Saccharomyces cerevisiae*. *Mol. Syst. Biol.*, 6, 430



A yeast cell expressing fluorescently labelled endocytic proteins. The first two images show a coat protein Sla1 (green) and an actin-binding protein Abp1 (red). The last image shows both channels merged. The spots at the cell surface reveal the transient accumulation of the proteins at endocytic sites during vesicle budding

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 signalling, cell motility, cytokinesis, and vesicle trafficking. Our laboratory's long-term goal is to contribute to the understanding of the general principles governing the assembly and function of these supramolecular machines.

More specifically, we are interested in the formation of cargo-loaded transport vesicles, such as endocytic vesicles. The formation of the endocytic vesicle is driven by 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, the recruitment of the cargo molecules, and the selection of the location and timing of endocytic events in the cell.

Our main experimental organism is the 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 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 learn 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. We have also started to investigate the evolution of the membrane-actin coupling in animals and fungi using a phylogenetic comparative approach.

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.

Cytoskeletal dynamics and function in oocytes

Péter Lénárt

PhD 2004, EMBL and University of Heidelberg.

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

Staff scientist at EMBL Heidelberg since 2008. Group leader since 2011.

Using starfish as a model organism, the Lénárt group combines biochemistry with imaging assays to investigate how the fertilisable egg cell develops from the oocyte.

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 organise cellular architecture in a very large cytoplasm. How the cytoskeleton adapts to this unusual size, and how these mechanisms differ 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, develop simply in seawater at room temperature, and are transparent – ideal for high-resolution imaging of cytoskeletal dynamics in live cells. 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 were able to 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

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 will employ high-resolution imaging methods, including electron tomography and super-resolution light microscopy, to resolve single actin-filaments and to 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.

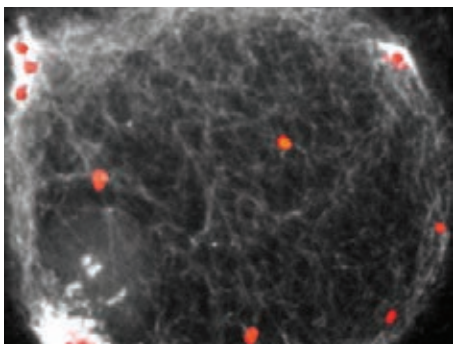
Selected references

Field, C.M. & Lénárt, P. (2011). Bulk cytoplasmic actin and its functions in meiosis and mitosis. *Curr. Biol.*, 21, R825-R830

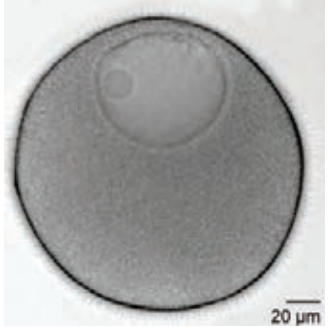
Mori, M., Monnier, N., Daigle, N., Bathe, M., Ellenberg, J. & Lénárt, P. (2011). Intracellular transport by an anchored homogeneously contracting F-actin meshwork. *Curr. Biol.*, 21, 606-11

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 insights 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, 812-18



The actin filament network (grey) embedding the chromosomes (red)



Transparent starfish oocytes are uniquely suited for imaging meiotic divisions

Cellular architecture

The Nédélec group develops in vitro *experiments and modelling tools to explore complex intracellular processes, such as mitosis.*

François Nédélec

PhD 1998, Université Paris-Sud II.

Postdoctoral research at EMBL.
Group leader since 2005.
Joint appointment with the Structural and Computational Biology Unit.

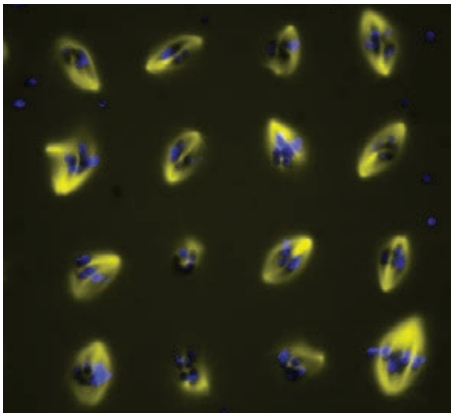
Selected references

Loughlin, R., Heald, R. & Nédélec, F. (2010). A computational model predicts *Xenopus* meiotic spindle organization. *J. Cell Biol.*, 191, 1239-49

Dinarina, A., Pugieux, C., Corral, M.M., Loose, M., Spatz, J., Karsenti, E. & Nédélec, F. (2009). Chromatin shapes the mitotic spindle. *Cell*, 138, 502-13

Jekely, G., Colombelli, J., Hausen, H., Guy, K., Stelzer, E.H.K., Nédélec, F. & Arendt, D. (2008). Mechanism of phototaxis in marine zooplankton. *Nature*, 456, 395-99

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



An array of mitotic spindles obtained in vitro with Xenopus laevis *egg extracts (Dinarina* et al.*,* Cell*, 2009)*

The metaphase spindle, a dynamic bipolar structure of filaments called microtubules (white) that are connected by molecular motors (orange). This simulation elucidates how a spindle can remain stable for hours, even though it is made of filaments that individually exist for less than a minute (Loughlin et al*. 2010)*

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, except for the chromosomes, all the components of a spindle — polar filaments called microtubules and associated proteins — are in rapid turnover. Microtubules grow, shrink and disappear in less than a minute and their associated proteins continuously and stochastically bind and unbind even faster. The resulting assembly although highly dynamic is remarkably precise: it can remain steady for hours waiting for the right signal, to eventually apply 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 seems within our reach, as testified by the rise of systems biology. Yet, collective behaviours in biology require more than statistical averages. They are challenging us for many reasons: 1) the diversity of molecular players is enormous; 2) their interactions are often dynamic and out-of-equilibrium, and 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. We develop an advanced simulation engine – called cytosim – to simulate ensembles of multiple polar fibers and associated proteins. Cytosim can simulate various problems involving microtubules, actin filaments or both. 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.

Systems biology of stem cell differentiation

Pierre Neveu

PhD 2007, Ecole Normale Supérieure, Paris.

Postdoctoral research at the Kavli Institute for Theoretical Physics and the Neuroscience Research Institute, Santa Barbara.

Group leader at EMBL Heidelberg since 2011.

The Neveu group takes an integrated systems biology approach to investigate the molecular changes that determine what a stem cell becomes.

Previous and current research

Pluripotent cells have the dual ability to self-renew and differentiate. Therefore, in pluripotent cells, the expression of hundreds of genes should be stable in the self-renewal case but gene expression can also be directed in a coordinated manner towards particular states upon external signaling cues (lineage commitment towards terminal differentiation). Deciphering this complex problem has garnered much attention at the systems level.

Tackling this challenge requires good characterisation of the pluripotent state. miRNAs are suitable marker candidates because they are excellent classifiers of tissue types or cellular states and they also play a crucial role in differentiation. By profiling miRNA expression in human cells, we have previously shown that pluripotency surprisingly emerges as a much more diverse state than previously believed: variability in miRNA expression is comparable to the one found in differentiated cells and cancer cells. We have also shown that it is possible to dramatically reduce the complexity of miRNA expression patterns to a few meaningful dimensions. This reductionist approach still allows us to quantitatively and robustly discriminate pluripotency, cancer and lineage commitment. More importantly it suggests that complex processes of the stem cell system, such as differentiation and reprogramming, can be mapped quantitatively.

Currently, we are employing a dynamic approach at the single cell level to resolve the dynamics of differentiation and the different molecular and cellular processes at play during fate determination. Indeed differentiation is intrinsically a dynamic process, where individual cells have to change from one state to another. Having developed fluorescent reporters to assess miRNA expression in single cells, we are characterising mouse ESC self-renewal using single cell live imaging.

Future projects and goals

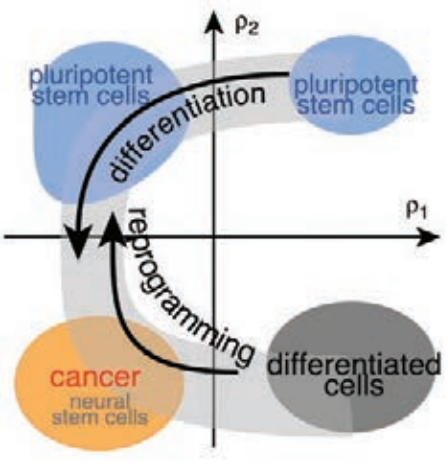
We plan to study the dynamics of differentiation at the single cell level both *in vitro* in mouse embryonic stem cells and *in vivo*. The ultimate goal is to dissect the transcriptional regulation and gene networks and the associated cellular changes underlying stem cell differentiation. We are taking an integrated systems biology approach that combines single cell live imaging of miRNA expression, image processing, perturbation approaches and mathematical modelling.

We wish to address the following questions:

- How dynamic is the pluripotent state?
- What are the *in vitro* dynamics of differentiation of mouse ESCs?
- How do *in vitro* findings compare to *in vivo* differentiation behaviour?

Selected reference

Neveu, P., Kye, M.J., Qi, S., Buchholz, D.E., Clegg, D.O., Sahin, M., Park, I.H., Kim, K.S., Daley, G.Q., Kornblum, H.I., Shraiman, B.I. & Kosik, K.S. (2010). MicroRNA profiling reveals two distinct p53-related human pluripotent stem cell states. *Cell Stem Cell*, 7, 671-81



Molecular cartography of stem cells: miRNA expression classifies pluripotent cells, cancer cells and differentiated cells. This map allows us to quantitatively follow changes in cell identity such as differentiation and reprogramming. It reveals that reprogramming involves a cancer-like behaviour

Membrane traffic in the early secretory pathway

The Pepperkok team develops novel approaches to investigate interactions between the endoplasmic reticulum and the Golgi complex.

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 Heidelberg since 1998.

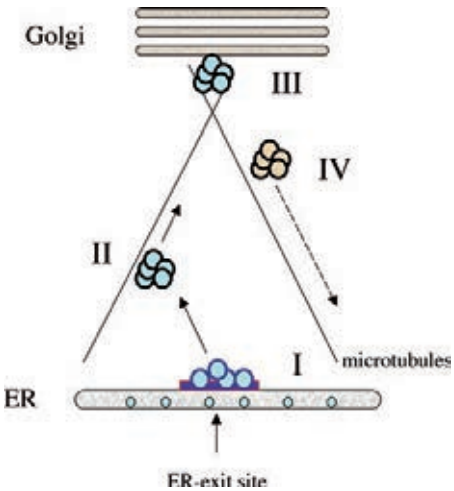
Selected references

Tangemo, C., Ronchi, P., Colombelli, J., Haselmann, U., Simpson, J.C., Antony, C., Stelzer, E.H., Pepperkok, R. & Reynaud, E.G. (2011). A novel laser nanosurgery approach supports *de novo* Golgi biogenesis in mammalian cells. *J. Cell Sci.*, 124, 978-87

Conrad, C., *et al*. (2011). Micropilot: automation of fluorescence microscopy-based imaging for systems biology. *Nat Methods*, 8, 246–49

Bartz, F., *et al*. (2009). Identification of cholesterol-regulating genes by targeted RNAi screening. *Cell Metab*, 10, 63-75

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



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.

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). 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.

Chemical cell biology

Carsten Schultz

PhD 1989, University of Bremen.

Postdoctoral research at the University of California, San Diego.

Habilitation 1997, Organic Chemistry, University of Bremen.

Group leader, MPI for Mol. Physiology, Dortmund.

Group leader at EMBL Heidelberg since 2001.

The Schultz group develops tools for imaging and for manipulating cellular enzyme activities, with a particular emphasis on the hereditary disease cystic fibrosis.

Previous and current research

Past projects: Before joining EMBL in 2001, our research focused on finding novel ways to stimulate chloride and water secretion of epithelial cells in understanding 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. Of particular significance was the development of chemical methods to convert highly polar signalling molecules like cyclic nucleotides, inositol phosphates and phosphoinositides to membrane-permeant, bioactivatable derivatives ('prodrugs') (Schultz, 2003; Laketa *et al*., 2009).

Current projects: Our interest in CF has shifted to the development of lung emphysema (the ultimate cause of death in the patient). In a truly translational collaboration with the Mall group (Molecular Medicine Partnership Unit), we developed FRET reporters to sense enzyme activities detrimental to lung tissue such as macrophage and neutrophil elastases. In *ex vivo* experiments, we are now able to monitor these enzyme activities on cells from both mouse models and patients (Cobos *et al*., 2009; Gehrig *et al*., submitted). At the cell biology level, our interest shifted to basic signalling networks regulated by G-protein-coupled as well as growth factor receptors. We developed a wide range of fluorescent reporter molecules either genetically encoded (Piljić & Schultz, 2011) or as small molecule fluorescent probes (see figure). We hope to provide a more complete picture of the signalling network and to help find compounds beneficial in unravelling basic principles in signal transduction and, ultimately, in ion and enzyme secretion relevant to CF patients. In addition, we prepared a large number of tools to manipulate signalling networks and are able to locally activate the important second messenger PIP3 and other signalling lipids with a light flash in subcellular resolution in living cells (Mentel *et al*., 2011). Alternatively, we switch on enzymes such as single G-proteins by translocating them to their site of action with the help of a chemical dimeriser (Putyrski *et al*., 2011).

Hot projects: Currently, we are very excited about performing bioorthogonal chemistry inside living cells. In collaboration with the Lemke group (page 62), we developed a new set of amino acids that can be site-specifically incorporated into a protein of interest by amber stop codon suppression and then labelled *in vivo* with a fluorogenic compound. This provides access to labelling with high quality dyes and minimal disruption of the protein structure (Plass *et al*., 2011; Plass *et al*., submitted). In collaboration with the Häring group (page 23), we developed a new method to cross-link proteins in a protein-protein interaction-dependent fashion in living cells by using FlAsH technology (Rutkowska *et al*., 2011).

Future projects and goals

In 2012, we will focus predominantly on lipid signalling and lipid-controlled cell biology and examine the effect of phospholipids on endocytosis and lipid trafficking in collaboration with the Pepperkok group (page 29). Most projects rely on organic chemistry and the group 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.

Selected references

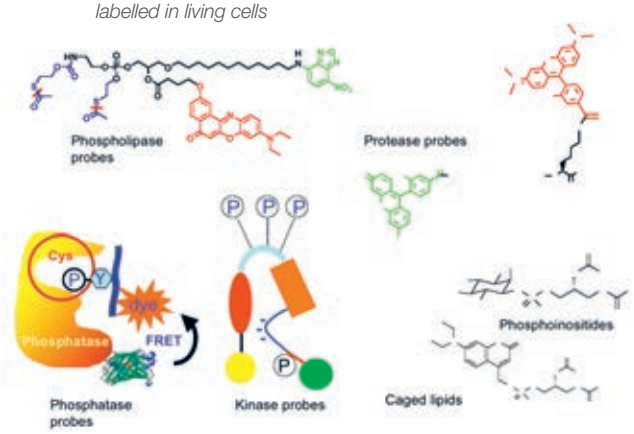
Rutkowska, A., Haering, C.H. & Schultz, C. (2011). A FlAsH-based cross-linker to study protein interactions in living cells. *Angew Chem. Int. Ed. Engl.*, 50, 12655-58

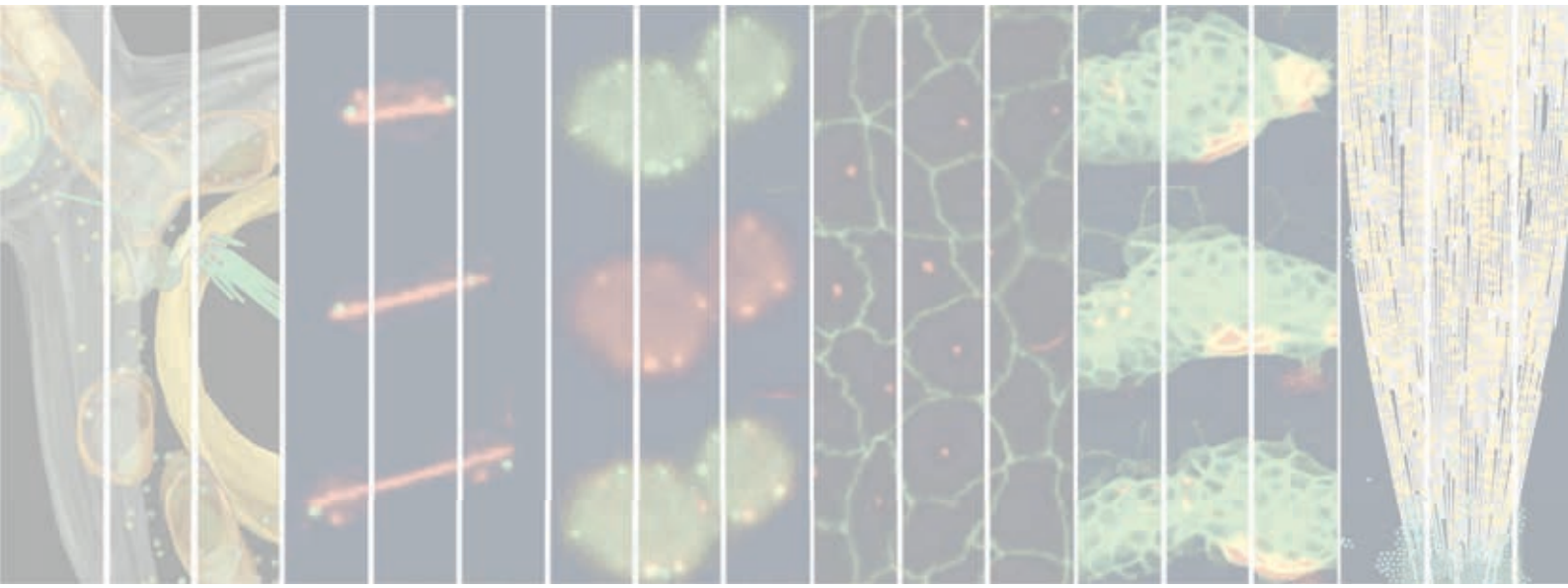
Putyrski, M. & Schultz, C. (2011). Switching heterotrimeric G protein subunits with a chemical dimerizer. *Chem. Biol*, 18, 1126-33

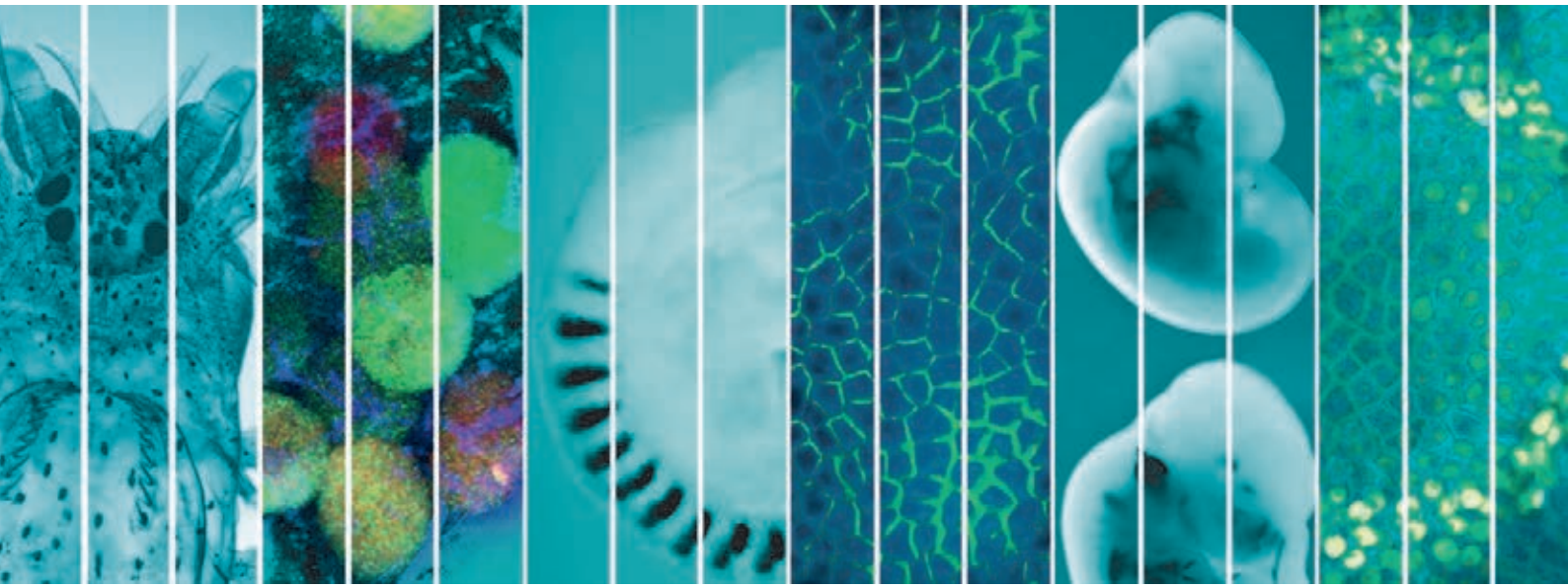
Plass, T., Milles, S., Koehler, C., Schultz, C. & Lemke, E.A. (2011). Genetically encoded copper-free click chemistry. *Angew Chem. Int. Ed. Engl.*, 50, 3878-81

Mentel, M., Laketa, V., Subramanian, D., Gillandt, H. & Schultz, C. (2011). Photoactivatable and cell-membrane-permeable phosphatidylinositol 3,4,5-trisphosphate. *Angew Chem. Int. Ed. Engl.*, 50, 3811-14

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

The development of living organisms requires precise coordination of all basic cellular processes, in space and time. Groups seek to elucidate the principles, mechanisms and dynamics of fundamental developmental events. Using animal and plant models, research in the Unit integrates numerous complementary approaches to understand how cellular and morphological processes are coordinated and evolve to shape and maintain living organisms in their environment.

A fundamental question in developmental biology is the mechanism by which symmetry is broken and cells with distinct fates are specified. In many organisms embryonic development begins before the onset of zygotic transcription, under the control of mRNAs and proteins localised asymmetrically in the egg. Cell polarity thus underlies embryonic asymmetry. Mechanisms underlying cell polarisation, mRNA transport, and translational control in *Drosophila* are under investigation in the unit. In plants, the polarised transport of auxin determines the positioning of lateral organs; and how auxin specifies different cell types in *Arabidopsis* is another topic of research. In mammals, in which polarity is absent in the egg, symmetry is broken during early embryogenesis, when stochastic processes may be involved in generating cellular heterogeneity. A systems-level understanding of the symmetry breaking processes operating in the early mouse embryo is another aim.

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 goal. Using the mouse model, the mechanisms controlling overall developmental rate at an organismal level, as well as the timing of individual patterning processes and 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 major remaining mysteries in animal evolution: the evolution of the central nervous system.

Several groups 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

Cell polarity and RNA localisation

Anne Ephrussi

PhD 1985, Massachusetts Institute of Technology.

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

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

The Ephrussi group seeks to understand the mechanisms regulating basic cellular processes in a developmental context, in the fruit fly.

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.

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.

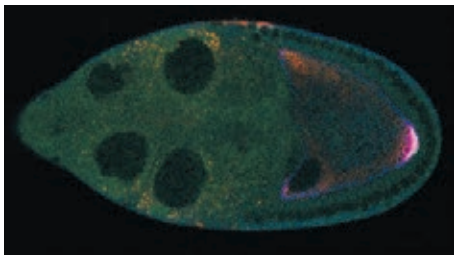
Selected references

Jambor, H., Brunel, C. & Ephrussi, A. (2011). Dimerization of *oskar* 3' UTRs promotes hitchhiking for RNA localization in the *Drosophila* oocyte. *RNA*, 17, 2049-57

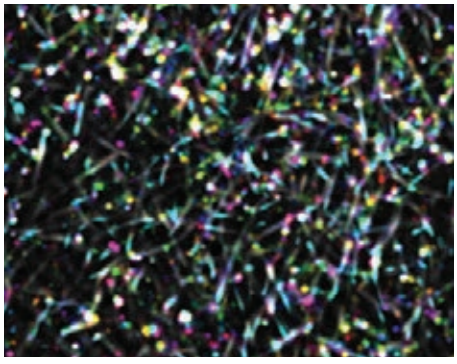
Krauss, J., Lopez de Quinto, S., Nusslein-Volhard, C. & Ephrussi, A. (2009). Myosin-V regulates *oskar* mRNA localization in the *Drosophila* oocyte. *Curr. Biol.*, 19, 1058-63

Chekulaeva, M., Hentze, M.W. & Ephrussi, A. (2006). Bruno acts as a dual repressor of *oskar* translation, promoting mRNA oligomerization and formation of silencing particles. *Cell*, 124, 521-33

Hachet, O. & Ephrussi, A. (2004). Splicing of *oskar* RNA in the nucleus is coupled to its cytoplasmic localization. *Nature*, 428, 959-63



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



Right: 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 grey with cyan tips, indicating plus ends) to take fast, long linear runs*

Evolution of the nervous system in Bilateria

By studying and comparing simple marine organisms, the Arendt group looks to understand the origin and evolution of our central nervous system.

Detlev Arendt

PhD 1999, Albert-Ludwigs-Universität, Freiburg.

Team leader at EMBL since 2002.
Group leader and senior scientist since 2007.
Academic mentor, postdoctoral training since 2007.
ERC investigator since 2012.

Selected references

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

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

Arendt, D., Hausen, H. & Purschke, G. (2009). The 'division of labour' model of eye evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 364, 2809-17

Jekely, G., Colombelli, J., Hausen, H., Guy, K., Stelzer, E.H.K., Nedelec, F. & Arendt, D. (2008). Mechanism of phototaxis in marine zooplankton. *Nature*, 456, 395-99

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 being? What did it first look like 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*. Genomic resources and molecular techniques have been generated that make it a model marine invertebrate for ocean biology and for organismal systems biology. *Platynereis* is amenable to high-throughput imaging techniques and functional interference approaches (first genetic knock-out lines have been generated). With the recent development of the PrImR (Profiling by Image Registration) resource, it is the first animal model for which gene expression profiling data can be obtained in cellular resolution for the whole organism. 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 – findings that revolutionise our understanding of brain evolution. The clear picture is emerging that the *Platynereis* brain harbours many cell types so far known only for the vertebrates, but in a much more simple and different overall arrangement.

To broaden our comparative approach, we study two other model species (amphioxus and *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 nervous system evolution.

Future projects and goals

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 sea anemone and amphioxus. 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 Centre, we will 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 (early developmental as well as differentiation stages).

In 2012, we are starting an ERC-funded project, 'BrainEvoDevo', to generate a neuron type atlas of the annelid larval brain, combining neuronal morphologies, axonal projections and cellular expression profiling for an entire bilaterian brain. Building on the Atlas, we will dissect *Platynereis* chemosensory-motor forebrain circuits, by laser ablation of GFP-labelled single neurons, gene knockout studies and behavioural assays based on microfluidics. Our aim is to explore duplication, divergence and expansion of neural circuits in CNS development and evolution.

We are also interested in exploring population genetics and the variability of development and differention in different habitats and we are collecting strains of *Platynereis* and amphioxus as part of the TARA Oceans expedition and EMBL oceans team.

As a 'living fossil', Platynereis *represents an ideal connecting link between vertebrates and the fast evolving protostome models,* Drosophila *and* Caenorhabditis

Timing of mammalian embryogenesis

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.

The Aulehla group studies how the precise timing and sequence of events that unfold as an embryo develops are controlled.

Previous and current research

During an embryo's journey from a single cell to a complex organism, countless patterning processes unfold with remarkable precision, both spatially but also in respect to their temporal sequence, or timing. 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 intrinsic and extrinsic factors responsible for controlling these oscillations within the segmentation process. The insight gained from studying this specific oscillation phenomenon will be combined with our efforts that address the mechanisms that control the overall timing of development.

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:

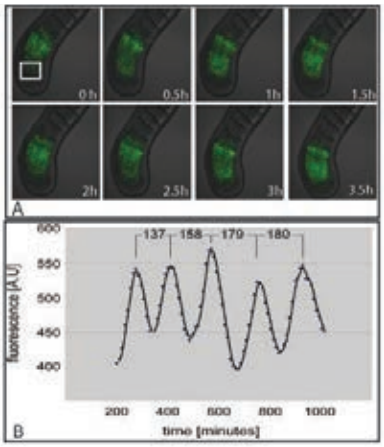
- identification of extrinsic and intrinsic signals controlling the timing of mouse development
- generation of a real-time imaging reporter system for Wnt-signalling oscil lations in mouse embryos using embryonic stem cell technology
- analysis of the mechanisms underlying Wnt-signalling oscillations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

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



(A) Data from two-photon real-time imaging experiments performed in a transgenic reporter mouse embryo. (B) Quantification of fluorescent signal within the tail region (white box in A) identifies striking oscillations

In situ hybridisation of mouse embryo at day nine of development

Membrane dynamics during tissue morphogenesis and differentiation

The de Renzis group investigates how the machinery that controls trafficking within cells is reorganised as tissues form, and how that affects embryonic development.

Stefano de Renzis

MD 1997, University Federico II, Naples.
PhD 2002, EMBL Heidelberg.
Postdoctoral work at Princeton University.
Group leader at EMBL Heidelberg since 2008.

Selected references

Tang, D., Xiang, Y., De Renzis, S., Rink, J., Zheng, G., Zerial, M. & Wang, Y. (2011). The ubiquitin ligase HACE1 regulates Golgi membrane dynamics during the cell cycle. *Nat. Commun.*, 2, 501

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. (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

Previous and current research

Our research is focused on understanding how machineries controlling intracellular trafficking are re-organised during cell and tissue morphogenesis and how this, in turn, impacts on specific cell and tissue behavior during embryonic development. We address these questions using the early *Drosophila* embryo as model system during cellularisation and early gastrulation stages.

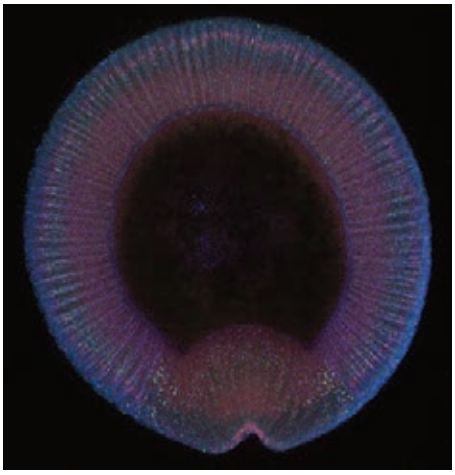
Cellularisation of the *Drosophila* embryo provides an excellent system to study mechanisms linking membrane trafficking and cell/tissue morphogenesis. It takes around an hour for a syncytium of ~6000 nuclei to complete the process of cellularisation, a particular form of cytokinesis involving a massive mobilisation of intracellular membranes. The end result of this process is the formation of a fully polarised ephitelium. Concomitantly, the embryo undergoes extensive remodelling of gene expression characterised by the activation of zygotic transcription. 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 have 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 polarised activation of Notch trafficking in the early embryo (see figure). We are now combining high-resolution imaging methods to follow the spatio-temporal modulation of trafficking pathways in live *Drosophila* embryos at the subcellular scale.

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 during morphogenesis. Our long-term goal is to analyse the differentiation of intracellular pathways in other cell types and tissues. We wish to elucidate how machineries controlling intracellular trafficking reorient during differentiation and how this in turn 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 signalling receptor Notch is endocytosed (green dots) together with its ligand Delta (blue dots) specifically in cells undergoing invagination (ventral furrow formation).*



Developmental patterning in plants

Marcus Heisler

PhD 2000, Monash University, Australia.

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

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

Group leader at EMBL since 2009.

Using A. thaliana *as a model, the Heisler group seeks to understand patterning in plant development and how it is established and regulated.*

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.

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 we are 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 patterning processes associated with dorsiventral cell-type specification. Our main questions 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. 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:** Our goal is to integrate single-cell transcriptomics 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.

Selected references

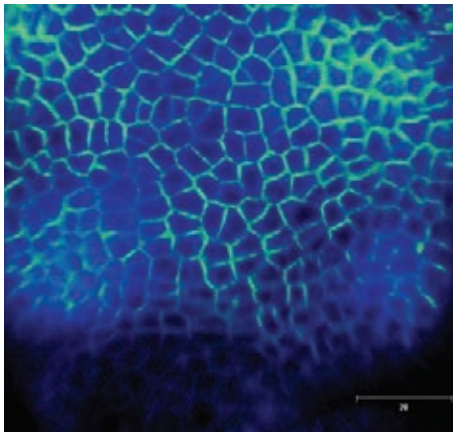
Heisler, M.G., Hamant, O., Krupinski, P., *et al*. (2010) Alignment between PIN1 polarity and microtubule orientation in the shoot apical meristem reveals a tight coupling between morphogenesis and auxin transport. *PLoS Biol*., 8, e1000516

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, 1650-55

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

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, 1633-38

Confocal projection showing polar localisation of the auxin efflux carrier PIN1 fused to GFP. At organ inception PIN1 polarities are directed away from adjacent organ sites and towards the new site



Systems-level understanding of early mammalian development

Looking at the molecular, cellular and systems levels, the Hiiragi group studies how, early in mammal development, the embryo is shaped from a spherical mass of cells.

Takashi Hiiragi

PhD 2000 Kyoto University, Japan.

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

Group leader at the MPI of Immunobiology 2002-2007.

Independent group leader at the MPI for Molecular Biomedicine, Muenster, 2007-2011.

Group leader at EMBL since 2011.

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, 1407-14

Dietrich, J. E. and Hiiragi, T. (2007) Stochastic patterning in the mouse pre-implantation embryo. Development, 134, 4219-31

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, 1081-92

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, 360-64

Figure 1: Computer simulation of blastocyst morphogenesis

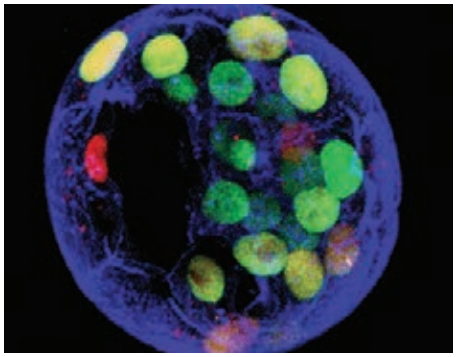


Figure 2: Unprecedented molecular heterogeneity during mouse blastocyst patterning. Cells expressing Nanog (green), Cdx2 (red) or both (yellow)

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 surrounded by one-cell layer of the trophectoderm. The former lineage generates the embryo proper, while the latter yields the extra-embryonic tissue. Despite its importance for understanding mammalian development and for stem cell research, the molecular mechanism of blastocyst patterning has long been elusive. How is the symmetry broken in the mammalian embryo? How is the definitive embryonic pattern established?

We have developed a live-imaging system for mouse pre-implantation embryos, demonstrating unexpectedly high dynamicity during early morphogenesis. Our studies characterised key principles underlying early mammalian development: absence of polarity in the egg, and localised determinants play little or no role in generating the asymmetry; mechanical and structural context plays a key role (figure 1); stochastic processes generate dynamic heterogeneity (figure 2). Collectively, an attractive hypothesis is that early mammalian embryogenesis may be 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 the mechanisms of early mammalian development, it will be essential to address how the diverse inputs acting on individual cells are integrated in the embryo at the systems level. Thus we have recently established necessary tools and multi-disciplinary strategies: a fluorescence-based gene-trap mouse lines that visualise molecular dynamics during embryonic patterning; gene expression profiles of every single cell in the mouse pre-implantation embryo; computer simulation of the blastocyst morphogenesis (figure 1). Mouse pre-implantation embryo is suitable for systems-level study because: it is an isolated system composed of a limited number of cells (up to 64) and cell populations (four); the development can be recapitulated and live-imaged *in vitro*; and micro-manipulation is applicable. We aim at understanding general principles and robustness underlying early mammalian development.

Future projects and goals

We adopt a wide variety of experimental strategies including molecular genetics, live-imaging, cell physics, and modelling, to address fundamental questions in development and cell biology at a molecular, cellular and systems levels. Our goals are:

- identification of the symmetry breaking cue in the mouse embryo
- understanding the molecular mechanism leading to the first lineage establishment
- 4D imaging microscopy to digitally reconstruct embryogenesis
- evaluating the significance of dynamic molecular heterogeneity.

Microglia: the guardians of the developing brain

Francesca Peri

PhD 2002, University of Cologne.

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

Group leader at EMBL since 2008.

The Peri group combines genetic approaches with quantitative imaging techniques to study interactions between neurons and the microglia that eliminate cellular debris in the 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 in order 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.

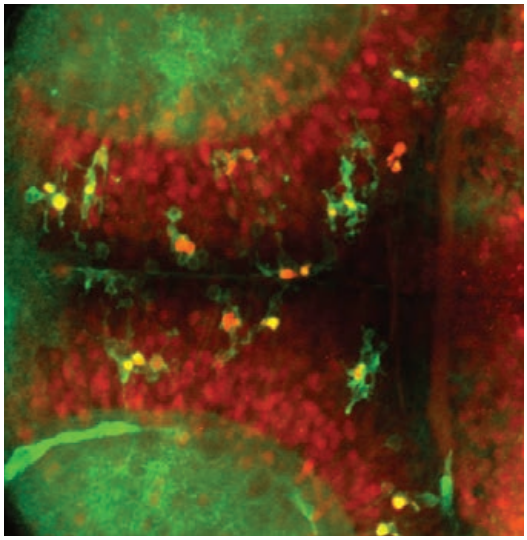
Selected references

Schlegelmilch, T., Henke, K. & Peri, F. (2011). Microglia in the developing brain: from immunity to behaviour. *Curr. Opin. Neurobiol.,* 21, 5-10

Peri, F. (2010). Breaking ranks: how leukocytes react to developmental cues and tissue injury. *Curr. Opin. Genet. Dev.*, 20, 416-19

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, 916-27

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



Gene regulation and genome architecture

The Spitz group aims to understand where in the chromosome regulatory sequences are located, and how their arrangement controls gene expression.

François Spitz

PhD 1997, Institut Cochin de Génétique Moléculaire, Paris.

Postdoctoral research at the University of Geneva.

Group leader at EMBL since 2006.

Selected references

Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., de Laat, W., Spitz, F. & Duboule, D. (2011). A regulatory archipelago controls hox genes transcription in digits. *Cell*, 147, 1132-45

Ruf, S., Symmons, O., Uslu, V.V., Dolle, D., Hot, C., Ettwiller, L. & Spitz, F. (2011). Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nat. Genet.*, 43, 379-86

Rowe, H.M., Jakobsson, J., Mesnard, D., *et al.* (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature*, 463, 237-40

Friedli, M., Nikolaev, S., Lyle, R., Arcangeli, M., Duboule, D., Spitz, F. & Antonarakis, S.E. (2008). Characterization of mouse Dactylaplasia mutations: a model for human ectrodactyly SHFM3. *Mamm. Genome*, 19, 272-78

Previous and current research

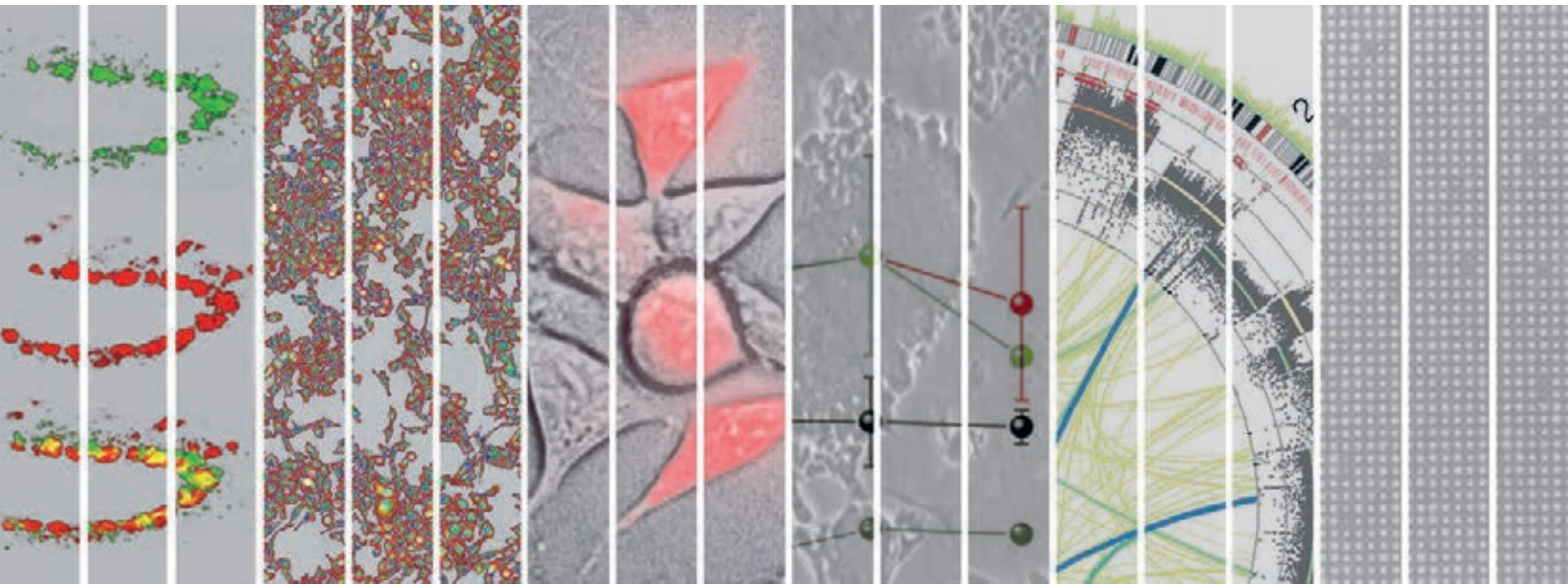
While the major part of vertebrate genomes is not coding for proteins, it nevertheless contains functionally important information. In particular, the tightly regulated expression of many developmental genes is achieved through the coordinated action of multiple cis-regulatory elements, which could be 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.

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. Our lab has developed several experimental approaches in the mouse that allow us to explore and assess the regulatory organisation of the mammalian genome 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. These approaches supply mouse models to study the molecular etiology of these genomic disorders, and 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 can be brought close together 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 such specific long-range chromatin interactions are still poorly understood. By combining advanced genomic engineering with chromatin profiling and conformation analyses (using next-generation sequencing and imaging), we aim to learn 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.



Genome Biology

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 transcriptional, post-transcriptional and post-translational. 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 regulatory steps. Their translation results in the production of proteins, whose functions define the characteristics of different cell types, or cellular phenotypes. Not all RNAs are translated, however. In recent years, multiple types of non-coding RNAs have been discovered that display diverse functionality. Genetic variation in non-coding and protein-coding genes alike, as well as the regulatory elements that govern their expression, can adversely affect the function of these genes, leading to diseases such as cancer. Groups within the Unit are investigating various aspects of genome biology in order to understand these processes leading from genotype to phenotype.

A notable strength of the Unit is its ability to address questions at different scales, ranging from detailed mechanistic studies (using biochemistry, genetics, microfluidics and chemistry) to genome-wide studies (using functional genomic, proteomic and computational approaches), often with developing new enabling technologies. For example, the development and integration of chemistry and microfluidic devices with the recent advances in next-generation 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

Regulatory networks driving cell fate decisions: Dissecting the logic

Eileen Furlong

PhD 1996, University College Dublin.

Postdoctoral research at Stanford University.

Group leader at EMBL since September 2002.
Joint Head of Genome Biology Unit and senior scientist since 2009.

The Furlong group aims to understand fundamental principles of transcription, focusing on the processes that determine what a cell becomes during embryonic development.

Previous and current research

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

Our research includes studies of the mechanism of enhancer function (figure 1) and the interplay of transcription factors and chromatin state (figure 2), as well as studies of 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 transcription and 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 have developed a very accurate and sensitive method to investigate cell type specific changes in chromatin status and chromatin binding protein occupancy in the context of a multicellular embryo's development (figure 2). We are currently using this method to examine the interplay between changes in chromatin remodelling with dynamic changes in transcription factor occupancy and developmental transitions.

Variation and plasticity in cis-regulatory networks: Variation in cis-regulatory elements can affect gene expression and account for individual differences in phenotypes like taste sensation and olfactory sensitivity. However, little is known about how much variation in gene expression or transcription factor function can be tolerated for essential developmental process during embryonic development. We plan to investigate this by extending our current knowledge of the transcriptional network regulating cell fate choices during mesoderm development to many *Drosophila* individuals (isogenic *Drosophila* strains) whose genomes have been fully sequenced.

Predictive models of embryonic development: Our previous work demonstrated that using only information on combinatorial occupancy of transcription factors is sufficient to predict spatio-temporal cis-regulatory activity (*Nature*, 2009) and that information on chromatin state and RNA polymerase II occupancy on enhancers can predict the precise timing and location of active enhancer elements *de novo* (*Nature Genetics*, 2012). We plan to build on this by 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.

Selected references

Junion, G.*, Spivakov, M.*, Girardot, C., Braun, M., Gustafson, E.H., Birney, E., Furlong, E.E. (2012). A transcription factor collective defines cardiac cell fate and reflects the developmental history of this cell lineage. *Cell*, February 2012

Bonn, S.*, Zinzen, R.P.*, Girardot, C.*, *et al*. (2012). Tissue specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature Genetics*, January 2012

Furlong, E.E. (2011). A fly in the face of genomics. *Nature,* 471, 458-9

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

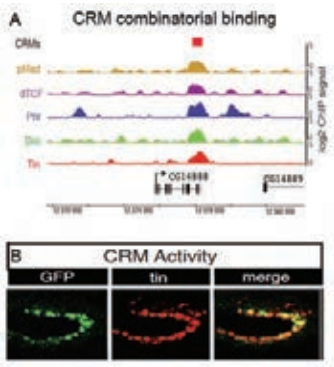


Figure 1: Enhancers can function by highly cooperative transcription factor occupancy using very flexible motif content and organisation (Junion, Spivakov, et al*, 2012)*

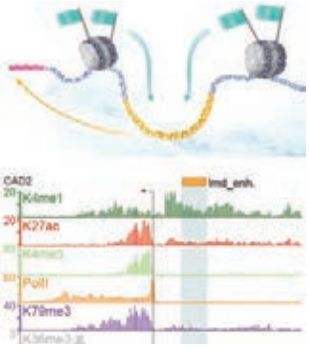


Figure 2: Chromatin state and RNA polymerase II occupancy on enhancers (yellow) is highly predictive of enhancers activity and are very dynamic mirroring that of dynamic enhancer usage during development (Bonn, Zinzen, Girardot, et al*, 2012)*

Systems genetics

The Steinmetz group bridges diverse domains of genome science, from deciphering the structure and function of genomes to the application of these insights in understanding diseases.

Selected references

Xu, Z., Wei, W., Gagneur, J., Clauder-Munster, S., Smolik, M., Huber, W. & Steinmetz, L.M. (2011). Antisense expression increases gene expression variability and locus interdependency. *Mol. Syst. Biol.*, 7, 468

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, 147-50

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, 1033-37

Mancera, E., Bourgon, R., Brozzi, A., Huber, W. & Steinmetz, L.M. (2008). High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature*, 454, 479-85

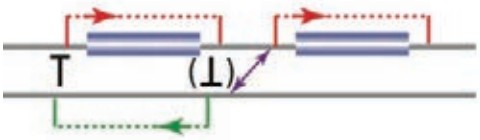


Figure 1: Antisense transcription enables local dispersion of regulatory signals via bidirectional promoters (Xu et al.*, 2011,* Molecular Systems Biology*).*

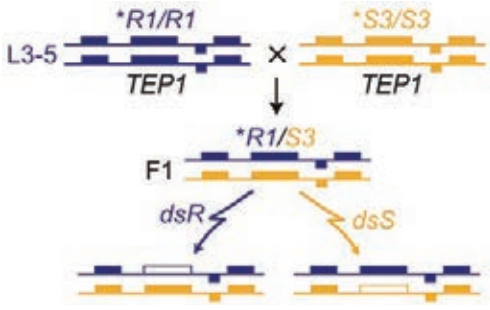


Figure 2: Reciprocal allele-specific RNAi identifies the mosquito gene TEP1 as a major contributor to resistance to malaria parasites (Blandin et al.*, 2009,* Science*).*

Lars Steinmetz

PhD., Stanford University, 1997-2001.

Postdoctoral research at Stanford Genome Technology Center, 2002-2003.
Visiting group leader at Stanford Genome Technology Center since 2003.

Group leader at EMBL since 2003.
Joint Head of Genome Biology Unit and senior scientist since 2009.

Previous and current research

One of the most daunting challenges in medicine is the complex nature of most common diseases (including cancer, diabetes, and heart disease) due to interactions between multiple genetic variants and environmental influences. Our research is directed at understanding such complex traits; to do so, we develop novel genomic approaches to investigate the molecular processes that link genotype to phenotype, identify the underlying factors, and quantify their contributions. We investigate variation at the level of the genome, transcriptome and proteome, which we integrate with higher-level phenotypes. We also use the resulting molecular networks to predict and evaluate intervention points that enable modulation of phenotype.

To this end, we are using budding yeast and human cell culture as model systems for functional genomics. Using the strand-specific yeast tiling microarray that we developed, we have discovered pervasive transcription of the genome and shown that much of this transcription originates from bidirectional promoters, which enable the spread of regulatory signals via antisense transcription (figure 1). We are currently carrying out functional and mechanistic studies of these non-coding RNAs as well as developing new sequencing based technologies to define transcriptome architecture. Furthermore, we have dissected the genetic architecture of complex traits such as high-temperature growth in yeast as well as mosquito resistance to the malaria parasite (figure 2) and are continuing with further traits such as drug resistance. To clarify the process of trait inheritance, we have generated a high-resolution map of meiotic recombination in yeast and are now studying this process in humans via whole genome sequencing.

Future projects and goals

We are developing new technologies to carry out rapid, high-resolution dissection of the genetic and environmental factors underlying complex traits; we will combine this knowledge with other 'omic' measurements that will allow the computational prediction of phenotype from genotype. We will continue testing novel genetic and chemical therapeutics in experimental models of mitochondrial and neurodegenerative diseases. Furthermore, we have begun working with patient-derived iPS cells to measure the phenotypic consequences of disease-associated alleles using whole genome sequencing, transcriptome profiling, and RNAi. Ultimately, by integrating genetics, genomics, systems biology, and computational modelling, we aim to develop approaches that unravel disease mechanisms and predict effective therapeutics, enabling personalised and preventive medicine.

Computational biology and genomics

Wolfgang Huber

PhD 1998, Statistical Physics, University of Freiburg.

Postdoctoral at IBM Research, California and DKFZ, Heidelberg.

Group leader at EMBL-EBI 2004-2009 and at EMBL Heidelberg since 2009.
Senior scientist since 2011.

The Huber group develops computational and statistical methods to design and analyse novel experimental approaches in genetics and cell biology.

Previous and current research

Our aim is to understand biological systems through systems-wide maps and quantitative models. Our main tool is statistics - the science of computing with uncertainty, making rational inference based on incomplete and noisy data. Together with its sister discipline, machine learning, it helps humans to discover patterns in large datasets and to infer underlying mechanisms, predictive and causal relationships.

Our aim is biological discovery – to understand genetic and phenotypic variation between individuals on a genome-wide scale. We have projects in the areas of gene expression and regulation, in the genetics of complex phenotypes and genetic interactions, in cell division and cell migration and in cancer genomics.

The group brings together expertise from quantitative disciplines – mathematics, statistics, physics and computer science – and from different areas of biology to design and analyse genomic experiments and their biological interpretation. Computational and statistical methods are 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

An emphasis of the group's work is on project-oriented collaborations with experimenters. We aim to develop the computational techniques needed to make new types of experiments feasible and to turn the data into biology. Among our current projects are:

- Large-scale systematic maps of gene-gene and gene-environment interactions by automated phenotyping, using image analysis, machine learning, sparse model building and causal inference.
- DNA-, RNA- and ChIP-Seq and their applications to gene expression regulation: statistical and computational foundations.
- Cancer genomics, genomes as biomarkers, cancer phylogeny.
- Image analysis for systems biology: measuring the dynamics of cell cycle and of cell migration of individual cells under normal conditions and many different perturbations (RNAi, drugs).
- Systematic mapping of molecular interactions and life cycles within single cells.
- Open source software for genomics, high-throughput phenotyping and statistical bioinformatics, to support reproducible research and wide dissemination of state-of-the-art methods.

Right: 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

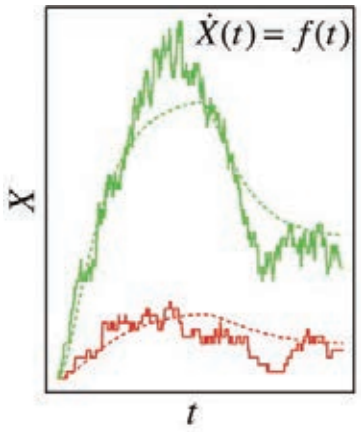
Selected references

Horn, T., Sandmann, T., Fischer, B., Axelsson, E., Huber, W. & Boutros, M. (2011). Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. *Nat. Methods*, 8, 341-46

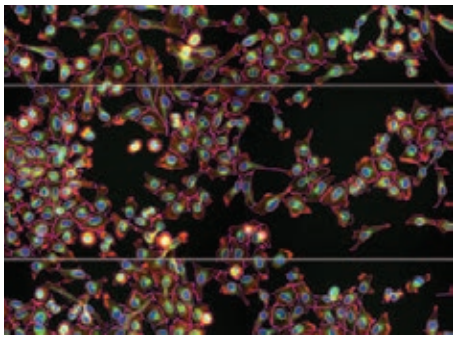
Anders, S. & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.*, 11, R106

Bourgon, R., Gentleman, R. & Huber, W. (2010). Independent filtering increases detection power for high-throughput experiments. *Proc. Natl. Acad. Sci. USA*, 107, 9546-51

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., *et al.* (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature*, 457, 1033-37



Above: Dynamical modelling of protein life cycles from single cell fluorescence timer imaging



Investigation of phosphatases using chemical biology tools

The Köhn group combines molecular biology, biochemistry and synthetic chemistry to develop new approaches to study phosphatases, which can play a major role in cancer.

Maja Köhn

PhD 2005 MPI for Molecular Physiology, Dortmund.

Postdoctoral research at Harvard University, Cambridge, Massachusetts.

Group leader at EMBL since 2007.

Selected references

McParland, V., Varsano, G., Li, X., Thornton, J., Baby, J., Aravind, A., Meyer, C., Pavic, K., Rios, P. & Köhn, M. (2011). The metastasis-promoting phosphatase PRL-3 shows activity toward phosphoinositides. *Biochemistry*, 50, 7579-90

Meyer, C. and Köhn, M. (2011). Efficient scaled-up synthesis of N-a-Fmoc 4-Phosphono(difluoromethyl)-Lphenylalanine and its incorporation into peptides, *Synthesis*, 20, 3255-60

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, 958-68

Köhn, M. (2009). Immobilization strategies for small molecule, peptide and protein microarrays. *J. Pept. Sci.*, 15, 393-97

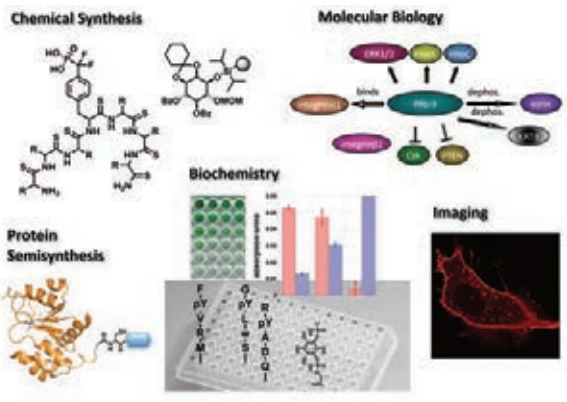
Previous and current research

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 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. In collaboration with Mathieu Bollen (KU Leuven), 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 the Wilmanns group (page 112), and the Thornton group (page 78), we use computational, biochemical and structural approaches to view, predict and validate these networks.



Investigation of disease-promoting phosphatases

Future projects and goals

We are interested in further developing chemical methods to stabilise peptides and 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 the Márquez group (page 104). 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.

Origin and function of genetic variation

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.

The Korbel group combines experimental and computational biology to decipher the function and origin of genetic variation, with a particular focus on genomic structural variants.

Previous and current research

Polymorphic structural variations (SV) are a major source of variation in the human genome, exceeding single nucleotide polymorphisms in their net effect on DNA sequence (number of nucleotides affected). A main focus of our group is deciphering genetic variation in humans, research that we complement by model organism studies.

We recently developed next-generation DNA sequencing-based approaches for identifying SVs (paired-end mapping, split-read analysis, and read-depth analysis). As participants of the 1000 Genomes Project, we are using these tools with the aim of constructing a state-of-the-art genetic variation ('variome') map in two thousand individuals. First versions of the map provide initial insights into the mechanistic origin of SVs and enable the delineation of mutational hotspots in the human genome. Our group dissects the functional impact of genetic variation by integrating cellular (transcriptome and protein-DNA binding) and individual (e.g. clinical) phenotype data.

Dissecting mechanisms of cancer: We participate in cancer genome projects in the context of the International Cancer Genome Consortium, including those focused on early-onset prostate cancer and paediatric brain tumours. Childhood medulloblastomas harbour only a few point mutations (1 to 17 mutations affecting genes per tumour), which raised the question of which molecular events actually caused these cancers to develop. Using whole-genome sequencing, we found that medulloblastomas commonly develop as an immediate consequence of catastrophic DNA rearrangements that occur as a single molecular event, termed 'chromothripsis', involving up to hundreds of breaks typically localised on a single chromosome. This phenomenon is strongly associated with germline mutations of the *TP53* gene, encoding the p53 protein – the vast majority of medulloblastoma patients with chromothripsis harbour such heritable mutations. This discovery has clinical implications and relevance for assessing the feasibility of 'personalised medicine' (i.e., the use of deep sequencing to guide medical decisions): Specifically, extensive cancer screening in families with heritable *TP53* mutations can increase patient survival. Furthermore, current cancer treatments, including radiotherapy and DNA damaging agents, may trigger secondary cancers in the context of heritable *TP53* mutations (due to deficient DNA repair in various cell types) – calling for the assessment of other treatments in these patients.

Future projects and goals

- Constructing a complete map of human genome variation, including presently difficult-to-ascertain genomic regions, using 'third generation' (long-range) DNA sequencing.
- Development of wet-lab and computational approaches for deciphering the molecular origin and function of SVs in humans and model organisms (yeast and flies).
- Dissecting the effect of genetic variation on biological systems in healthy states and in cancer, by integrating 'variomes' with expression (e.g., transcriptome) and epigenetics (e.g., methylome) data.
- Deciphering the mechanistic basis of 'chromothripsis', a rearrangement phenomenon thought to cause the development of 2-3% of human cancers, especially highly aggressive malignancies.

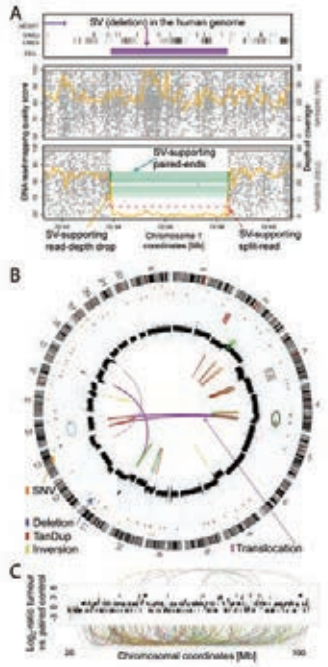
Selected references

Rausch T, Jones DT, Zapatka M, *et al.* (2012) Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with *TP53* mutations, *Cell*, 148, 59-71

Schlattl, A., Anders, S., Waszak, S.M., Huber, W. & Korbel, J.O. (2011). Relating CNVs to transcriptome data at fine resolution: Assessment of the effect of variant size, type, and overlap with functional regions. *Genome Res.*, 21, 2004-13

Mills, R.E., Walter, K., Stewart, C., *et al.* (2011). Mapping copy number variation by population-scale genome sequencing. *Nature*, 470, 59-65

Kasowski, M., Grubert, F., Heffelfinger, C., *et al.* (2010). Variation in transcription factor binding among humans. *Science*, 328, 232-5



(A) Mapping genomic structural variation by paired-end mapping, read-depth analysis, and split-read analysis (Mills et al., Nature *2011). (B) Mutational landscape in a childhood medulloblastoma genome. (C) Catastrophic chromosome rearrangements resulting from chromothripsis (Rausch* et al.,Cell *2012)*

Quantitative proteomics

The Krijgsveld team uses a combination of biochemistry, analytical chemistry, mass spectrometry and bioinformatics to study the role of proteins in cell behaviour.

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., Dormeyer, W., Braam, S.R., *et al.* (2010). Identification of cell surface proteins for antibody-based selection of human embryonic stem cell-derived cardiomyocytes. *J. Prot. Res.*, 9, 1610-18

Van Hoof, D., Munoz, J., Braam, S.R., Pinkse, M.W., Linding, R., Heck, A.J., Mummery, C.L. & Krijgsveld, J. (2009). Phosphorylation dynamics during early differentiation of human embryonic stem cells. *Cell Stem Cell*, 5, 214-26

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,* 8,1566-78

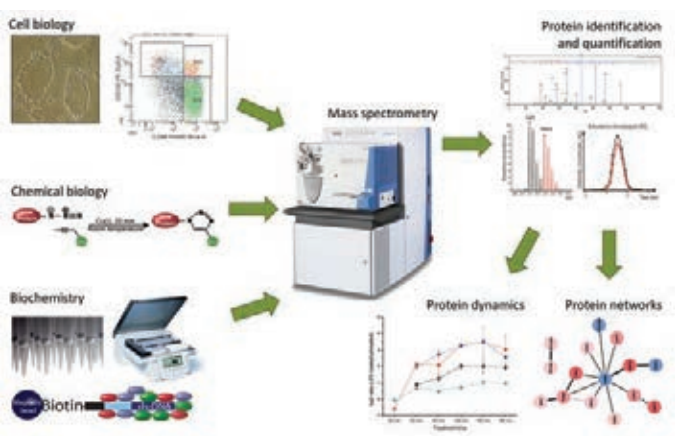
Van Hoof, D., Pinkse, M.W., Oostwaard, D.W., Mummery, C.L., Heck, A.J. & Krijgsveld, J. (2007). An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics. *Nat. Methods*, 4, 677-8

Previous and current research

Proteins fulfill most of the functions that are crucial for cellular behaviour. 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 centered on quantitative proteomics, combining biochemistry, analytical chemistry, mass spectrometry and bioinformatics. Mass spectrometry coupled to liquid chromatography has matured to the stage that many thousands of proteins can be analysed simultaneously, encompassing entire proteomes for relatively simple organisms (e.g. yeast) while covering large portions of more complex proteomes (e.g. mammals). Our lab is equipped with state-of-the-art mass spectrometric technology (Thermo Orbitrap Velos Pro) 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 cellular differentiation and dedifferentiation, and to the initial steps in hematopoiesis. This should provide molecular and mechanistic clues as to how cells progress through subsequent developmental stages. 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 will continue to develop and implement tools to identify proteins interacting with enhancer elements and promoter regions, in order to understand how transcription of developmentally important genes is regulated.



Future projects and goals

- Studying the changing proteome during gain and loss of pluripotency. We will focus on differentiation of hematopoietic 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.

Miniaturising biology and chemistry in microfluidic systems

Christoph A. Merten

PhD 2004, University of Frankfurt.

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

Junior group leader at the Institut de Science et d´Íngénierie Supramoléculaire, Strasbourg.

Group leader at EMBL since 2010.

The Merten group develops novel approaches in microfluidic technology to address complex, multidisciplinary questions at the interface of biology, chemistry and engineering.

Previous and current research

Working on the micro scale offers some unique advantages:

- Drastically increased throughput (processing up to a million samples an hour).
- Superb spatio-temporal resolution (assays can be carried out on micrometer length scales and sub-millisecond timescales).
- Low material consumption enabling single organism, single cell or even single molecule assays.

Over the last couple of years we have developed powerful microfluidic platforms for cell-based and (bio)chemical assays. In our laboratory, we perform all steps ranging from the design/manufacturing of microfluidic chips and detection systems to the cultivation and study of human cells and multicellular organisms (prior knowledge in microfluidics is not obligatory for joining the group). Furthermore, we are interested in combinatorial chemistry, perform computational fluid dynamics simulations and develop novel software controlling our microfluidic systems. For many applications we use two-phase microfluidics, in which aqueous droplets within an immiscible oil phase serve as miniaturised reaction vessels. As they can be generated at kilohertz frequencies, they are of particular interest for high-throughput screens. Furthermore, the small assay volumes (pico- to nanoliters) facilitate the obtainment of high concentrations of nucleic acids (mRNA, DNA) or proteins (e.g. secreted antibodies) from individually encapsulated cells, paving the way for single cell assays. On the other hand, we use continuous-phase microfluidics to generate laminar flow patterns, where we expose cells and organisms (or even small parts thereof) to different chemical environments. Amongst other applications, this allows analysing signalling events in developing embryos.

Future projects and goals

Having a comprehensive microfluidic toolbox at hand (and expanding it continuously), we are now focusing on applications in three different research fields:

Cell biology: We are planning large-scale chemical pertubations to analyse stem cell differentiation and to map the interactions between different cellular pathways with a special focus on cellular senescence. In particular, we will analyse the crosstalk between autophagy, insulin signaling and mitochondrial activity.

Developmental biology: The exact regulation and timing of developmental steps during embryogenesis remain an enigma. To reveal the underlying mechanisms we develop microfluidic platforms allowing single cell transcriptomics. Furthermore we cultivate multicellular organisms in laminar flow systems to pertubate and analyse developmental steps in particular parts of the organism.

Combinatorial chemistry: The possibility of rapidly generating, mixing and analysing huge sample numbers allows the exploration of large areas of chemical structure space. Focusing on click chemistry and one-bead-one-compound libraries, we are interested in the identification of novel bioactive molecules such as antimicrobial peptides.

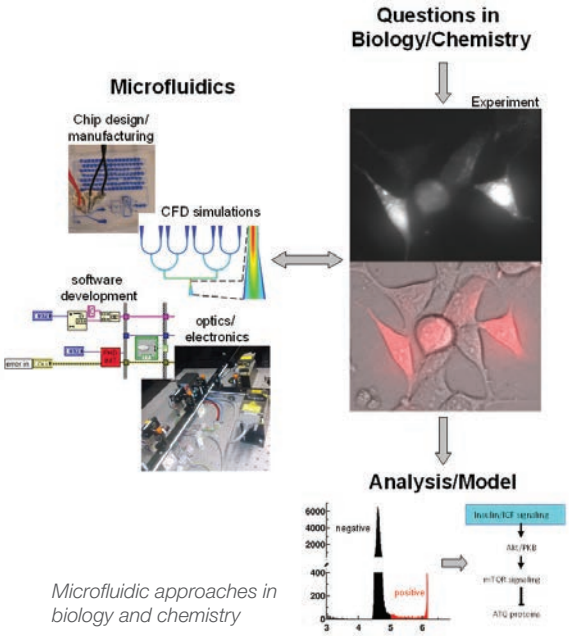
Selected references

Vyawahare, S., Griffiths, A.D. & Merten, C.A. (2010) Miniaturization and parallelization of biological and chemical assays in microfluidic devices. *Chem. Biol.*, 17, 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, 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, 229-35

Clausell-Tormos, J., Lieber, D., Baret, J.C., *et al.* (2008) Droplet-based microfluidic platforms for the encapsulation and screening of Mammalian cells and multicellular organisms. *Chem. Biol.* 15, 427-37. Research Highlight in *Nature Methods* 5, 580-81



Microfluidic approaches in biology and chemistry

Dissecting bacterial lifestyle and interspecies interactions with systems approaches

The Typas group develops and utilises high-throughput methods to study the cellular networks of different species of bacteria, and how these bacteria interact with the environment and with each other.

Athanasios Typas

Ph.D., 2006 Institute of Microbiology and Plant Physiology, Freie Universität Berlin, Germany.

Postgraduate research, University of California, San Francisco.

Group leader at EMBL since 2011.
Joint appointment with the Structural and Computational Biology Unit

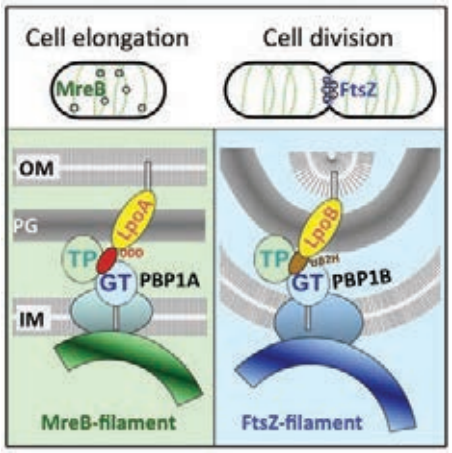
Selected references

Oh, E., Becker, A.H., Sandikci, A., Huber, D., Chaba, R., Gloge, F., Nichols, R.J., Typas, A., Gross, C.A., Kramer, G., Weissman, J.S. & Bukau, B. (2011). Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor *in vivo*. *Cell*, 147, 1295-1308

Nichols, R.J., Sen, S., Choo, Y.J., Beltrao, P., Zietek, M., Chaba, R., Lee, S., Kazmierczak, K.M., Lee, K.J., Wong, A., Shales, M., Lovett, S., Winkler, M.E., Krogan, N.J., Typas, A. & Gross, C.A. (2011). Phenotypic landscape of a bacterial cell. *Cell*, 144, 143-56

Typas, A., Banzhaf, M., van den Berg van Saparoea, B., Verheul, J., Biboy, J., Nichols, R.J., Zietek, M., Beilharz, K., Kannenberg, K., von Rechenberg, M., Breukink, E., den Blaauwen, T., Gross, C.A. & Vollmer, W. (2010). Regulation of peptidoglycan synthesis by outer-membrane proteins. *Cell*, 143, 1097-1109

Typas, A., Nichols, R.J., Siegele, D.A., Shales, M., Collins, S.R., Lim, B., Braberg, H., Yamamoto, N., Takeuchi, R., Wanner, B.L., Mori, H., Weissman, J.S., Krogan, N.J. & Gross, C.A. (2008). High-throughput, quantitative analyses of genetic interactions in *E. coli*. *Nat. Methods*, 5, 781-87



Two novel niche-specific lipoproteins control peptidoglycan synthesis from the outside of the sacculus (Typas et al., *Cell 2010)*

Previous and current research

The recent explosion of genomic sequence information provides a first step towards better understanding diverse bacteria, but also makes it crucial to develop large-scale phenotyping approaches to characterise functions of novel genes and to map them within pathways. Research pioneered in yeast indicates that some of the most powerful global phenotyping approaches are based on scaling up long-standing concepts in biology. Epistasis-genetic interactions assess how the function of one gene depends on the function of another; chemical genetic interactions measure how the function of one gene affects cellular responses to chemical stress. We have developed quantitative high-throughput versions of these approaches in *E. coli*, and used them to obtain a mechanistic understanding of key aspects of its envelope function and architecture.

The bacterial envelope is sandwiched between two membranes in gram-negative bacteria and constitutes the interface of the bacterium to the environment. This compartment acts as a protective barrier, but also allows bacteria to sense changes in and communicate with their environment, and is vital for pathogenesis, cell morphogenesis and cell developmental programs. Although many envelope structural components have been characterised, we often have limited information on how their biosynthesis and transport are interconnected, regulated, or linked to the overall status of the cell, how they sense perturbations, and how signals are transduced to achieve homeostasis.

Working at the intersection between systems genomic biology and mechanistic molecular biology, we have discovered key missing players of major envelope components, uncovered niche-specific regulation of conserved envelope processes, identified linking proteins that allow coordination between processes and mapped network rewiring under different stresses. Specific examples of the biological stories we have uncovered include: i) identification of the first regulators controlling peptidoglycan synthesis outside of the cell, thereby providing an opportunity for outer membrane status to influence the fundamental shape-forming structure in bacterial cells (see figure); ii) discovery of a new player that orchestrates outer membrane constriction during cell division; iii) a mechanism to transduce information about cell size/shape to the transcriptional machinery; iv) a mechanism to control the activity of lipoproteins.

Future projects and goals

Our main goal is to expand our efforts in two directions. First, we want to develop analogous high-throughput methods for other bacteria. This will enable us to generate comprehensive interaction datasets in different bacteria, enrich them with other available genome-wide resources and ultimately use them to create inferences about common and individual characteristics of biology in those organisms. Second, we plan to expand this technology to report on more social characteristics of bacteria. This will enable us to probe interspecies interactions and interactions with the host, shedding light on different mechanistic aspects of these interactions. Our specific aims include: i) generation of global genetic and chemical genetic interaction profiles for different pathogen and commensal bacteria and integration with other available global phenotyping data; ii) cross-species comparisons to map key features of divergence in the organisation of the bacterial envelop; iii) development of high-throughput approaches to study bacterial interspecies interactions and bacterial-host interactions.

Structural and Computational Biology

The 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. Dedicated large scale biochemistry, proteomics, chemical biology, biophysics, and cell biology approaches complement the structural biology activities and, in conjunction with a wide range of innovative computational biology activities, are integrated 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. Along those lines, groups in the 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. In the thermophilic fungus *Chaetomium thermophilum* spatial and temporal networks will be deduced using multidisciplinary approaches including structural studies, large scale biochemistry and computational biology. Together, they will provide insight into eukaryotic thermophily at the molecular and cellular 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, several groups based in other Units have shared appointments with the Unit.

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, 700 MHz, 600 MHz and 500 MHz NMR spectrometers; and several transmission electron microscopes, including a 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

Deciphering function and evolution of biological systems

Peer Bork

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

By analysing and comparing complex molecular data, the Bork group predicts function, gains insights into evolution, and makes connections between genes, organisms and ecosystems.

Previous and current research

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; 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 aspects 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, and also to understand microbial community interactions, with application potential for human health and wellbeing. For example, our recent discovery of enterotypes – three distinct community compositions in the human gut analogous to blood groups (Arumugam *et al*., 2011 *Nature*) – was considered as one of the breakthroughs of 2011 by *Science* because it might explain different responses of people to drug intake and diet. 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

Much of the group's focus in the coming years will be on the human gut. We aim to understand biological processes upon drug treatment, considering 'human' as a biological system with various readouts, from drug side-effects to toxicology data. We will explore networks between proteins and chemicals such as lipids or carbohydrates and link them to phenotypic data such as disease status. We will also look at the 2kg or so of bacteria 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. Potential applications could include microbial biomarkers for diseases or antibiotic resistance potential. We also want to understand how microbial communities evolve in us, how frequently they are transmitted parentally or horizontally, and how they communicate with each other and with our cells.

Other projects include involvement in collaborations studying various other systems, such as biodiversity (with the TARA Oceans project).

The group is partially associated with the Max Delbruck Center for Molecular Medicine in Berlin and with the Molecular Medicine Partnership Unit.

Selected references

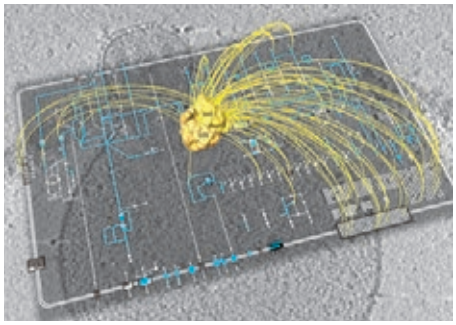
Arumugam, M., *et al.* (2011) Enterotypes of the human gut microbiome, *Nature*, 473, 174-80

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, 1235-40

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)*



Molecular mechanisms of transcriptional regulation and epigenetics

The Müller group uses biophysical and biochemical approaches to learn about the molecular mechanisms of transcription regulation in eukaryotes, where DNA is packaged into chromatin.

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.

Selected references

Glatt S., Létoquart J., Faux C., Taylor N.M.I, Séraphin B. and Müller, C.W. (2012). The elongator subcomplex Elp456 is a hexameric RecA-like ATPase. *Nat. Struct. Mol. Biol*., doi:10.1038/nsmb.2234

Glatt, S., Alfieri, C. & Müller, C.W. (2011). Recognizing and remodeling the nucleosome. *Curr. Opin. Struct. Biol.*, 21, 335-41

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

Moriniere, J., Rousseaux, S., Steuerwald, U., Soler-Lopez, M., Curtet, S., Vitte, A.L., Govin, J., Gaucher, J., Sadoul, K., Hart, D.J., Krijgsveld, J., Khochbin, S., Müller, C.W. & Petosa, C. (2009). Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature*, 461, 664-68

Previous and current research

In the context of chromatin, we are interested how different sequence-specific transcription factors assemble on DNA and how these 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-working of large molecular machines like RNA polymerases or chromatin modifying complexes involved in the transcription process. Finally, we would like to gain insight into how DNA sequence information and epigenetic modifications work 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 approaches. 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.

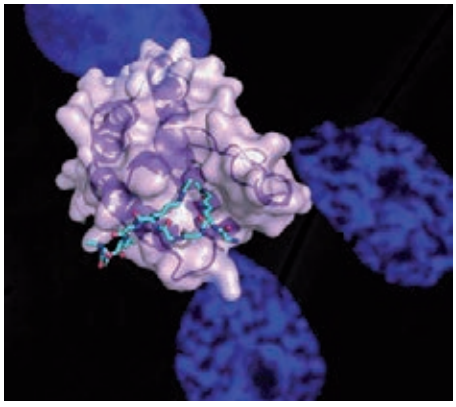
Chromatin modifying complexes: The accessibility of chromatin in eukaryotes is regulated by ATP-dependent chromatin remodelling factors and histone modifying enzymes. Both classes of enzymes use similar domains like bromodomains, chromodomains, MBT domains, PHD fingers and SANT domains for the controlled 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 nucleosomes and larger chromatin templates, and how their activities are regulated.

RNA polymerase III transcription: 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 regulators 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)



Mechanism of DNA recombination and its applications for research and therapy

Orsolya Barabas

PhD 2005, Eötvös Loránd University, Budapest, Hungary.

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

Group leader at EMBL since 2009.

The Barabas lab investigates how controlled DNA rearrangements – essential for survival at all levels of life, from individual cells to populations – are carried out at the molecular and cellular level.

Previous and current research

We mainly focus on DNA transposons, a class of mobile genetic elements that 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 can stably and heritably modify a target genome. Thus these 'jumping' 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 the structure of functional complexes, the chemistry they use to cut and paste DNA, as well as their target-site selection and regulation in the cell. Our techniques include structural biology (mainly X-ray crystallography), molecular biology, biochemistry and cell-based assays. We currently study: i) the movement of various DNA transposons; and ii) RNA-based regulatory pathways that control transposition.

Sleeping Beauty: This reactivated transposon has recently become a favoured genetic tool for forward mutagenesis screens, mapping regulatory landscapes, chromosomal engineering and even gene therapy (Ivics *et al*, 2009). We perform structural and functional studies to obtain a mechanistic understanding and, in collaboration with the Gavin group (page 60), we also investigate how it interacts with human cell.

Sequence-specific elements: One of the main obstacles of gene therapy is integration of the therapeutic gene at unwanted locations. Therefore, we seek recombinases or integrases that insert to a specific sequence and may provide a solution. Our recent work revealed the mechanism of the bacterial Insertion Sequence IS608, which integrates to a short specific sequence (Barabas *et al*, 2008). We found that target site recognition is achieved via base pairing between transposon and target DNA. The site of insertion can be altered by making point-mutations in the transposon (Guynet *et al*, 2009).We currently investigate if this target recognition can be extended to target potentially unique genomic sites. We are also studying a newly found mobile element, the plasticity zone transposon (TnPZ) from *Helicobacter pylori* (Kersulyte *et al*, 2009) and want to learn how it moves and integrates to a 7nt-long specific sequence.

Transposon regulation: To avoid deleterious outcomes, cells must keep transposons under control. Small RNAs can drive transposons to heterochromatin or destabilise their RNA intermediates. In collaboration with the Carlomagno (page 59) and Pillai (page 107) groups, we investigate these processes in prokaryotes and eukaryotes.

Future projects and goals

- We will select representative targets from less well known transposon and recombinase families to capture a broader picture of recombination pathways applied by nature.
- We will investigate the molecular mechanism of somatic genome assembly in ciliated protozoa (a large scale DNA rearrangement process that resembles transposition).
- Building on mechanistic insights, we will begin to develop novel genetic tools.

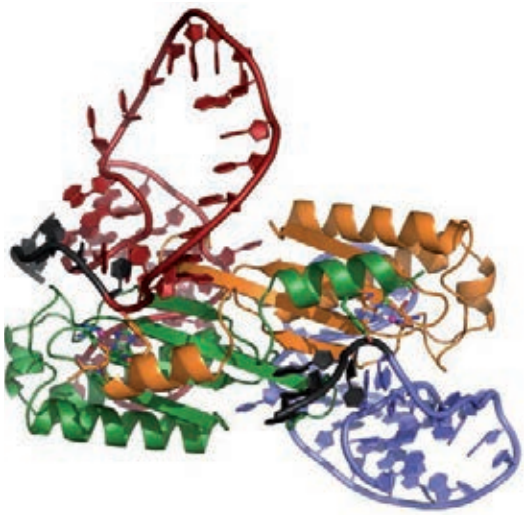
Selected references

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, 612-19

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, 302-12

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, 208-20



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

Structure and function of large macromolecular assemblies

Research in the Beck group combines biochemical approaches, proteomics and cryo-electron microscopy to study large macromolecular assemblies.

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.

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, 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, 762-65

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

Previous and current research

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 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

- 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.

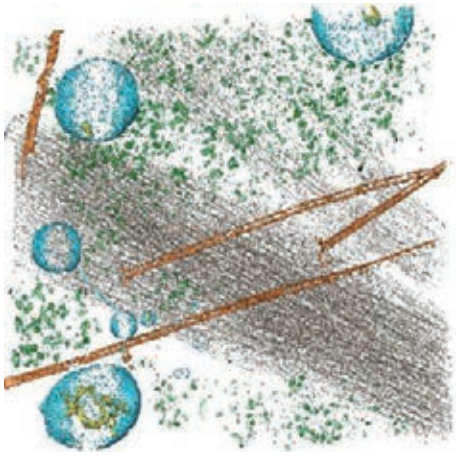


Figure 1: Cryo-electron tomogram of a fraction of the cytoplasm of a human cell. Microtubules are coloured in orange, stress fibres in grey, protein complexes in green, membranes in cyan and vesicular contents in yellow

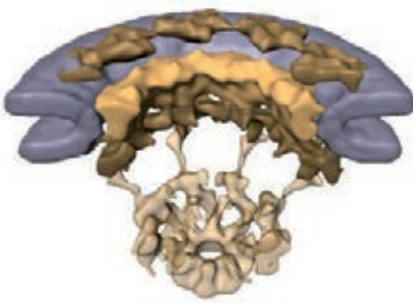


Figure 2: Structure of the nuclear pore complex. Membranes are coloured in grey, the scaffold structure in yellow and the nuclear basket in transparent brown

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

John Briggs

PhD 2004, Oxford University.

Postdoctoral research at the University of Munich.

Group leader at EMBL since 2006.

The Briggs group uses cryo-electron microscopy techniques to explore the mechanisms of assembly and budding of enveloped viruses and coated vesicles.

Previous and current research

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.

Selected references

Bharat, T.A., Riches, J.D., Kolesnikova, L., Welsch, S., Krahling, V., Davey, N., Parsy, M.L., Becker, S. & Briggs, J.A. (2011). Cryo-electron tomography of marburg virus particles and their morphogenesis within infected cells. *PLoS Biol.*, 9, e1001196

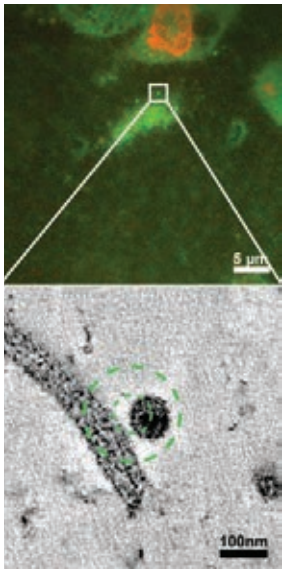
Beck, R., Prinz, S., Diestelkotter-Bachert, P., Rohling, S., Adolf, F., Hoehner, K., Welsch, S., Ronchi, P., Brugger, B., Briggs, J.A. & Wieland, F. (2011). Coatomer and dimeric ADP ribosylation factor 1 promote distinct steps in membrane scission. *J. Cell Biol.*, 194, 765-77

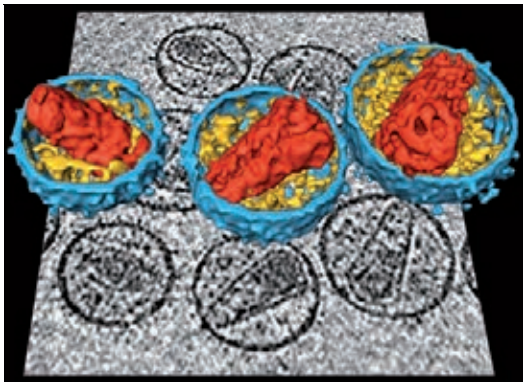
Kukulski, W., Schorb, M., Welsch, S., Picco, A., Kaksonen, M. & Briggs, J.A. (2011). Correlated fluorescence and 3D electron microscopy with high sensitivity and spatial precision. *J. Cell Biol.*, 192, 111-9

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

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.





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

Left: 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)*

Functional mechanisms of complex enzymes involved in RNA metabolism and methodology development for drug design

The Carlomagno group uses NMR spectroscopy in combination with biochemical and biophysical techniques to study the structure and dynamics of biomolecular complexes.

Selected references

Simon, B., Kirkpatrick, JP., Eckhardt, S., Reuter, M., Rocha, EA., Andrade-Navarro, MA, Sehr, P., Pillai, R., Carlomagno, T. (2011). Recognition of 2'-O-methylated 3'-end of piRNA by the PAZ domain of a Piwi protein, *Structure*, 19, 172-80

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, 6274-85

Orts, J., Tuma, J., Reese, M., *et al*. (2008). Crystallography-independent determination of ligand binding modes. *Angew Chem. Int. Ed. Engl.*, 47, 7736-40

Liu, S., Li, P., Dybkov, O., Nottrott, S., 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

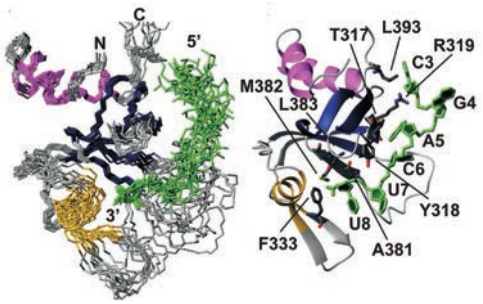


Figure 1: Overview of the structure of the complex between the Miwi-PAZ domain and the 3'-end 2'-O-methylated piRNA

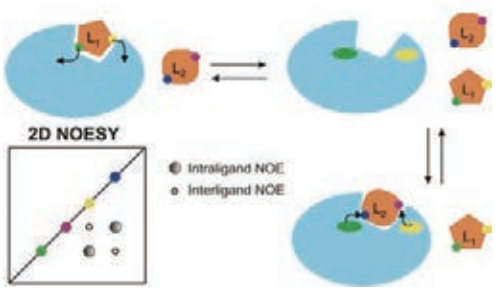


Figure 2: Schematic representation of the principle of the INPHARMA NOEs

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.

Previous and current research

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 107) 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.

Biomolecular networks

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.
Senior scientist since 2011.

The Gavin group focuses on detailed and systematic charting of cellular networks and circuitry at molecular levels in time and space.

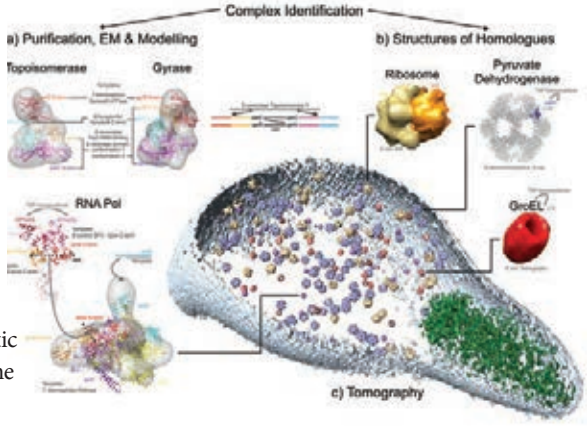
Previous and current research

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 main areas of interest are:

The charting of biological networks: Biological function at cellular levels is achieved by groups of interacting proteins or protein complexes that represent basic functional and structural units of proteome organisation. We systematically chart their dynamics using biochemical and quantitative mass spectrometry approaches in *S. cerevisiae*, *M. pneumoniae* and, in the future, thermophiles or other extremophiles. Datasets produced allow an unbiased overview of important biological principles. Collaborations with groups at EMBL and incorporation of structural models, single-particle EM and cellular electron tomograms provides supporting details for larger assemblies of protein complexes. We are also part of a network of EMBL groups tackling a range of biological networks in *M. pneumoniae*, where 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

i) Further development of chemical biology methods based on affinity purification to monitor protein-metabolites interaction; ii) global screen aiming at the systematic charting of the interactions taking place between the proteome and the metabolome in *S. cerevisiae* and *M. pneumoniae*; iii) develop new and existing collaborations 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, 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, 1235-40

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

Biological sequence analysis

The Gibson group investigates protein sequence interactions, undertakes computational analyses of macromolecules, and develops tools to enhance sequence analysis research.

Toby Gibson

PhD 1984, Cambridge University.

Postdoctoral research at the Laboratory of Molecular Biology, Cambridge.

At EMBL since 1986.
Team leader at EMBL since 1996.

Selected references

Dinkel, H., Michael, S., Weatheritt, R.J., et al. (2012). ELM – the database of eukaryotic linear motifs. *Nucleic Acids Res.*, 40, D242-D251

Sievers, F., Wilm, A., Dineen, D., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, 7, 539

Davey, N.E., Trave, G. & Gibson, T.J. (2010). How viruses hijack cell regulation. *Trends Biochem. Sci.*, 36(3):159-69

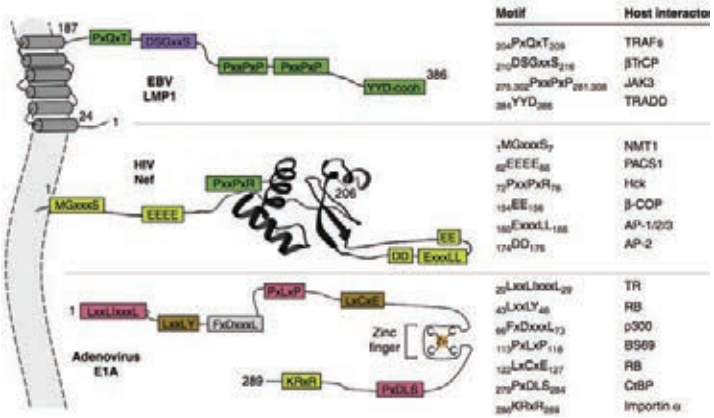
Gibson, T.J. (2009). Cell regulation: determined to signal discrete cooperation. *Trends Biochem. Sci.*, 34, 471-82

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 (ELM) resource 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). ELM data are now being used by many bioinformatics groups to develop and benchmark LM predictors. We are now actively hunting for new LM candidates. Recently, we proposed new candidate KEN anaphase destruction motifs, 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. Our collaborators Des Higgins (Dublin) and Julie Thompson (Strasbourg) have released Clustal Omega, a major update to the widely used multiple sequence alignment software. Additional public web servers include Phospho.ELM, a collection of some 42,000 reported phosphorylation sites and EpiC, a tool to aid in targeting epitopes for antibody selection.



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

Future projects and goals

We will continue to hunt for regulatory motifs and may survey individual gene families in depth, and we 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. As part of the consortia DiGtoP, SyBoSS and SYSCILIA we are looking at interaction networks and systems in stem cells and primary cilia.

Structural light microscopy - single molecule spectroscopy

Edward Lemke

PhD, Max Planck Institute for Biophysical Chemistry, Göttingen.

Research Associate, the Scripps Research Institute, USA.

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

The Lemke group combines advanced microscopy with modern chemical biology tools to elucidate the nature of naturally unfolded proteins in biological systems and disease mechanisms.

Previous and current research

Currently, more than 50,000 protein structures with atomic resolution are available from the protein databank and this 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 (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 strategies for probing protein structure are incompatible with molecular disorder and the highly dynamic nature of those systems. In complex biological systems, the situation is further complicated by the mosaic of molecular states and reaction systems, where unexpected phenotypes might arise where proteins behave differently to the average. One example are 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 and can easily lead to the generation of false or insufficient models. Single molecule and superresolution techniques – which directly probe the distribution of molecular events – can help to overcome this by revealing 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.

Future projects and goals

Recent studies have shown that even the building blocks of some of the most complex and precise machines with a critical role in the survival of the cell are built from IDPs rather than 'structured' proteins. A prominent exmaple is the nuclear pore complexe (NPC), which, with an approximate molecular weight of 120 megaDalton, constitutes the largest molecular machine in the cell. Malfunctions of this machinery are linked to diseases ranging from leukemia to HIV. The NPC is built from hundreds of proteins and can regulate the transport of ≈ 1000 proteins per second in a single pore. But how such floppy protein networks are formed and maintain high specificity, high exchange rates, low error rates and diversity at the same time, challenges our views on how nature designs biological machineries. Examples such as the NPC, but also DNA packing (Chromatin) and epigenetic mechanisms call for new approaches in biology. Consequently, we aim to: i) combine protein engineering tools with diverse high/superresolution and single molecule fluorescence techniques to explore the physical and molecular rationale behind the fundamental role of IDPs; ii) extend the spectrum of chemical biology and protein engineering tools to overcome limitations in single molecule fluorescence studies – with one of our primary strategies to genetically encode unnatural amino acids; iii) continue to develop new methods and recruit techniques from other disciplines (such as microfluidics) to meet our goals.

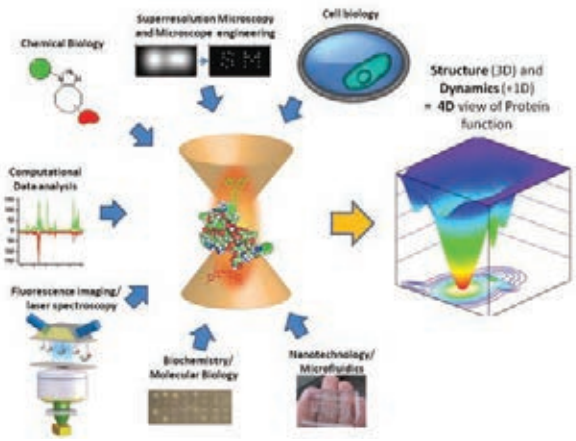
Selected references

Milles, S. & Lemke, E.A. (2011). Single molecule study of the intrinsically disordered FG-Repeat nucleoporin 153. *Biophys. J.*, 101, 1710-19

Lemke, E.A. & Schultz, C. (2011). Principles for designing fluorescent sensors and reporters. *Nat. Chem. Biol.*, 7, 480-3

Plass, T., Milles, S., Koehler, C., Schultz, C. & Lemke, E.A. (2011). Genetically encoded copper-free click chemistry. *Angew Chem. Int. Ed. Engl.*, 50, 3878-81

Gambin, Y., Vandelinder, V., Ferreon, A.C., Lemke, E.A., Groisman, A. & Deniz, A.A. (2011). Visualizing a one-way protein encounter complex by ultrafast single-molecule mixing. *Nat. Methods*, 8, 239-41



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

Architecture and regulation of metabolic networks

The Patil group uses a combination of modelling, bioinformatics, and experimental approaches to study metabolic networks and how they are controlled.

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.

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

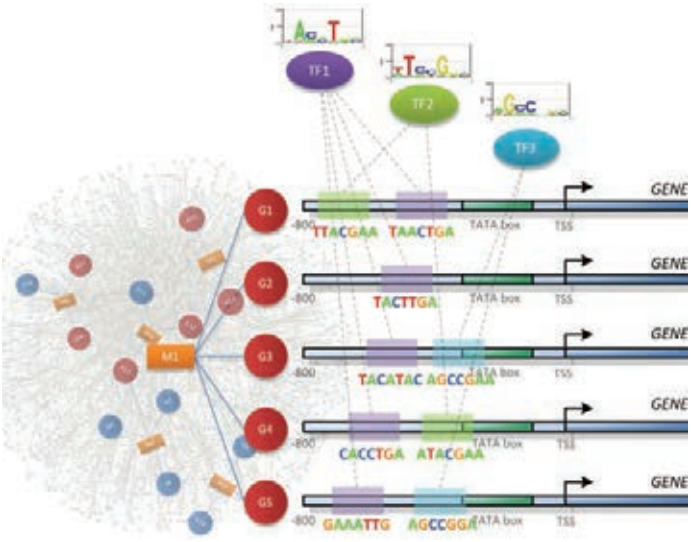
Cakir, T., Patil, K.R., Onsan, Z., Ulgen, K.O., Kirdar, B. & Nielsen, J. (2006). Integration of metabolome data with metabolic networks reveals reporter reactions. *Mol. Syst. Biol.*, 2, 50

Patil, K.R. & Nielsen, J. (2005). Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc. Natl. Acad. Sci. USA*, 102, 2685-89

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.



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 and squares - transcription factors and corresponding binding motifs

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.

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

Carsten Sachse

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

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

Group leader at EMBL since 2010.

The Sachse group uses electron cryo-microscopy to study protein aggregates, typical of neurodegenerative diseases such as Alzheimer's, and the mechanisms cells normally use to eliminate them.

Previous and current research

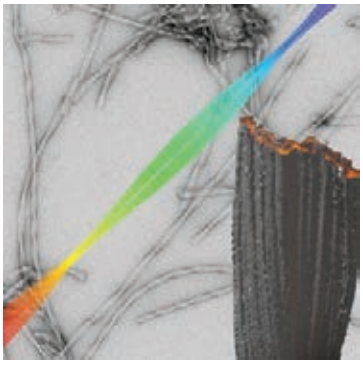
The molecular hallmark of neurodegenerative diseases, such as Alzheimer's and Parkinson's, is the formation of large protein aggregates called amyloid fibrils. The net build-up of these fibrillar aggregates is a result of an imbalance in the cellular production and clearance of misfolded polypeptides. 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?



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

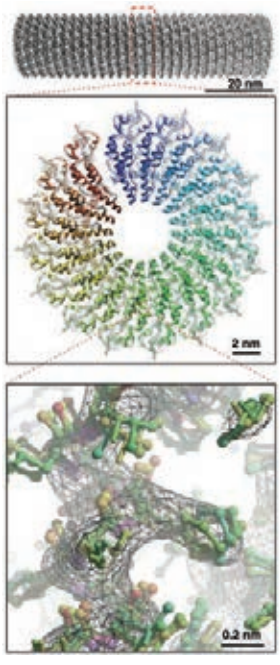
Selected references

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, 1321-3

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, 1342-52

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, 7462-6

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, 812-35



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

Data integration and knowledge management

Scientists in the Schneider team work to capture and centralise knowledge and optimise mining, browsing, and navigation of biological information.

Reinhard Schneider

PhD 1994, University of Heidelberg, Germany.
Postdoctoral research at EMBL.

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

Team leader at EMBL since 2004.

Head of Bioinformatics Core Facility, Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, since May 2011.

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.

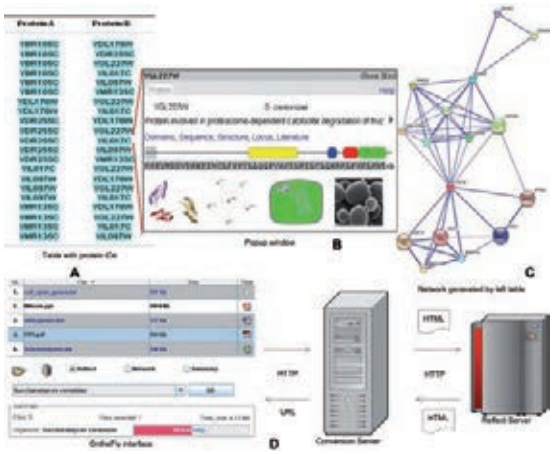
Selected references

Pavlopoulos, G.A., Secrier, M., Moschopoulos, C.N., Soldatos, T.G., Kossida, S., Aerts, J., Schneider, R. & Bagos, P.G. (2011). Using graph theory to analyze biological networks. *BioData Min.*, 4, 10

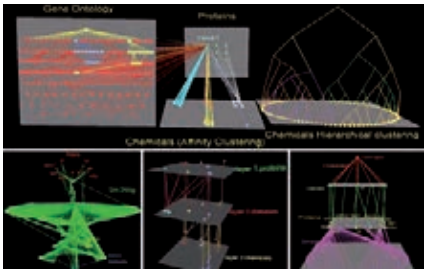
O'Donoghue, S.I.; Horn, H.; Pafilis, E.; Haag, S.;, Kuhn, M.; Satagopam, V.P.; Schneider, R., Jensen, L.J. (2010). Reflect: A practical approach to web semantics, *Web Semantics*. 8, 182-9

Gehlenborg, N., O'Donoghue, S.I., Baliga, N.S., et al. (2010). Visualization of omics data for systems biology. *Nat. Methods*, 7, S56-S68

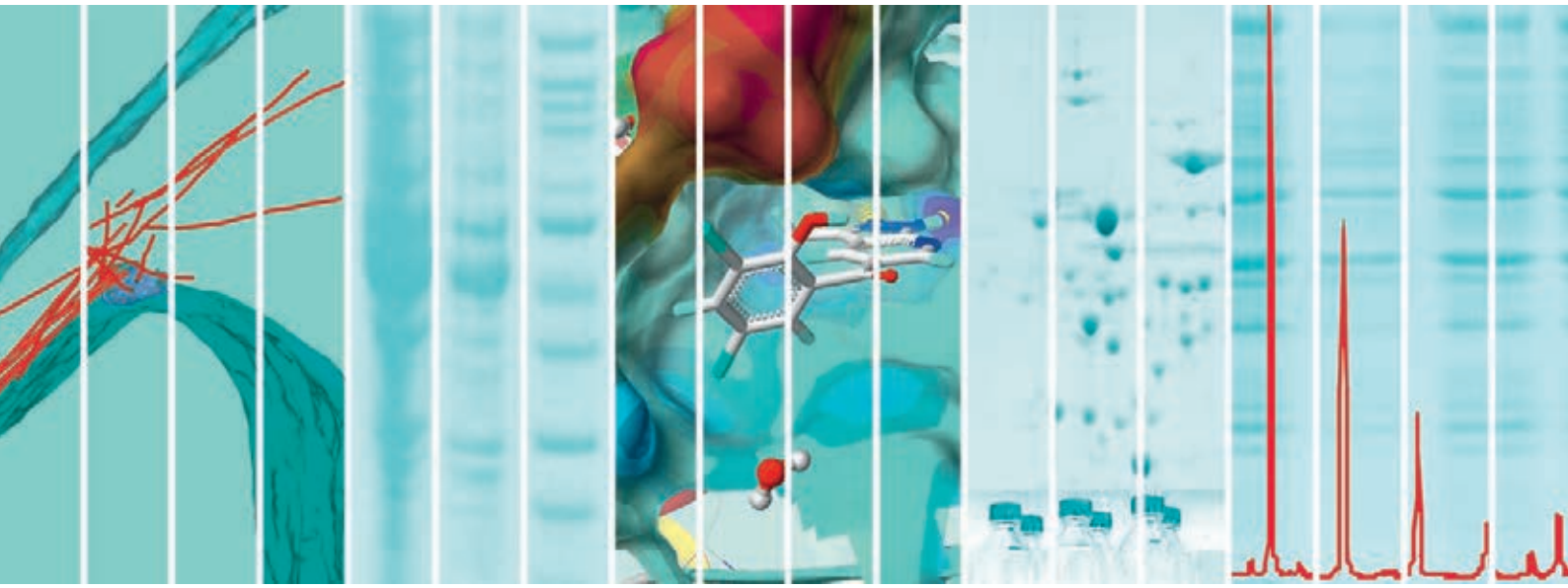
Soldatos, T.G., O'Donoghue, S.I., Satagopam, V.P., Jensen, L.J., Brown, N.P., Barbosa-Silva, A. & Schneider, R. (2010). Martini: using literature keywords to compare gene sets. *Nucleic Acids Res.*, 38, 26-38



Above: 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



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



Core Facilities

The EMBL model for Core Facilities has developed a first-rate reputation in the European life sciences community. The Core Facilities 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.

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, 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'.

Christian Boulin
Director, Core Facilities and Services

Advanced Light Microscopy Core Facility

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.

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 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 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.

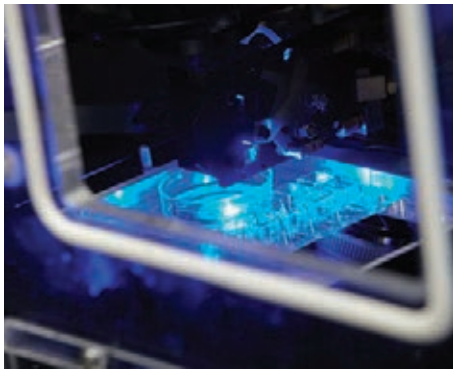
Selected references

Conrad, C., Wunsche, A., Tan, T.H., Bulkescher, J., Sieckmann, F., Verissimo, F., Edelstein, A., Walter, T., Liebel, U., Pepperkok, R. & Ellenberg, J. (2011). Micropilot: automation of fluorescence microscopy-based imaging for systems biology. *Nat. Methods*, 8, 246–9

Neumann, B., Walter, T., Heriche, J.K., *et al*. (2010). Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature*, 464, 721-7

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

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



The ALMF manages 19 advanced microscope systems and nine High-Content Screening microscopes

Chemical Biology Core Facility

The facility can assist groups in developing primary and secondary assays for screening against the in-house compound library and guide them in developing tool compounds for their specific target.

Joe Lewis

PhD 1991, Institute of Molecular Pathology, Vienna.

Postdoctoral research at EMBL.

Group and Global HCV project leader at Anadys Pharmaceuticals, Heidelberg.

MBA 2008 Mannheim Business School.

Facility head at EMBL since 2004.

Small molecules play essential roles in many areas of basic research and are often used to address important biological questions. 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 (since February 2012) 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 approximately 80 000 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
- computational screening using ligand-based and structure-based design strategies.

Partners

- Technology partners: Perkin Elmer, IDBS, GE, TTP Labtech.
- Chemistry partners: Tripos L.P., AMRI, Chembridge and Enamine.

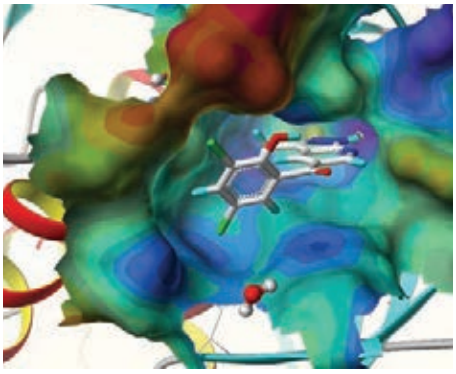
Selected references

Bartonova, V., Igonet, S., Sticht, J., Glass, B., Habermann, A., Vaney, M.C., Sehr, P., Lewis, J.D., Rey, F.A. & Krausslich, H.G. (2008). Residues in the HIV-1 capsid assembly inhibitor binding site are essential for maintaining the assembly-competent quaternary structure of the capsid protein. *J. Biol. Chem.*, 283, 32024-33

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, 500-6

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

Sehr P, Pawlita M, Lewis, J.D. (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



Ligand docked into target protein

Electron Microscopy Core Facility

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.

The facility provides training on sample preparation and use of advanced electron microscopes, and develops approaches for cell and developmental biology.

The EMCF gives EMBL scientists the chance to learn sample preparation for EM and provides training on advanced electron microscopes and instrumentation, such as electron tomography (ET) setup (F30 300 kV TEM). Techniques can be applied and adapted to achieve EM-resolution at the level of cell organisation. We are also developing correlative microscopy approaches for cellular and developmental biology.

Major projects and accomplishments

Our electron tomography equipment, which has been 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, but can also be used as a cryo-microscope. Specialised EM engineers manage the F30, 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.

We have upgraded our Correlative Light & Electron Microscopy (CLEM) technology with a new Zeiss LM microscope (Zeiss-Observer Z1), situated next to a high-pressure freezing machine (HPM-010) and devoted to perform CLEM, using cryofixation of our samples. New projects are therefore welcome using this technology, possibly in combination with the methods introduced at EMBL by the Briggs group (see page 58).

External collaborations include the study of SPB duplication in meiotic fission yeast (Kayoko Tanaka, University of Leicester); spindle formation in vertebrate cells (Maria Koffa, Democritus University of Thrace); SPBs in fission yeast (Anne Paoletti, Institut Curie-UMR144 CNRS); budding yeast SPBs (Schiebel group, DKFZ–ZMBH); and centriole biogenesis (Hoffmann group, DKFZ). We also support investigations on *in vivo* dengue virus replication (Bartenschlager group, Heidelberg University).

Services provided

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

Technology partners

FEI Company (advanced electron microscopes, including the new tomography microscope), Leica Microsystems (supplied our portable EMPACT2 HP Freezer, as well as ultramicrotomes units for plastic samples or cryo-sectioning) and Zeiss (for a light microscopy setup devoted to correlative LM/EM microscopy).

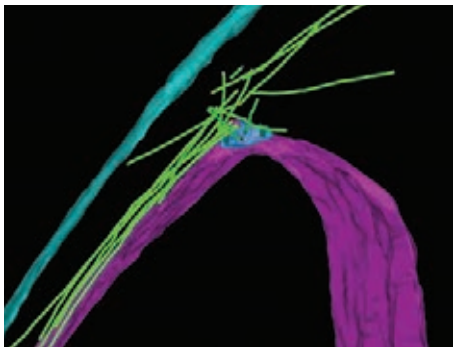
Selected references

De Fazio, S., Bartonicek, N., Di Giacomo, M., *et al*. (2011). The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature*, 480, 259-63

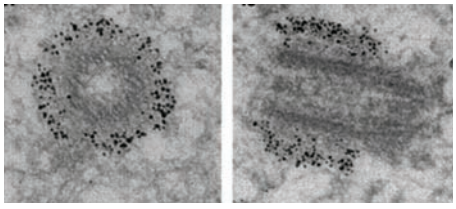
Tängemo C., Ronchi P., Colombelli J., *et al*. (2011). A novel laser nanosurgery approach provides support for *de novo* Golgi biogenesis in mammalian cells. *J. Cell Sci.*, 124, 978-87

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, 731-9

Reuter, M., Berninger, P., Chuma, S., Shah, H., Hosokawa, M., Funaya, C., Antony, C., Sachidanandam, R. and Pillai, R.S. (2011). MIWI catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature*, 480, 264-7



Model of meiotic SPB-Microtubule interaction in S.pombe *cells from a tomography 3D reconstruction using IMOD (Univ. Boulder CO). (Picture by Charlotta Funaya, collaboration with Kayoko Tanaka, University of Leicester)*



Immunogold labelling of Cep152, a centriolar protein in Hela cells. (Picture by Uta Haselmann, collaboration with Ingrid Hoffmann, DKFZ. J. Cell Biol. *191:731-9)*

Flow Cytometry Core Facility

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

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.

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 Zn2+ 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

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.

Major projects and accomplishments

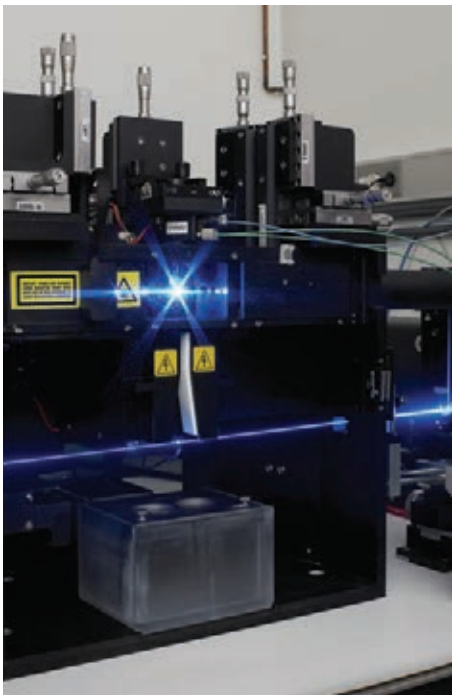
- The analysis of algae life cycle project relied on the intrinsic fluorophores in algae to identify life cycle stages. It required photosaturation of the photosynthetic units in the algae.
- Cell cloning by Darwinian selection required a series of single cell sorts of a target population into 96-well 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 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.



The facility provides key services, such as sorting heterogeneous single cell populations into homogenous populations for experiments

Genomics Core Facility

Vladimír Beneš

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

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.

Major projects and accomplishments

GeneCore provides its services to a broad range of users ranging from small research groups to large consortia (ICGC, for example). Training is a crucial aspect of our work and staff tutor individual researchers and organise practical courses on corresponding subjects. The acquisition of new-generation sequencing technology was a vital step to ensure EMBL remains at the forefront of European research. Since 2010, our massively parallel sequencing (MPS) suite has been upgraded and expanded significantly and now includes HiSeq2000 and cBot instruments as well as an Ion Torrent instrument. Preparation of MPS libraries for various applications is facilitated by a robust framework of other instruments (e.g. Covaris, Bioanalyzer, Qubit).

GeneCore provides the following analyses in a single- or a pair-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 (including strand-specific libraries)
- discovery of small non-coding RNAs: ncRNA-Seq
- genome-wide DNA methylation analysis: Methyl/BS-Seq
- *de novo* sequencing & resequencing of genomic DNA
- targeted enrichment (sequence capture) in solution coupled with MPS.

We continue 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 2 terabases of MPS sequence 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, upon demand, customised arrays.

In addition to three qPCR instruments managed by GeneCore, in 2011 GeneCore's qPCR capacity has been considerably enhanced by a Fluidigm Biomark HD instrument – a device capable of the quantitation of transcripts on a single-cell level.

Technology partners

MPS is still a very dynamic and rapidly evolving technology. We collaborate with several MPS-related companies and test their products in our workflows. GeneCore is a member of the early-access program of Illumina, Agilent, NuGEN, Beckman Coulter and Epicentre.

Selected references

Sanchez, M., Galy, B., Schwanhaeusser, B., Blake, J., Bahr-Ivacevic, T., Benes, V., Selbach, M., Muckenthaler, M.U. & Hentze, M.W. (2011). Iron regulatory protein-1 and -2: transcriptome-wide definition of binding mRNAs and shaping of the cellular proteome by iron regulatory proteins. *Blood*, 118, e168-e179

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J. & Wittwer, C.T. (2011). Primer Sequence Disclosure: A Clarification of the MIQE Guidelines. *Clin. Chem.*, 57, 919-921

Castoldi, M., Vujic Spasic, M., Altamura, S., et al. (2011). The liver-specific microRNA miR-122 controls systemic iron homeostasis in mice. *J. Clin. Invest.*, 121, 1386–1396

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.*, 39, 2073-91

The facility provides access to cutting-edge instruments and provides full user support

Protein Expression and Purification Core Facility

The facility produces and purifies proteins from E. coli, *insect, mammalian cells and sera, using a variety of chromatographic methods.*

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.

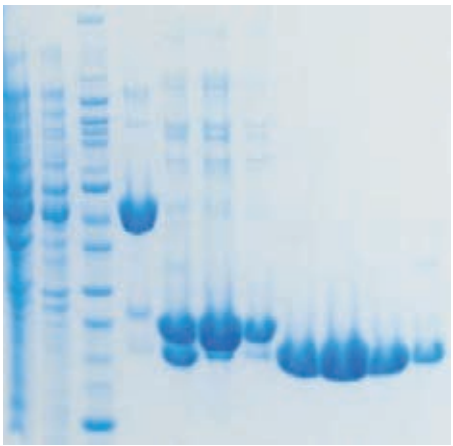
Selected references

Mackereth, C.D., Madl, T., Bonnal, S., Simon, B., Zanier, K., Gasch, A., Rybin, V., Valcarcel, J. & Sattler, M. (2011). Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature*, 475, 408-11

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

Hothorn, M., Neumann, H., Lenherr, E.D., *et al*. (2009). Catalytic core of a membrane-associated eukaryotic polyphosphate polymerase. *Science*, 324, 513-6

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



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

Following each purification, we can perform biophysical analyses to ensure the quality of the purified sample in terms of correct folding and stability. Our group also develops or evaluates new techniques and advanced protocols for protein production and purification and there is significant focus on 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 pETM-series expression vectors for *E. coli* that can now be used for sequence and ligation-independent cloning (SLIC). We have adapted vectors for insect and mammalian cells for the same cloning protocol. Using a single PCR product with the gene of interest, you can integrate the insert into all of the vectors due to the universal overlaps that are present in the linearised vectors and the PCR product. A lethal gene insert in the original template vectors inhibits the growth of false positive colonies, which reduces the number of clones to test for the correct insert. With this new vector set, one can test the expression of a gene in 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 can obtain pure, untagged proteins that are otherwise difficult to express or purify and can be used, for example, as antigens for immunisation.

Services provided

- Expression and purification: 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 marker for general use within EMBL.
- Developing and testing new vectors and protocols.
- Access to protocols and vector sequence information on the website.
- Scientific and technical advice to users at EMBL and external researchers.
- Caring for equipment for protein production and analysis.
- Providing quality analysis and biophysical characterisation of purified proteins (e.g. analytical ultracentrifugation).

Technology partners

We are open to collaborations with academic or industrial partners to evaluate new products or technological developments. Furthermore, we have initiated a network of protein facilities across Europe (Protein Production and Purification Partnership in Europe (P4EU)) to improve information exchange and evaluation of new technologies.

Proteomics Core Facility

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.

The Proteomics Core Facility provides a full proteomics infrastructure for the identification and characterisation of proteins.

Infrastructure in the the facility is centered on state-of-the-art mass spectrometry for MS and LC-MSMS experiments. It 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 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 (Waters Q-tof).
- Determination of N- and C-termini of proteins and products of limited proteolysis.
- Verify incorporation of non-natural amino acids.
- Protein separation by 1- and 2-dimensional electrophoresis using gel systems from 7 to 24 cm.

Proteomics:

- Protein identification from gel or in solution.
- Protein quantification by stable-isotope labelling (SILAC and dimethyl labelling).
- Identification of post-translational modifications.
- Enrichment of phosphopeptides (TiO2 and IMAC).
- 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 Pro) for identification and quantification of proteins in complex mixtures.
- Multi-dimensional peptide separation (isoelectric focusing and liquid chromatography).
- Pipeline for differential proteome analysis by 2D gels (incl. DIGE).

Technology partner

- BIO-RAD

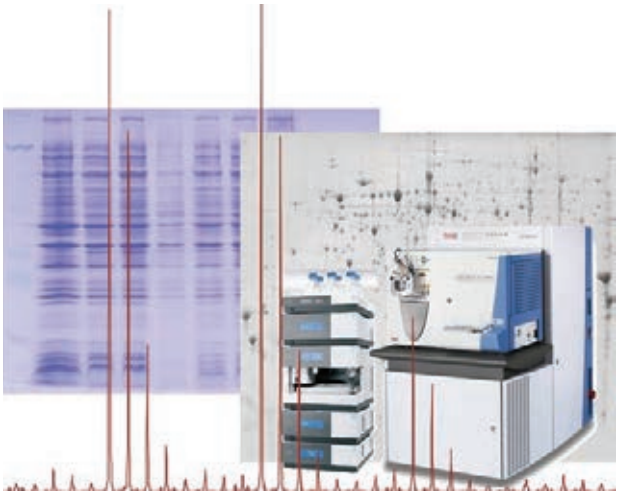
Selected references

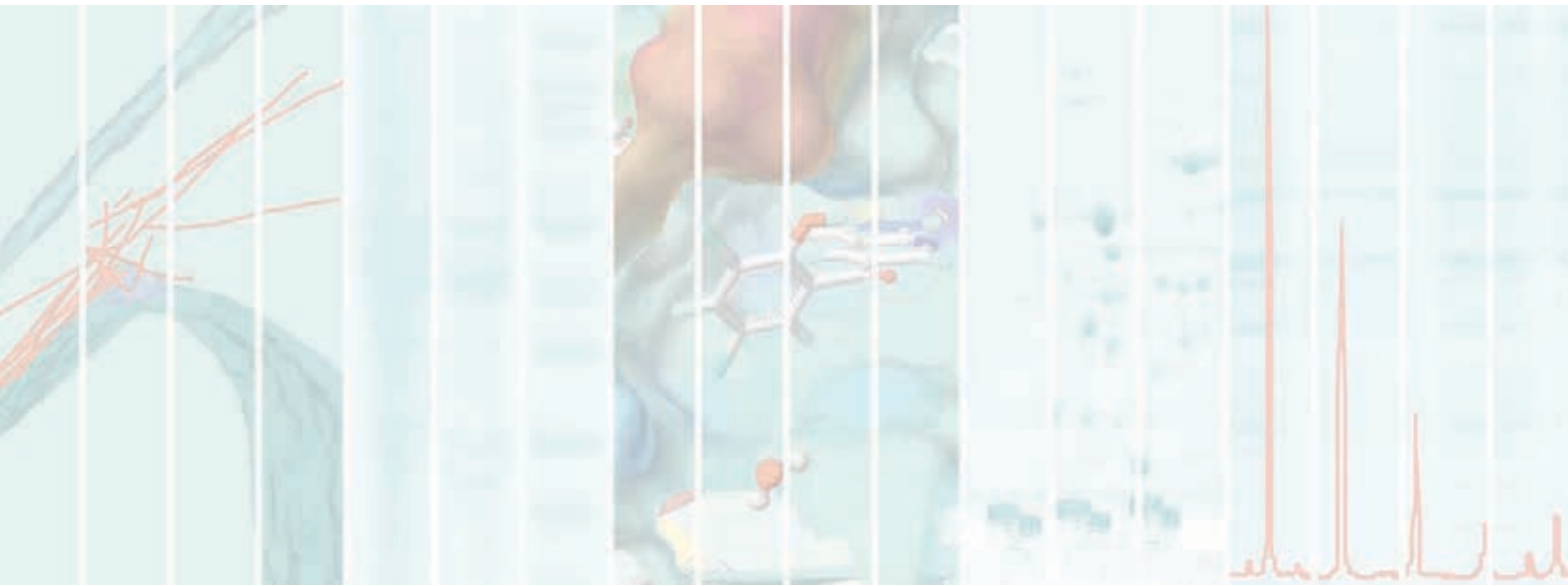
Lannert, H., Franz, T., Able, T., Leicht, S., Lenze, M., Eckstein, V., Saffrich, R., Ho, A.D. (2008). Regulation of migration - Oncoprotein (Op18, Stathmin) overexpression in G-CSF mobilized stem cells. *Onkologie*, 31

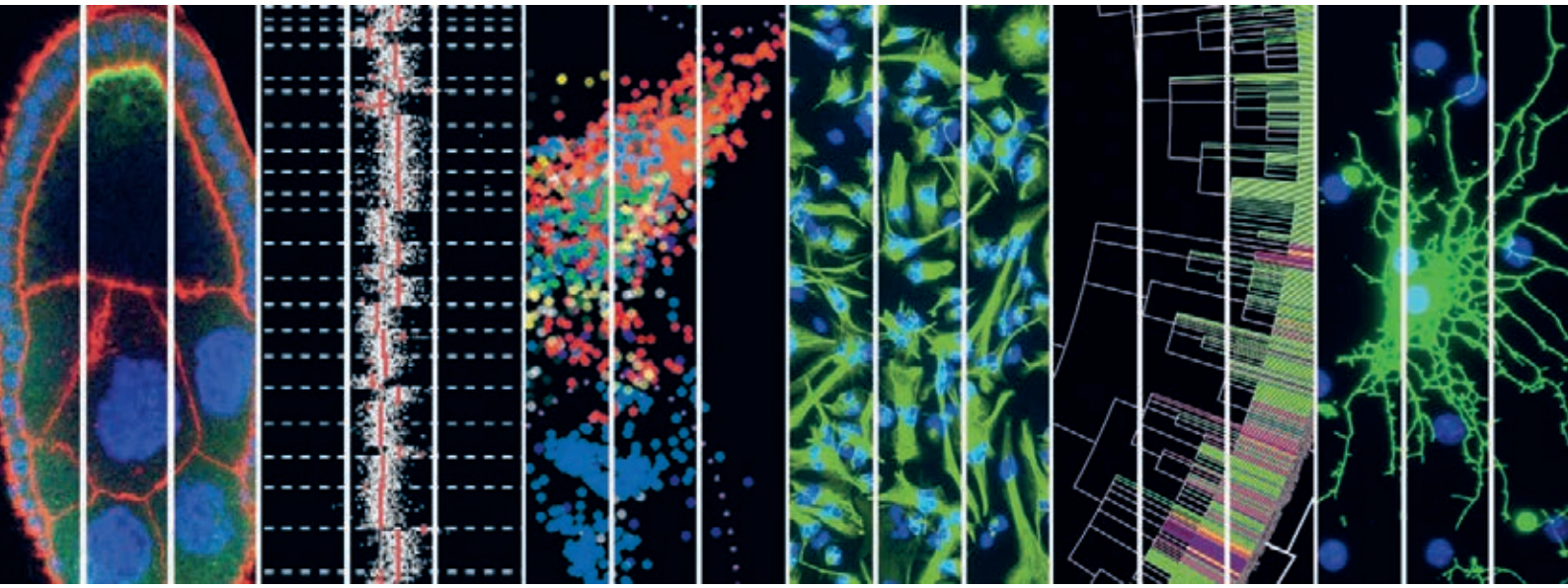
Koyama, S., Hata, S., Witt, C.C., Ono, Y., Lerche, S., Ojima, K., *et al*. (2008). Muscle RING-finger protein-1 (MuRF1) as a connector of muscle energy metabolism and protein synthesis. *J. Mol. Biol*., 376, 1224-36

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, 43-6

Funk, J., Li, X. & Franz, T. (2005). Threshold values for detergents in protein and peptide samples for mass spectrometry. *Rapid Commun. Mass Spectrom.*, 19, 2986-8







European Bioinformatics Institute

The EBI is probably best known worldwide for its provision of biological information and bioinformatics services. However, about 20% of the institute is devoted to investigator-led research using computational approaches to unravel the secrets of life.

The development of new technologies provides a constant driver for innovative research into processing and analysing the data generated. For example, the wide uptake of next-generation sequencing by life scientists has led to unprecedented growth in sequence data. These data require novel algorithms to turn them into reliable information, and perhaps even more challenging is to use these new data to obtain novel insights into biological processes.

Research at EBI is carried out both in groups devoted solely to research and in some of the larger service teams that have associated research activities. All researchers have computational approaches as their major focus, but most also collaborate closely with experimentalists and often generate experimental data themselves. Our research is highly collaborative within EMBL as well as with many external colleagues. We are highly interdisciplinary; our faculty comprises scientists who originally trained in biology, physics, chemistry, engineering, medicine or mathematics. We develop novel algorithms and protocols for handling data, such as checking the quality of the data; interpret data and integrate data to generate new knowledge. We use this information to develop novel hypotheses about the basic molecular processes of life.

Although we are united in using computers, the biological questions we address and the algorithms we develop and use are very diverse. We explore biological questions spanning genome evolution, transcriptional regulation and systems modelling of basic biological processes and disease. For example, different groups are investigating the molecular basis of ageing; the differentiation of stem cells; the basis for neuronal plasticity; and the early development of brain structure. Others are exploring regulation through epigenetics or RNA processing; how phenotype is related to genotype both in mice and humans; and how new enzyme reactions appear during evolution. All our discoveries are published in peer-reviewed journals but in addition, as part of these studies, our researchers often develop novel bioinformatics services, which are usually made freely available for all users so that our work helps facilitate new discoveries throughout the global scientific community.

Increasingly, much of our work is related to problems of direct medical significance and with the emergence of personal genomes we are very conscious of the need to contribute to the translation of the new knowledge into medicine and the environment. This process is just beginning and will provide many challenges to computational biologists over the coming years.

Janet Thornton
Director, EMBL-EBI

Computational biology of proteins: structure, function and evolution

Janet Thornton

PhD 1973, King's College and NIMR.

Postdoc at Oxford University, NIMR and Birkbeck College. Lecturer, Birkbeck College, 1983-1989.

Prof. of Biomolecular Structure, UCL since 1990.

Bernal Prof. at Birkbeck College, 1996-2002.
Director, Centre for Structural Biology, Birkbeck College and UCL, 1998-2001.

Director of EMBL-EBI since 2001.

The Thornton group aims to understand more about the 3D structure and evolution of proteins, for example studying how enzymes perform catalysis, and the insulin signalling pathway in ageing.

Previous and current research

The goal of our research is to understand more about how biology works at the molecular level, with a particular focus on proteins and their 3D structure and evolution. Currently we are exploring how enzymes perform catalysis, involving the gathering of relevant data from the literature and the development of 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. 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 understanding 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 or mice and relate these observations to effects on life span, by combining information on function, context (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, analysis of survival curves and combining transcriptome data sets with network analysis for flies, worms and mice, to compare the different pathways and ultimately explore effects related to human variation and age.

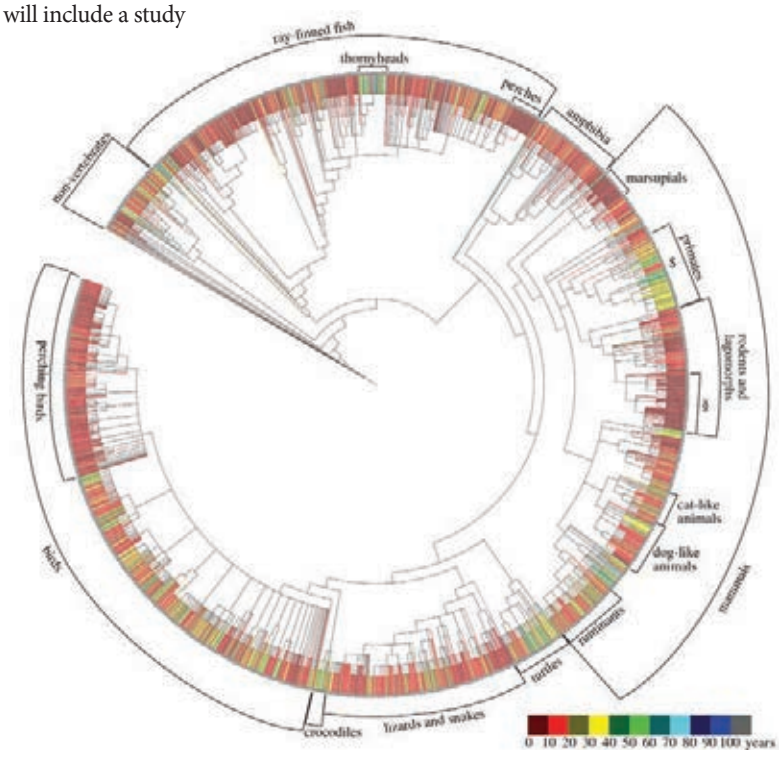
Selected references

Furnham, N., Sillitoe, I., Holliday, G.L., *et al*. (2012). FunTree: A resource for exploring the functional evolution of structurally defined enzyme superfamilies. *Nucleic Acids Res.*, 40, D776-D782

Holliday, G.L., Fischer, J.D., Mitchell, J.B.O., *et al*. (2011). Characterizing the complexity of enzymes on the basis of their mechanisms and structures with a bio-computational analysis. *FEBS J.*, 278, 3835-45

Holliday, G. L., Andreini, C., Fischer, J.D., *et al*. (2011). MACiE: Exploring the diversity of biochemical reactions. *Nucleic Acids Res.*, 40, D783-D789

Wieser, D., Papatheodorou, I., Ziehm, M., *et al*. (2011). Computational biology for ageing. *Philos. Trans. R. Soc. Lond. B Biol. Sci*., 366, 51-63



Bioinformatics analysis of phylogenetic clustering of maximal lifespan records. The figure was created with the help of the iTOLWebservice using longevity data from AnAge and phylogeny data from NCBI Taxonomy. Similar maximal lifespans can be found in phylogenetically distant species. Within the large variety of maximal lifespans in evolutionary clades, clusters of similar maximal lifespan can be observed. This hints to a strong genetic influence on ageing. $ indicates Callitrichinae (a family of New World monkeys), § indicates mouse-like animals

Protein, cheminformatics and metabolism data

The Apweiler team is responsible for the protein and proteomics-related activities at EMBL-EBI, which include the production of protein sequence and protein family databases.

Rolf Apweiler

PhD 1994 University of Heidelberg.

At EMBL since 1987, at EMBL-EBI since 1994. Joint team leader, Protein and Nucleotide Data Group (Proteins).

Selected references

Alam-Faruque, Y., Huntley, R.P., Khodiyar, V.K., et al. (2011). The impact of focused Gene Ontology curarion of specific mammalian systems. *PLoS One* 6, e27541

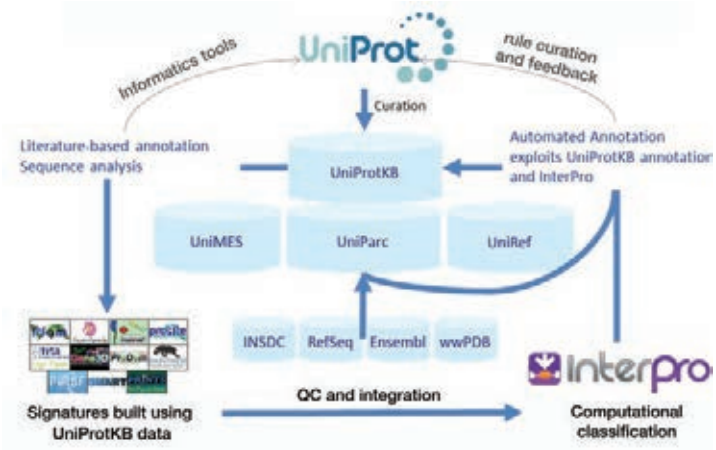
O'Donovan, C., and Apweiler, R. (2011). A guide to UniProt for protein scientists. *Methods Mol. Biol.* 694, 25-35

Previous and current research

The Protein and Nucleotide Data group, (led jointly by Rolf Apweiler and Ewan Birney) focuses on the production of protein-sequence, protein-family and nucleotide-sequence databases at EMBL-EBI. The team maintains and hosts the UniProt protein resource, the InterPro domain resource and a range of other biomolecular databases. These efforts fall into two major categories: proteins, and cheminformatics and metabolism. The group has a complementary research component.

Our activities 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 European Nucleotide Archive (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 as well as automatic annotation approaches. Rolf Apweiler supervises one PhD student directly.



UniProt sequence and annotation data flow

Future projects and goals

In 2012 we intend to improve integration and synchronisation of all EMBL-EBI 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 results; and applying the new methodologies to UniProtKB/TrEMBL annotation.

Pluripotency, reprogramming and differentiation

Paul Bertone

PhD 2005, Yale University.

At EMBL-EBI since 2005.
Joint appointments in Genome Biology and Developmental Biology Units.

The Bertone group investigates the cellular and molecular processes underlying mammalian stem cell differentiation and induced pluripotency.

Previous and current research

We investigate the cellular and molecular processes underlying mammalian stem cell biology 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.

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

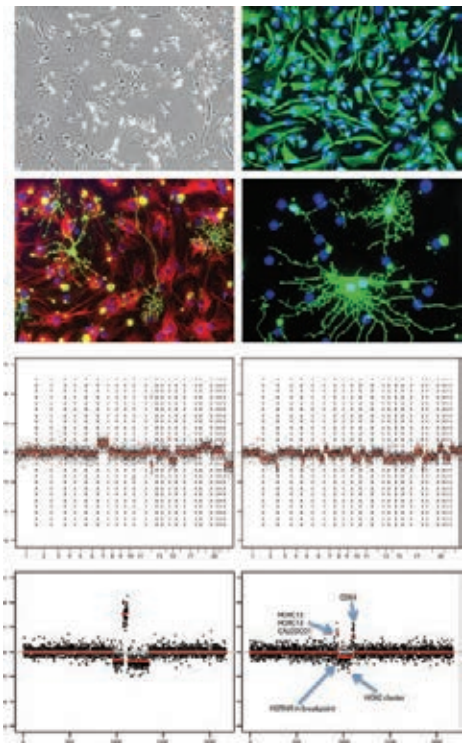
We will continue working to understand the molecular mechanisms that support pluripotency in ground-state embryonic stem cells, and to map the transition between the pluripotent state and early lineage commitment. We also plan to use the ChIP-seq approach to capture the epigenetic status of cells undergoing 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.

Selected references

Reynolds, N., *et al*. (2011) NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. *EMBO J.* 31, 593-605

Albers, C., *et al*. (2011) Exome sequencing identifies NBEAL2 as the causative gene for gray platelet syndrome. *Nat. Genet.,* 43, 735-7

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



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)

Nucleotide and genome data

The Birney group's research focuses on developing sequence algorithms and using intra-species variation to study basic biology.

Ewan Birney

PhD 2000, Sanger Institute.

At EMBL since 2000.
Joint team leader, Protein and Nucleotide Data Group (Nucleotides).

Selected references

Birney, E. (2011). Chromatin and heritability: how epigenetic studies can complement genetic approaches. *Trends in Genetics*, 27, 172-76

Fritz , M., Leinonen, R., Cochrane, G., *et al*. (2011). Efficient storage of high-throughput DNA sequencing data using reference-based compression. *Genome Research*, 21, 734-40

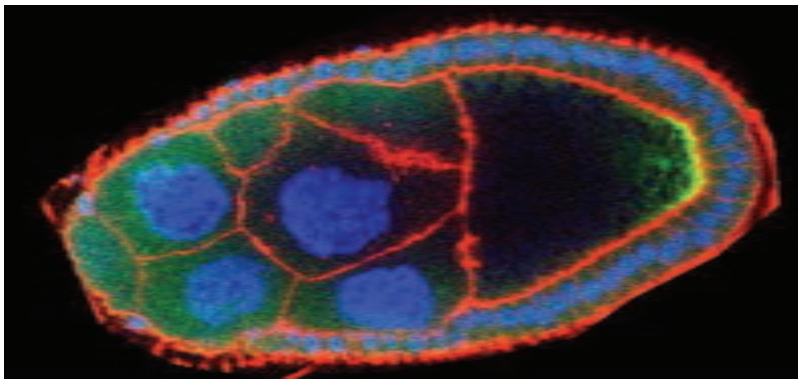
Overview of research

DNA sequence remains at the heart of molecular biology and bioinformatics. Our work focuses on developing algorithmic methods for genome analysis and the use of genetic association techniques to understand basic biology. Our work provides strategic oversight across all nucleotide resource groups. The three main branches of activity in our 'Nucleotides' team are: vertebrate genomics (Ensembl); non-vertebrate genomics (Ensembl Genomes); and nucleotide sequences (European Nucleotide Archive).

Development of sequence algorithms: Over time students have developed multiple alignment, assembly and more recently sequence compression tools for DNA sequences. In the future we hope 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.*



Functional genomics

Alvis Brazma

PhD in Computer Science 1987, Moscow State University.

Postdoctoral research at New Mexico State University, USA.

At EMBL-EBI since 1997.

Research in the Brazma team focuses on analysis of gene expression, development of new algorithms and methods for integrative data analysis, and building a global map of gene expression.

Previous and current research

Our teams focus on functional genomics data services, analysis of data from high-throughput sequencing and gene expression experiments; and research and development related to biomedical informatics and systems microscopy. We run several of EMBL-EBI's core resources: the ArrayExpress Archive of Functional Genomics Data, the Gene Expression Atlas and the BioSample Database.

Our PhD students focus on data analysis, building models for systems biology and developing new methods and algorithms. Integration of data across multiple platforms and types of data is another important area of activity. We also contribute substantially to training in transcriptomics, ontologies and the general use of EBI tools.

Our on-going research projects are related to regulation of gene expression and analysis of large-scale functional-genomics data. In addition to publishing several papers on RNA-seq analysis, our team has made advances in developing methods for gene expression data meta-analysis, in particular for discovery of diabetes candidate genes and new integrative analysis of RNAseq data. Part of our work involves analysing RNAseq and DNA data from kidney and other cancers, as a part of the International Cancer Genome Consortium. In collaboration with Helsinki University of Technology, we have developed and published new methods for data-driven information retrieval.

Future projects and goals

In 2012 we will work to develop methods for RNA-seq data analysis and processing, and apply these to address important biological questions such as ubiquity of gene expression, the role of alternative splicing and splicing mechanisms. With our collaborators from the International Cancer Genome Consortium we will be seeking new insights into cancer genomes and their impacts on functional changes in cancer development. In this area we will focus on discovery and analysis of fusion genes and their role in cancer development.

Selected references

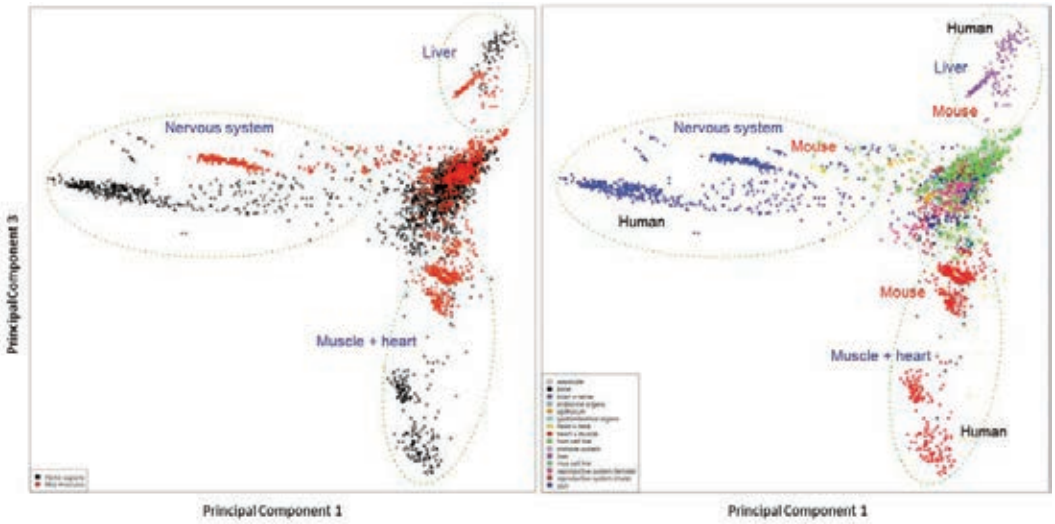
Caldas, J., Gehlenborg, N., Kettunen, E., *et al*. (2012). Data-driven information retrieval in heterogeneous collections of transcriptomics data links SIM2s to malignant pleural mesothelioma. *Bioinformatics*, 28, 246-53

Nicholson, G., Rantalainen, M., Li, J. V., *et al*. (2011). A genome-wide metabolic QTL analysis in Europeans implicates two loci shaped by recent positive selection. *PLoS Genet.*, 7, e1002270

Kutter, C., Brown, G. D., Gonçalves, A., *et al*. (2011). Pol III binding in six mammals shows conservation among amino acid isotypes despite divergence among tRNA genes. *Nat. Genet.*, 43, 948-55

ENCODE Project Consortium, *et al*. 2011. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol*. 9, e1001046

A combined human and mouse gene-expression data matrix (principal components 1 and 3). Each dot represents a sample, which is labelled by (a) species and (b) tissue type



Functional genomics and analysis of small RNA function

The Enright group focuses on small non-coding RNAs and develops computational tools, systems, and algorithms to predict the function and interactions of small RNAs.

Anton Enright

PhD in Computational Biology 2003, University of Cambridge.

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

At EMBL-EBI since 2008.

Selected reference

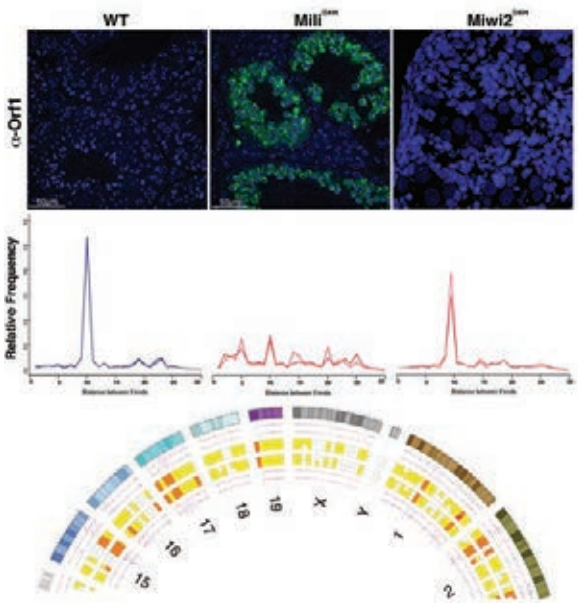
De Fazio, S., Bartonicek N., Di Giacomo M., *et al*. (2011). The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature*, 480, 73-6.

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 unannotated 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 involves the development of algorithms, protocols and datasets for functional genomics. We focus on determining the functions of regulatory RNAs including microRNAs, piwiRNAs and long non-coding RNAs. We collaborate extensively with experimental laboratories on commissioning experiments and analysing experimental data. Some laboratory members take advantage of these close collaborations to gain hands-on experience in the wet lab.

Future projects and goals

Our long-term goal is to combine regulatory RNA target prediction, secondary effects and upstream regulation into complex regulatory networks. Predoc Leonor Quintais will develop strategies for dealing with large-scale CLIP assays for microRNA target analysis. We will continue to build an accurate database of piRNA loci in animals and explore the importance and evolution of these molecules. We are extremely interested in the evolution of regulatory RNAs and developing phylogenetic techniques appropriate for short non-coding RNA. We will continue to build strong links with experimental laboratories that work on miRNAs in different systems. This will allow 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.



Top row: Immunoflourescent staining (green) for the protein (ORF1) expressed from LINE1 transposons when active and blue DAPI staining for nuclei in mouse testes sections.The wild-type sample shows no expression while the Mili DAH sample shows rampant transposon activity when the slicing function of Mili is mutated. The final column shows that no LINE1 activity is present when the slicing activity is mutated in Miwi2. The second row shows the frequency of 5'/3' overlaps of specific lengths from piwiRNAs sequenced from each sample. Normal 'ping-pong' activity is observed in the wild type, while this is defective in MiliDAH. Again the Miwi2DAH sample behaves similarly to wild type indicating that the slicing activity of this enzyme is not required for successful transposon silencing. The semi-circular plot shows how sequenced piwiRNAs map across a section of the mouse genome, with piwiRNA hotspots shown as orange and red areas on the heatmap

Vertebrate genomics

Paul Flicek

DSc 2004, Washington University.

Honorary Faculty Member, Wellcome Trust Sanger Institute since 2008.

At EMBL-EBI since 2005.
Team leader since 2008.

The Flicek team develops large-scale bioinformatics infrastructure, explores the evolution of transcriptional regulation and develops algorithms to study epigenomic data.

Previous and current research

Our team creates and maintains the genomic resources of the Ensembl project and is responsible for data management for a number of large-scale international projects, including the 1000 Genomes Project and, in collaboration with the Brazma team (page 82), the International Mouse Phenotyping Consortium. We also maintain and develop EMBL-EBI's major variation databases, including the European Genome-phenome Archive (EGA) and the DGVa database of copy number and structural variation. All of these resources are made publicly available and are widely used by the scientific community and by the team itself as part of our research into evolution, epigenetics and transcriptional regulation.

Our specific research projects (largely done in collaboration with Duncan Odom's group, University of Cambridge) focus on the evolution of transcriptional regulation. Recently we have expanded 'comparative regulatory genomics' techniques including mapping the same DNA–protein interactions in matched tissues in multiple species to understand how gene regulation has evolved while the tissue level functions are largely conserved. We are also interested in the role of chromatin conformation in tissue-specific gene regulation and have investigated both the CTCF and cohesin complex in this context.

Selected references

Flicek, P., Amode, M.R., Barrell, D., *et al*. (2011). Ensembl 2011. *Nucleic Acids Res*. 39, D800-D806

Lindblad-Toh, K., Garber, M., Zuk, O., *et al*. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, 478, 476-482

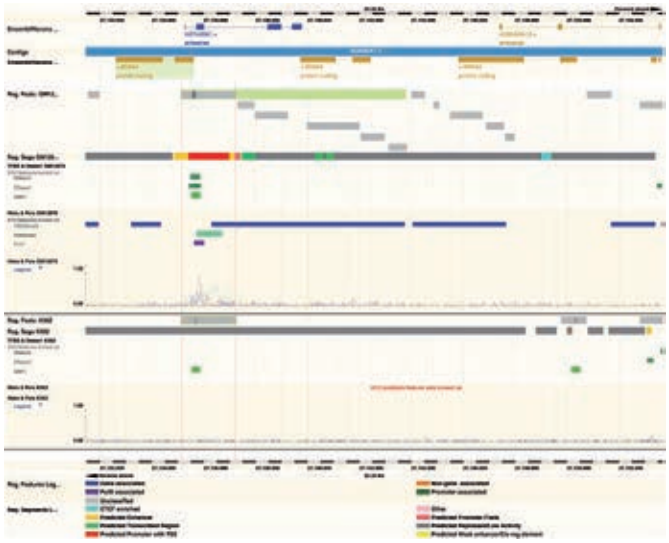
Locke, D.P., Hillier, L.W., Warren, W.C., *et al*. (2011). Comparative and demographic analysis of orang-utan genomes. *Nature*, 469, 529-33

Marth, G.T., Yu, F., Indap, A.R., *et al*. (2011). The functional spectrum of low-frequency coding variation. *Genome Biol.* 12, R84

Future projects and goals

The rapidly growing volume and diversity of data across the scope of genomics research is the major challenge for projects like Ensembl. We see an ever-increasing number of whole-genome sequences as well as comprehensive variation, regulatory, disease and phenotype data in human and other species. We have created a number of analysis and visualisation methods to summarise and present dense and complex regulation data (see figure) and will continue to develop expand these to other species and disease states. At the same time, the EU Blueprint project and the NIH-funded KOMP2 project are both expected to produce their first major data sets in 2012. This means that we will continue to play an end-to-end role in major genomics projects from raw-data management for the project to summary-data presentation to the wider scientific community.

Our research projects are expanding in number of species, tissues and specific DNA–protein interactions. We will also focus on understanding the differentiation process and components of cell- and tissue-specific regulation. These questions will be addressed both in the context of our established collaborative projects with the Odom group and as part of other collaborations, including larger EU-funded projects.



Raw, processed and summarised data from two human cell lines in region around the HOXA1 gene on chromosome 7 as presented in Ensembl. Lymphoblastoid (GM12878) and myelogenous leukemia (K562) cell lines from the ENCODE project show differences in raw data (wiggle tracks), processed data (coloured rectangular boxes in the tracks labelled DNase1) and summary level multi-coloured segmentation tracks demonstrating the difference at the chromatin and regulatory level between the active gene in GM12878 cells and the inactive gene in K562 cells

Evolutionary tools for genomic analysis

The Goldman group is developing improved mathematical and statistical methods for analysing DNA and amino acid sequences to study how these sequences evolve.

Nick Goldman

PhD 1992, University of Cambridge.

Postdoctoral work at National Institute for Medical Research, London, 1991-1995, and University of Cambridge, 1995-2002.

Wellcome Trust Senior Fellow, 1995-2006.

At EMBL-EBI since 2002.
EMBL Senior Scientist since 2009.

Selected references

Albers, C. A., Cvejic, A., Favier, R., *et al*. (2011). Exome sequencing identifies NBEAL2 as the causative gene for gray platelet syndrome. *Nature Gen*. 43, 735-737

Jordan, G, and Goldman, N. (2011). The effects of alignment error and alignment filtering on the sitewise detection of positive selection. *Mol. Biol. Evol.,* in press. Published online 1 November

Kosiol, C., and Goldman, N. (2011). Markovian and non-Markovian protein sequence evolution: aggregated Markov process models. *J Mol. Biol.*, 411, 910-923

Lindblad-Toh, K., Garber, M., Zuk, O., *et al.* (2011). A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, 478, 476-482

Previous and current research

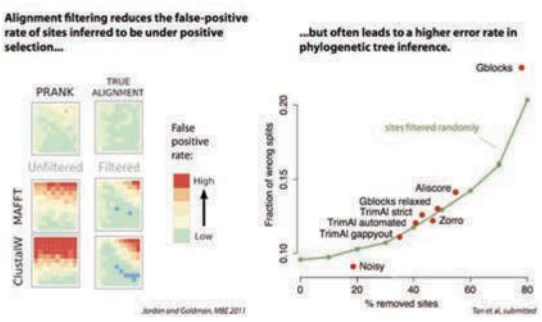
Our research concentrates on the mathematics and statistics of data analyses that use evolutionary information in sequence data and phylogenies to infer the history of living organisms, describe and understand processes of evolution and make predictions about the function of genomic sequence. We aim to increase our understanding of the process of evolution and provide new tools to elucidate the function of biological molecules as they change over evolutionary timescales. Our three main research activities are: developing new evolutionary models and methods; providing these methods to other scientists via stand-alone software and web services; and applying such techniques to tackle biological questions of interest. In recent years, collaborations with sequencing consortia have provided essential state-of-the-art data and challenges to inspire, develop, and apply novel methods. Collaborations between group members who are 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. Traditionally, the group has been strong in examining the theoretical foundations of phylogenetic reconstruction and analysis. In 2011 we continued to gain expertise in the analysis of next-generation sequencing (NGS) data. This vast source of new data promises great gains in understanding genomes and brings with it many new challenges.

Future projects and goals

In 2012 we plan to: further characterise common sources of alignment error in genome-wide studies; develop new methods for detecting and mitigating such errors; and design more advanced methods for simulating protein-coding sequences. At a theoretical level, we will complete and draw conclusions from an ongoing study in phylogenetic inference, of 'long-branch attraction', which refers to the apparent propensity of long branches to 'congregate' in estimated phylogenies. Specifically, we hypothesize that long-branch attraction might result from the accumulation of independent deviations attributable to each long branch but not arising from any interactions among them. We will also attempt to assess the irreversibility of substitution processes observed in real biological sequences by comparing similarly parametrised reversible and irreversible Markov models.

We will continue to analyse raw data from emerging sequencing technologies, with an eye to improving the reads produced and understanding the correct way to represent the inherent uncertainties of the data in downstream analyses. The remaining time on our Wellcome Trust sequencing technologies grant will be spent looking at whether techniques for spectral correction of short reads can also be applied to recalibration.

As high-throughput RNA sequencing (RNA-seq) has become the most important method in transcriptomics, data analysis methods are proliferating – each accounting for more biological factors. To enable fair and insightful benchmarking of novel methods, we started development of a simulation tool for RNA-seq library construction and sequencing; release is expected in early 2012.



Alignment filtering reduces the false-positive rate of sites inferred to be under positive selection (left) but often leads to a higher error rate in phylogenetic tree inference (right)

Computational and evolutionary genomics

John Marioni

PhD in Applied Mathematics, 2008, University of Cambridge.

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

At EMBL-EBI since September 2010.

The Marioni group develops statistical tools that exploit data generated using next-generation sequencing to understand the evolution and regulation of gene expression.

Previous and current research

Next-generation sequencing has revolutionised genetics, facilitating the generation of high-resolution maps of entire genomes, transcriptomes, and regulatory features such as 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 is motivated by the need to develop methods for modelling heterogeneity in gene expression profiles between populations of cells and also at the single-cell level. In both cases the expression profiles are typically generated using next-generation sequencing. We also focus on finding ways to incorporate a spatial dimension into our models. We collaborate closely with outstanding empirical labs such as those of Detlev Arendt (evolution of Bilateria, page 35) and Marcus Heisler (understanding plant development, page 38) in EMBL Heidelberg, and those of Duncan Odom (modeling gene regulation in mammals) at the University of Cambridge and Simon Tavaré (understanding heterogeneity in expression and genetic variation within Glioblastomas) from Cancer Research UK–Cambridge Research Institute (CRUK-CRI). We work together to frame biological questions of interest, design appropriate studies, and analyse and interpret the data generated.

Selected references

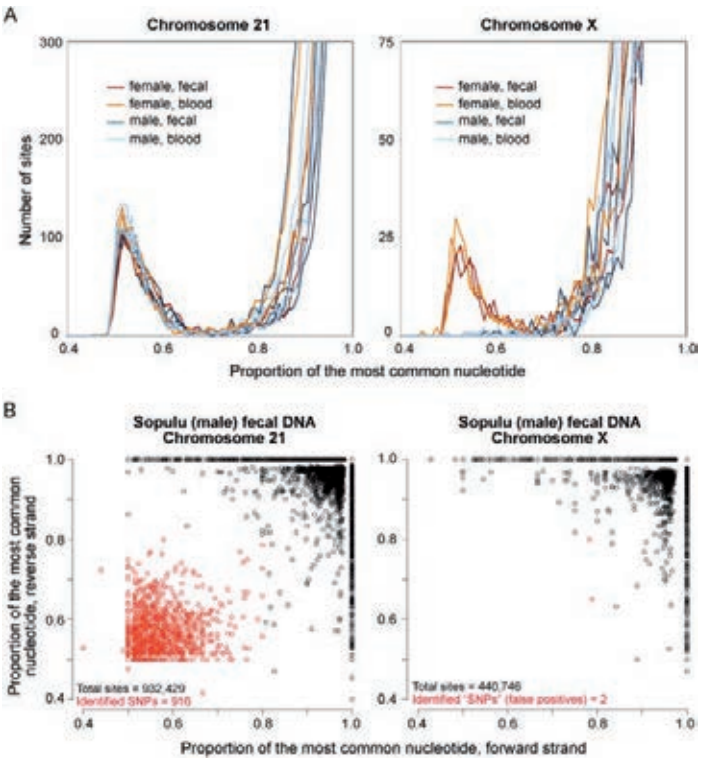
Marioni, J.C., *et al*. (2008). RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.*, 18, 1509-17

Perry, G.H., Melsted, P., Marioni, J.C. (2011). Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome Res.*, (in press); published online 29 December

Perry, G.H., Marioni, J.C, *et al*. (2010). Genomic-scale capture and sequencing of endogenous DNA from faeces. *Mol. Ecol*., 19, 5332-44

Future projects and goals

In 2012 we will continue to develop the methods outlined above, and to collaborate in applying these methods to relevant and important biological questions. From a computational perspective, modelling single-cell transcriptomics data will increase in importance. Methods for storing, visualising, interpreting and analysing the data generated will be critical if we are to exploit these data to the fullest extent. We will also work on methods for analysing conventional next-generation sequencing data, building on work that we have performed previously. These efforts will focus on methods for normalising the data to ensure samples from different individuals are comparable, and on developing statistical approaches for downstream analyses.



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)

ChEMBL: a database of bioactive drug-like small molecules

The Overington group develops and manages ChEMBL. Other projects include analysis of drug attrition, biotherapeutic drug discovery and open-source competitive intelligence.

John Overington

PhD in Crystallography 1991, Birkbeck College, London.

Postdoctoral research, ICRF, 1990-1992. Pfizer 1992-2000. Inpharmatica 2000-2008.

At EMBL-EBI since 2008.

Selected references

Krüger, F.A. and Overington, J.P (2012). Global analysis of small molecule binding to related protein targets. PLoS Comp. Biol. 8, e1002333

Gaulton, A., Bellis, L.J., Bento, A.P., *et al*. (2011). ChEMBL: A large-scale bioactivity database for chemical biology and drug discovery. *Nucl. Acids Res.,* (Database Issue). Published online 23 September

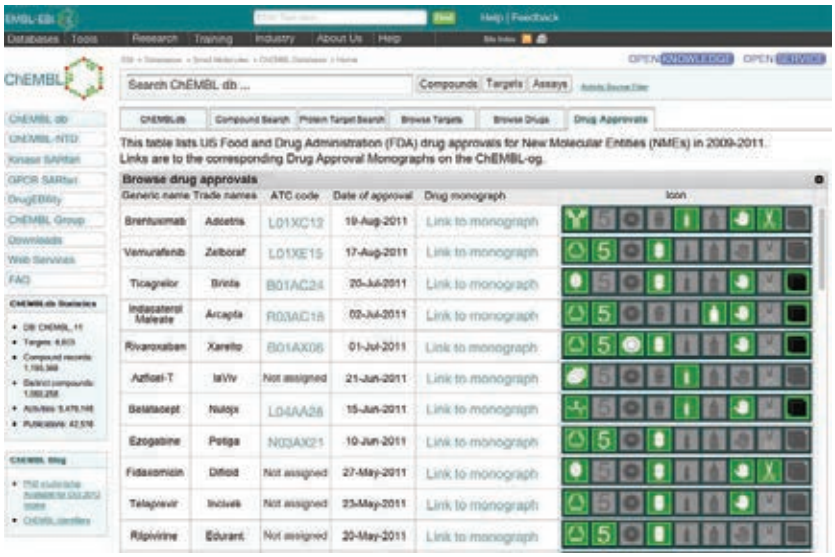
Gleeson, M.P., Hersey, A., Montanari, D. *et al.* (2011). Probing the links between *in vitro* potency, ADMET and physicochemical parameters. *Nature Rev. Drug Disc.*, 10, 197-208

Orchard, S., Al-Lazikani, B., Bryantet, S., *et al*. (2011). Minimum information about a bioactive entity (MIABE). *Nature Rev. Drug Discov.*, 10, 661-669

Previous and current research

The ChEMBL group develops and manages the EBI's database of quantitative small molecule bioactivity data. Synthetic small molecules form the majority of life-saving drugs, and the discovery of safe and efficacious new drugs acting through novel biological mechanisms is a major goal for life science research. The ChEMBL database stores curated two-dimensional chemical structures, abstracted bioactivity data (i.e. binding constants, pharmacology and ADMET data) alongside calculated molecular properties. ChEMBL data are abstracted and curated from the primary scientific literature and cover a significant fraction of the structure–activity relationship data for the discovery of modern drugs.

Our research interests focus on applying the ChEMBL data to drug-discovery challenges. One major area is the analysis of the properties of successful peptide-derived drugs in which physicochemical properties are very different to those of classic Lipinski-like synthetic molecules. Another area has been the exploration of the properties of 'tool compounds', the small-molecule probes of cells and pathways that are established as a core part of chemical biology research. We looked specifically at the quantitative conservation and differences of binding of compounds between established rodent models and human systems. We also explored the development of a novel integrated theoretical-, X-ray-, and NMR-based approach to binding-site characterisation. Other research has included the mapping of temporal expression-level changes of drug-target and metabolic systems from prenatal to geriatric stages for rodents and humans.



The ChEMBL database

Future projects and goals

In 2012 we will complete the rollout of UniChem across the institute as a standard method for cross-resource chemical structure integration, and then add new complementary sources including patent data and commercially available compounds. We will continue to develop and extend approaches to predict the inherent tractability of new target systems for drug discovery, including extension of the data used to cover clinical development stage compounds. We will also develop new informatics approaches to discover and curate key clinical stage molecular and target validation data from the broad literature. Finally, we will explore the integration of ChEMBL data with human variation data for target validation, drug re-use and drug safety purposes.

Systems biomedicine

Julio Saez-Rodriguez

PhD 2007, University of Magdeburg.

Postdoctoral work at Harvard Medical School and MIT.

At EMBL-EBI since 2010.
Joint appointment, Genome Biology Unit (EMBL-HD).

The Saez-Rodriguez group develops computational methods and tools to analyse signalling networks and mathematical models to better understand how signalling is altered in human disease.

Previous and current research

Our group aims to achieve a functional understanding of signalling networks and their deregulation in disease, and to apply this knowledge to novel therapeutics. 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 has practical applications, as alterations in the functioning of these networks underlies the development of diseases such as cancer and diabetes. Considerable effort has been devoted to identifying proteins that can be targeted to reverse this deregulation. However, their benefit is often unexpected: 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 research is hypothesis-driven and tailored towards producing mathematical models that integrate diverse data sources. To this end, we collaborate closely with experimental groups. Our models integrate a range of data (from genomic to biochemical) with various sources of prior knowledge, with an emphasis on providing both predictive power of new experiments and insights into the functioning of the signalling network. We combine statistical methods with models describing the mechanisms of signal transduction either as logical or physico-chemical systems. For this, we develop tools and integrate them with existing resources. We then use these models to better understand how signalling is altered in human disease and predict effective therapeutic targets.

Selected references

Saez-Rodriguez, J.*, Alexopoulos, L.G.*, Stolovitzky, G. (2011). Setting the standards in signal transduction research. *Science Signaling*, 4, pe10

Prill, R.J.,*, Saez-Rodriguez, J.*, Alexopoulos, L.G., Sorger, P.K., Stolovitzky, G. Crowdsourcing network inference: the DREAM predictive signaling network challenge. *Science Signaling*, 4, mr7

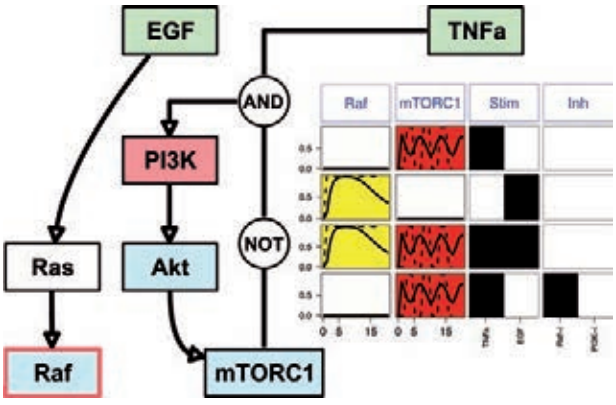
J Saez-Rodriguez*, LG Alexopoulos*, M Zhang, MK Morris, DA Lauffenburger, PK Sorger. (2011). Comparing signaling networks between normal and transformed hepatocytes using discrete logical models. *Cancer Research*, 71, 1-12

Morris, M.K., Saez-Rodriguez, J., Clarke, D., Sorger, P.K., Lauffenburger, D.A. (2011) Training signaling pathway maps to biochemical data with constrained fuzzy logic: quantitative analysis of liver cell responses to inflammatory stimuli. *PLoS Comp. Biol.*, 7, e1001099

Future projects and goals

We will continue to develop methods and tools to understand signal transduction in human cells, as well as their potential to yield insights of medical relevance. Our main focus will be on modelling signalling networks using phospho-proteomics data with our tool CellNOpt, and finding ways to employ different proteomics technologies and sources of information about pathways. We will also continue to develop methods to infer 'drug mode of action' and 'drug repurposing' by integrating genomic and transcriptomic data with drug screenings. Using 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? 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?



An illustration of how we use our logic modelling method CellNOpt to better understand deregulation of signal transduction in disease. Left: simple pathway model; right: experimental data and match between model simulations and data

Cheminformatics and metabolism

The Steinbeck group studies small molecule metabolism in biological systems, including reconstruction of metabolic networks from genomic data, and structure elucidation and identification of metabolites.

Selected references

Alcantara, R., Axelsen, K. B., Morgat, A., *et al*. (2011). Rhea: a manually curated resource of biochemical reactions. *Nucleic Acids Res.*, 40, D754-D760

de Matos, P., Adams, N., Hastings, J., *et al*. (2012). A database for chemical proteomics: ChEBI. *Methods in Molecular Biology*, 803, 273-96

Orchard, S., Al-Lazikani, B., Bryantet, S., *et al*. 2011. Minimum information about a bioactive entity (MIABE). *Nature Rev. Drug Discov*. 10, 661-9

Christoph Steinbeck

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

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.

Previous and current research

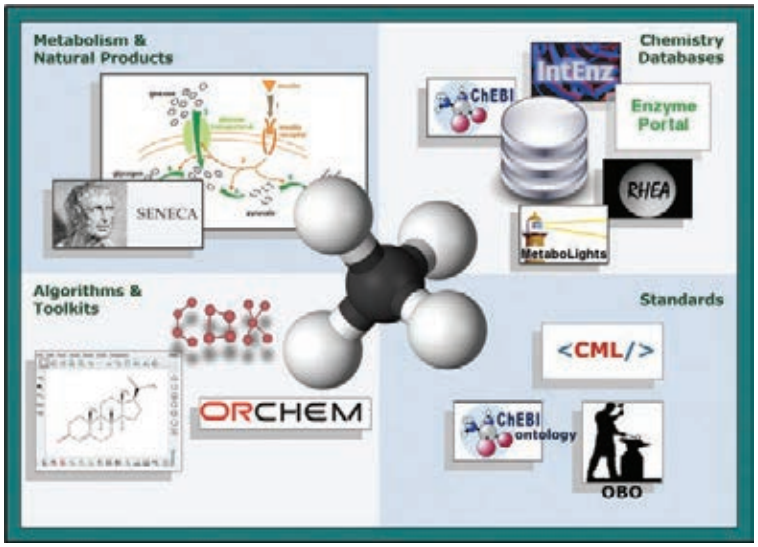
Our team provides the biomedical community with information on small molecules and their interplay with biological systems. Our database portfolio includes ChEBI, EMBL-EBI's database and ontology of chemical entities of biological interest, as well as Rhea and IntEnz, our enzyme-related resources. We develop methods to decipher, organise and publish the small-molecule metabolic content of organisms. We 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 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.

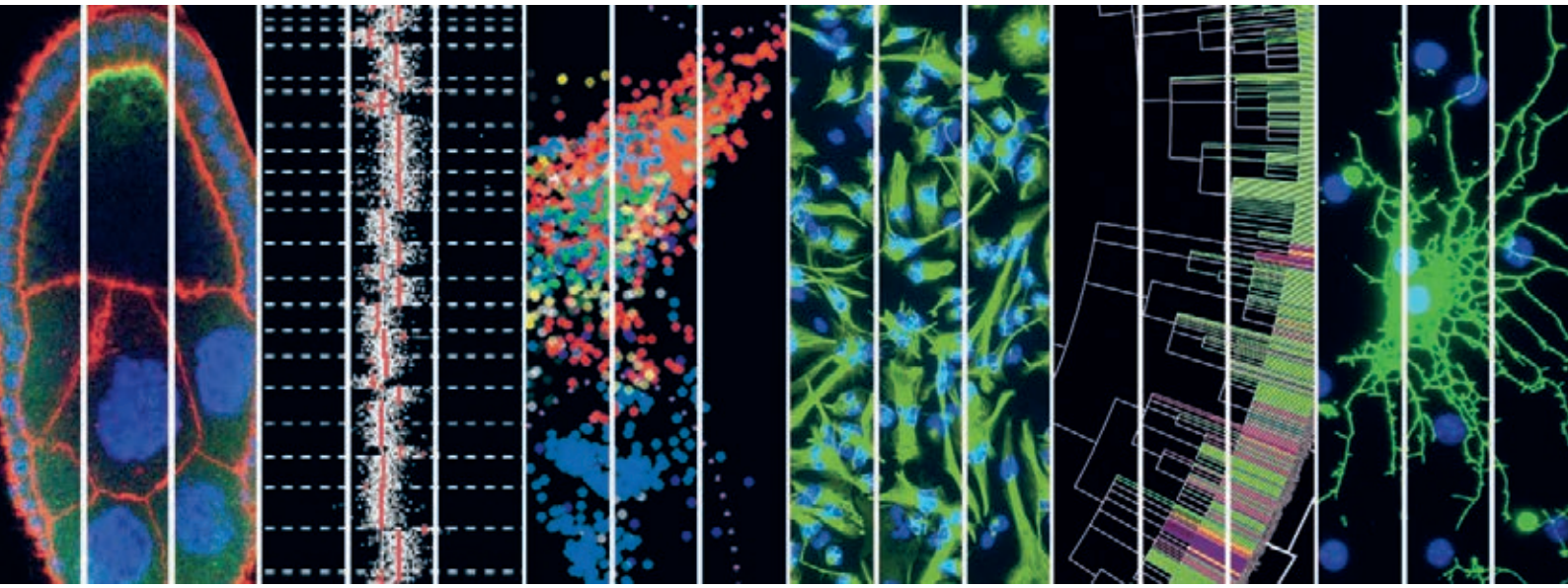
We are also investigating the extraction of chemical knowledge from the printed literature by text- and graph-mining methods. With our collaborators we have developed a number of well established and widely used open-source cheminformatics software packages. For example, the Chemistry Development Kit (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 Lego™-like manner, as well as OrChem, our structure-registration and -searching system for the Oracle™ -database. In collaboration with partners in Uppsala, we initiated Bioclipse, an award-winning rich client for chemo- and bioinformatics.

Future projects and goals

In early 2012 we will release two new major resources: The Enzyme Portal and the MetaboLights database archive. We will gather user feedback and work on stabilising and further developing these resources. A strong focus will be on developing the MetaboLights reference layer, which will contain information on individual metabolites and their chemical, spectroscopic and biological properties. This layer will strongly interface with existing resources at EMBL-EBI including ChEBI, Reactome, UniProtKB, the BioSamples Database, and the Expression Atlas. We are also working with the metabolomics community on data-exchange formats and mechanisms. As part of our efforts developing metabolomics and metabolism resources, the ChEBI team is working on a foundational release of a 'natural-products' collection in ChEBI. This first release will consist of approximately 5000 natural products, their structures, and information about the biological source.

A summary of the activities conducted within the Cheminformatics and Metabolism group which includes natural products and metabolism, chemistry databases, standards and cheminformatics toolkits





Bioinformatics Services

Service teams at EMBL-EBI focus on gathering, adding value to and presenting important collections of biological and chemical information for the benefit of the larger research community. Their work is enhanced by the input of basic researchers, some of whom are embedded in service teams. Master's students and other visiting scientists who conduct their research at EMBL-EBI have the opportunity to work in a unique environment, exploring how we can use an incredibly diverse range of information to understand life on a fundamental level.

Building on more than 20 years' experience in bioinformatics, EMBL-EBI maintains the world's most comprehensive range of molecular databases. We are the European node for globally coordinated efforts to collect and disseminate biological data. Many of our databases are household names to biologists – they include EMBL-Bank (DNA and RNA sequences), Ensembl (genomes), ArrayExpress (microarray-based gene-expression data), UniProt (protein sequences), InterPro (protein families, domains and motifs) and PDBe (macromolecular structures). Others, such as IntAct (protein–protein interactions), Reactome (pathways), ChEBI and ChEMBL (small molecules), help researchers understand not only the molecular parts that go towards constructing an organism, but how these parts combine to create systems. The details of each database vary, but they all uphold the same principles of service provision: accessibility, compatibility, comprehensive datasets, portability, and quality.

European Nucleotide Archive

Guy Cochrane

PhD 1999, University of East Anglia.

At EMBL-EBI since 2002.
Team leader since 2009.

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 including sequence similarity search, text search, browsing, rich integration with data resources beyond ENA, provided both over the web and through an increasingly sophisticated programmatic interface. These services are focused towards users who approach ENA data and services directly, and those who provide secondary services, such as UniProt, Ensembl, Ensembl Genomes and ArrayExpress, that build on ENA content. 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 ongoing focus 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.

Proteomics services

Henning Hermjakob

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

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

At EMBL-EBI since 1997.

Our 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 Organisation (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 RICORDO project, we contribute to the development of an interoperability framework that bridges physiology and molecular biology. As a result of long-term engagement with the proteomics community, journal editors and funding organisations, proteomics data deposition in PSI-compliant data resources such as IntAct and PRIDE is becoming a strongly recommended part of the publishing process. This has resulted in a rapid increase in the data content of our resources. In addition, the Proteomics curation teams ensure consistency and appropriate annotation of all data, whether from direct depositions or literature curation, to provide the community with high-quality reference datasets. We also contribute to the development of data integration technologies using the Distributed Annotation System (DAS) and web services across a range of projects, including EU Apo-Sys, LipidomicNet, SLING, and the NIH UCLA Proteomics Center.

InterPro

Sarah Hunter

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

InterPro is used to classify proteins into families and predict the presence of domains and functionally important sites. The project integrates signatures from eleven major protein signature databases into a single resource. During the integration process, InterPro rationalises instances where more than one protein signature describes the same protein family or 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 (UniParc) proteins using the InterProScan software, and displays the matches to the UniProt KnowledgeBase (UniProtKB) in various formats, including XML files and web-based graphical interfaces. 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.

Metagenomics is the study of the sum genetic material found in an environmental sample or host species, typically using Next Generation Sequencing (NGS) technology. The new metagenomics resource recently established at EBI enables metagenomics researchers to submit sequence data and associated descriptive metadata to the public nucleotide archives. Deposited data is subsequently functionally analysed using an InterPro-based pipeline and the results generated are visualized via a web interface.

Non-vertebrate genomics

Paul Kersey

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

The Ensembl Genomes team provides services based on the genomes of non-vertebrate species. The falling costs of DNA sequencing have led to an explosion of reference genome sequences and genome-wide measurements and interpretation. Ensembl Genomes (Kersey et al., 2010) provides portals for bacteria, protists, fungi, plants and invertebrate metazoa, offering access to these data through a set of programmatic and interactive interfaces, exploiting developments originating in 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 than if they develop their own tools. Our leading collaborators include VectorBase (Megy *et al.*, 2011), a resource focused on the annotation of invertebrate vectors; WormBase (Yook *et al.*, 2011), a resource for nematode biology; and PomBase (Wood *et al.*, 2011), focused on the fission yeast *Schizosaccharomyces pombe*. In the plant domain, we collaborate closely with Gramene in the US and with a range of European groups in the transPLANT project. 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. We have developed a new portal for plant pathogen data, PhytoPath (launched in early 2012), and are involved in the development of Microme, a new resource for bacterial metabolic pathways.

The Protein Data Bank in Europe

Gerard Kleywegt

PhD 1991, University of Utrecht.

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) is 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 Protein Data Bank (PDB) and the Electron Microscopy Data Bank (EMDB).

The major goal of PDBe is to provide integrated structural data resources that evolve with the needs of biologists. To that end, our team endeavours to: handle deposition and annotation of structural data expertly; provide an integrated resource of high-quality macromolecular structures and related data; and maintain in-house expertise in all the major structure-determination techniques (e.g. X-ray crystallography, Nuclear Magnetic Resonance spectroscopy and 3D electron microscopy). Our specific focus areas are: advanced services, ligands, integration, validation and experimental data.

UniProt development

Maria J. Martin

BSc in Veterinary Medicine, Universidad Complutense, Madrid, 1990.

PhD in Molecular Biology (Bioinformatics) 2003, Universidad Autonoma, Madrid.

At EMBL-EBI since 1996.
Technical team leader since 2009.

Our team provides the bioinformatics infrastructure for the databases and services of the Universal Protein Resource (UniProt). The team consists of software engineers and bioinformaticians responsible for the UniProt software and database development, and the study of novel automatic methods for protein annotation and representation. A user-experience analyst in the team coordinates the user-request gathering process for the design and development of the website. The team is also responsible for the maintenance and development of tools for UniProt curation. We work in a fully complementary fashion with Claire O'Donovan's UniProt Content group (see page 95) to provide essential resources to the biological community so that our databases can remain an integral part of the tools researchers use regularly for their work.

Literature services

Johanna McEntyre

PhD in Plant Biotechnology 1990, Manchester Metropolitan University.

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.

The scientific literature often represents both the start and end point of a scientific project. Placing the literature within the context of related public data resources will better equip researchers 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 – will become a necessity if we are to leverage the investment in scientific research to greater potential. Text-mining represents a high-throughput approach to the identification of biological concepts in articles, which can then form the basis for the development of new applications and stimulate precise, deep linking to related data resources in the future. Our team's goal is to build text-based resources for the life sciences that are integrated with other public-domain data resources hosted at EMBL-EBI. To this end, we run two literature databases: CiteXplore and UK PubMed Central. CiteXplore contains over 26 million abstracts and includes PubMed and data from Agricola and the European Patent Office. UK PubMed Central comprises over 2 million full text articles, of which about 400 000 are open access. The databases are linked to a number of EMBL-EBI data resources using the references appended to database records by curators and submitters as well as through text mining to identify terms of interest and linking with appropriate databases. We also calculate citation-network information for the records we hold: over 10 million articles have been cited at least once, representing the largest public-domain citation network in the world. We plan to use this infrastructure to develop novel and useful search and browse features for publications mapped to data, and to share the article content and annotation as widely as possible, both programmatically and for individual users.

UniProt content

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.

The UniProt Content Team comprises biocurators and bioinformaticians working on the Universal Protein Resource (UniProt). Biocuration involves the translation and integration of information relevant to biology into a database or resource that enables integration of the scientific literature as well as large data sets. Accurate and comprehensive representation of biological knowledge, as well as easy access to this data for working scientists and a basis for computational analysis, are the primary goals of biocuration. The goals of biocuration are achieved thanks to the convergent endeavors of biocurators and software developers and our team works in a fully complementary fashion with Maria Jesus Martin's UniProt Development group (see page 94) to provide essential resources to the biological community such that databases have become an integral part of the tools researchers use on a daily basis for their work.

Protein Data Bank in Europe: Databases and services

Tom Oldfield

DPhil 1990, University of York.

Postdoctoral research at GlaxoSmithKline, 1990-1993.

Principal Scientist at Accelrys Inc., 1993-2002.

At EMBL-EBI since 2002.
Team Leader since 2010.

The Protein Data Bank in Europe (PDBe) is one of six core databases located at EMBL-EBI and is also a partner in the Worldwide Protein Data Bank organisation (wwPDB) along with the RCSB and BMRB in the United States and PDBj in Japan. The PDBe team manages two production systems: the weekly update of deposited data, and the weekly increment of new released data. These production data systems are managed within multiple Oracle™ databases and support a large number of integrated web resources to collect data and disseminate information to the wider community. With the future provision of new services based on structure validation data there will be an emphasis in 2012 to extend the core databases and infrastructure. This will require an optimisation of the loading tools to manage both the increased amount of depositions expected and also the breadth of data required to support new services.

Functional genomics production

Helen Parkinson

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

At EMBL since 2000.

We manage data content and user interaction for three EBI databases: the ArrayExpress Archive, the Gene Expression Atlas and the Biosamples Database. All three have complex meta-data representing experimental types, variables and sample attributes for which we require semantic mark-up in the form of ontologies. We develop both ontologies and software for the annotation of complex biological data including the Experimental Factor Ontology (EFO) for functional genomics annotation, the Software Ontology (SWO), Ontology for Biomedical Investigation (OBI), the Coriell Cell Line Ontology and the Vertebrate Anatomy Ontology (VBO). ArrayExpress is a driving biological project for the National Center for BioOntology and we are developing ontology manipulation tools and tools for the semantic web to support functional genomics data integration. We have developed new formats to represent BioSamples that are used to solicit and load submissions to the BioSamples database. In the context of the the KOMP2 project, we manage, analyse, and distribute complex phenotypic data from 20 000 knockout mouse lines that will be generated over the course of five years. We also collaborate to develop tools for the annotation and production of the NHGRI's genome-wide association study catalogue.

In addition to microRNA data analysis, our research activities have focused on analyses (in R) of RNA-seq data in collaboration with the Geuvadis consortium and genotype imputation for individuals from the 1000 Genomes Project. Semantic web projects include development of an RDF representation of Gene Expression data from the Gene Expression Atlas and ArrayExpress.

Functional genomics software development

Ugis Sarkans

PhD in Computer Science 1998, University of Latvia.

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

At EMBL-EBI since 2000.

Our team has been developing software for ArrayExpress since 2001. As of January 2012, ArrayExpress holds data from more than 770 000 microarray hybridisations and is 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; and MIAMExpress (a data annotation and submission system). We participate in building the BioSamples database, a new EBI resource, as well as in a number of collaborative multi-omics projects in a data-management capacity.

Our team participates in several pilot projects in a data management role. We believe that being close to large consortia that generate different types of high-throughput data places us in a better position to fulfil our main objective of developing ArrayExpress and BioSamples infrastructures. Our participation in diXa, a toxicogenomics data management project, will enable us to build better links between BioSamples, ArrayExpress and other EMBL-EBI assay data resources. We are also beginning to work with a project devoted to autism research, which will contribute to a deeper understanding of ways to manage complex endophenotype data, such as imaging data.

Protein Data Bank in Europe: Content and integration

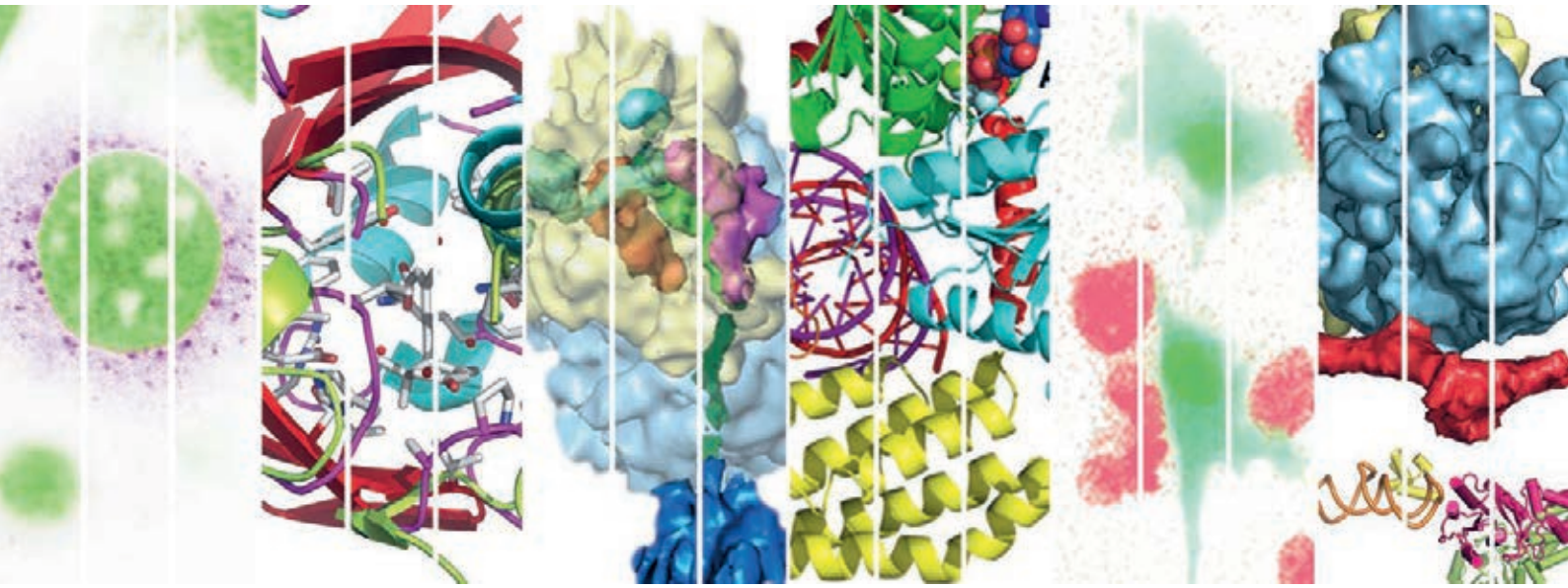
Sameer Valenkar

PhD 1997, Indian Institute of Science. Postdoctoral researcher, Oxford University, UK, 1997-2000.

At EMBL-EBI since 2000.

The Protein Data Bank in Europe (PDBe) manages the worldwide biomacromolecular structure archive, the Protein Data Bank (PDB) and is a founding member of the Worldwide Protein Data Bank (wwPDB). We accept and annotate worldwide depositions of biomacromolecular structures determined using X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy, 3D Electron Microscopy (EM) and other structure determination methods. PDBe is also a founding member of EMDataBank, which manages the deposition and annotation of electron microscopy data in EMDB.

Our goal is to ensure that PDBe truly serves the needs of the biomedical community. As part of that effort, we are constantly improving the web interface for existing tools and services and designing new tools to make structural data available to all. In the context of the SIFTS project, we integrate structural data with other biological data in the interests of facilitating discovery. These integrated data form the basis for many query interfaces that allow macromolecular structure data to be presented in its biological context. Our specific focus areas are: data integrity, data quality, integration and data dissemination to the non-expert biomedical community.



Structural Biology

EMBL Grenoble is a key player in the Partnership for Structural Biology (PSB) which integrates the activities of several Grenoble institutes involved with structural biology and thus is able to provide a uniquely comprehensive range of platforms for both inhouse research and external users.

A cornerstone of the PSB (established with the Institut Laue-Langevin, the French national Institut de Biologie Structurale and the European Synchrotron Radiation Facility (ESRF)) is the close interaction of EMBL Grenoble with ESRF, which involves collaboration on building and operating beamlines for macromolecular crystallography and small-angle scattering, developing the associated automated instrumentation and techniques, and providing expert help to external visitors. The highly automated ESRF crystallography beamlines are equipped with EMBL-designed high-precision micro-diffractometers 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 identified 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. All these platforms are now available to external users under the EU funded P-CUBE and BioStruct-X projects.

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 holo-translocon with 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 and chromatin-modification complexes is continuing with groups working on TFIID, p300, 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. In addition there are two projects related to the pathway of activation of interferon in response to viral infection, one on the viral RNA pattern recognition receptor RIG-I and the other on the downstream kinase TBK1. The CNRS–Grenoble University–EMBL Unit of Virus Host Cell Interactions (UVHCI) is situated in a building next to the outstation.

Scientists at EMBL Grenoble 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 top-end Polara electron microscope with cryo-tomography capability.

Stephen Cusack
Head of EMBL Grenoble

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

Stephen Cusack

PhD 1976, Imperial College, London, UK.

Postdoctoral work at EMBL Grenoble.
Staff scientist at EMBL Grenoble 1980-1989.
Group Leader, Senior Scientist and Head of Outstation since 1989.

Director of EMBL-Grenoble University-CNRS Unit for Virus Host Cell Interactions since 2007.

The Cusack group uses X-ray crystallography to study the structural biology of protein-RNA complexes involved in RNA metabolism, translation, virus replication and epigenetics.

Previous and current research

Since our goal is to relate protein-RNA complex structures to biological function we are more and more engaged in functional studies, including in live cells. Current major focuses are on negative strand RNA virus polymerases, innate immune system receptors for viral RNA and the X chromosome dosage compensation complex.

We determined the structure of cap-bound human nuclear cap-binding complex (CBC), a 90KDa heterodimeric protein and have studied several other proteins involved in cap-dependent processes, such as the vaccinia virus capping enzyme and PHAX. Once in the cytoplasm, mRNAs are subject to a quality control check to detect premature stop-codons (nonsense mediated decay), which crucially depends (in all eukaryotic organisms studied) on the three conserved Upf proteins (Upf1, Upf2 and Upf3). In mammals, it is linked to splicing. We obtained first structural information on binary complexes of these proteins whose ternary complex formation triggers decay.

In recent years we have focussed on leucyl-tRNA synthetase which 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 structurally characterised the large conformational changes required to switch from the aminoacylation to the editing configurations. We have also collaborated in the elucidation of the mechanism of action of a new boron containing anti-fungal compound that targets the editing site of leucyl-tRNA synthetase and have now exploited this, using structure-based approaches, to design new anti-bacterial compounds that are active against multi-drug resistant strains, including tuberculosis.

Future projects and goals

Ongoing projects related to RNA metabolism include continued studies on PHAX and ARS2, both of which bind CBC and are linked to the metabolism of small RNAs. A major focus is on structure determination of the influenza virus RNA-dependent RNA 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. These results give some insight into the polymerase mutations required for an avian virus to adapt to be able infect humans and also permit structure-based antiviral drug design. We have co-founded a spin-off company to pursue this. This work is now being extended to the polymerases of other segmented RNA viruses such as bunyaviruses. With the Ellenberg (page 20) and Briggs (page 58) groups we are engaged in confocal and cross-correlation fluorescence studies as well as correlative EM microscopy of the assembly and trafficking of the influenza polymerase and RNPs in living, infected cells. We also work on the RIG-I like helicases, which are cytoplasmic pattern recognition receptors of the innate immune system which signal interferon production upon detection and binding of viral RNA. We recently published in *Cell* the first structure-based mechanism of activation of RIG-I and are continuing work on MDA5 and LGP2, members of the same family. Finally we have new projects concerning ncRNAs: the structure and mechanism of the X-chromosome dosage compensation complex, which in *Drosophila* contains an essential non-coding RNA and secondly, in collaboration with the Pillai group (page 107), structural studies of proteins involved in the piRNA pathway.

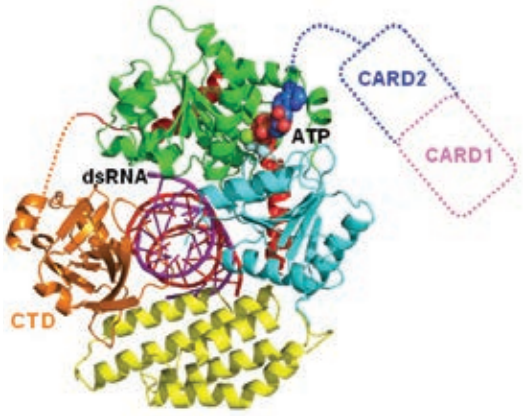
Selected references

Kowalinski, E., Lunardi, T., McCarthy, A.A., Louber, J., Brunel, J., Grigorov, B., Gerlier, D. & Cusack, S. (2011). Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell*, 147, 423-35

Kadlec, J., Hallacli, E., Lipp, M., Holz, H., Sanchez-Weatherby, J., Cusack, S. & Akhtar, A. (2011). Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1. *Nat. Struct. Mol. Biol.*, 18, 142-9

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

Rock, F.L., Mao, W., Yaremchuk, A., Tukalo, M., Crepin, T., Zhou, H., Zhang, Y.K., Hernandez, V., Akama, T., Baker, S.J., Plattner, J.J., Shapiro, L., Martinis, S.A., Benkovic, S.J., Cusack, S. & Alley, M.R. (2007). An antifungal agent inhibits an aminoacyl-tRNA synthetase by trapping tRNA in the editing site. *Science*, 316, 1759-61



Model of the activated state of RIG-I with bound dsRNA (centre) and ATP (top-right). The helicase domains (green and cyan), the insertion domain (yellow) and the C-terminal domain (gold) all contribute to RNA binding, which displaces the CARD domains thus allowing downstream signalling and interferon expression

Structural complexomics of eukaryotic gene expression

The Berger group studies eukaryotic multiprotein assemblies in transcription regulation, develops technologies to produce them recombinantly and subjects them to high-resolution structural and functional analyses.

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 Grenoble since 2007.

Previous and current research

An intense focus of current biological research efforts is the elucidation of protein interaction networks (interactome). Many large multiprotein complexes are discovered. This poses considerable challenges for molecular level studies, in particular for eukaryotic multiprotein complexes with intracellular quantities refractory to large-scale extraction from source.

Our major methodologies include molecular biology and X-ray crystallography. However, we readily apply techniques from other fields that are required for our research, both by ourselves and through collaborative efforts. A prerequisite for successful structural study of many complexes, both by electron microscopy and X-ray crystallography, is production of homogeneous, stable specimens. Present recombinant expression methods often require considerable investment in both labour and materials before multiprotein expression, and after expression and biochemical analysis do not provide flexibility for expressing an altered multiprotein complex. To meet these demands, we have developed our MultiBac system, a modular, baculovirus based technology specifically designed for eukaryotic multiprotein expression. MultiBac is now being used successfully in many laboratories worldwide, not only by structural biologists but also for applications as diverse as the development of vaccines and gene therapy vectors. In our lab, we have recently harnessed homologous and site-specific recombination methods in tandem for all steps involved in multigene assembly, and we have successfully implemented all steps involved in a robotics setup by developing ACEMBL, a proprietary automated platform for multigene recombineering on our TECAN EvoII workstation. By using our technology, we produced numerous large multiprotein assemblies in sufficient quantity and quality for structural studies, including large multicomponent membrane protein complexes and a 1MDa core assembly of human TFIID general transcription factor. Also, we have successfully expanded our multiprotein expression strategies to prokaryotic and mammalian hosts.

Future projects and goals

At EMBL Grenoble, we continue to advance our expression technologies to entirely automate and standardise the process of production for eukaryotic gene regulatory multiprotein complexes including the entire human TFIID holocomplex, its various isoforms and other components of the preinitiation complex. In collaboration with the Schaffitzel Team (page 108) and the Schultz Group (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.

Using state-of-the-art mass spectrometric methods from systems biology, we are developing MultiTRAQ, a new technology 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). Another recent project line in our lab exploits synthetic biology techniques for genome engineering, with the aim of creating disruptive platforms for recombinant protein production, for both academic and industrial applications.

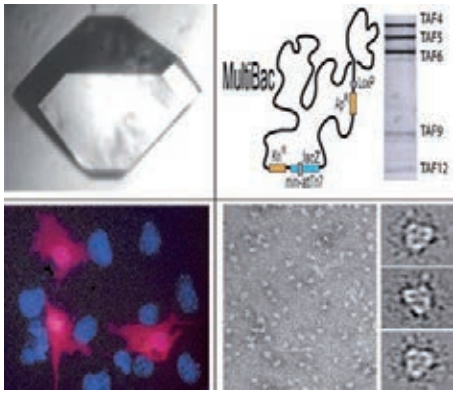
Selected references

Bieniossek, C., Imasaki, T., Takagi, Y. & Berger, I. (2012). MultiBac: expanding the research toolbox for multiprotein complexes. *Trends Biochem. Sci.*, 37, 49–57

Imasaki, T., Calero, G., Cai, G., Tsai, K.L., Yamada, K., Cardelli, F., Erdjument-Bromage, H., Tempst, P., Berger, I., Kornberg, G.L., Asturias, F.J., Kornberg, R.D. & Takagi, Y. (2011). Architecture of the Mediator head module. *Nature*, 475, 240-3

Kriz, A., Schmid, K., Baumgartner, N., Ziegler, U., Berger, I., Ballmer-Hofer, K. & Berger, P. (2010). A plasmid-based multigene expression system for mammalian cells. *Nat. Commun.*, 1, 120

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



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

Diffraction instrumentation team

Florent Cipriani

BSc 1974, Physics, University of Grenoble.

Senior engineer in nuclear and medical industries.

At EMBL Grenoble since 1991.
Team leader since 2003.

The Cipriani team develops instruments and methods for X-ray scattering experiments and works with industry to make them available to scientists worldwide.

Previous and current research

Our team develops instruments and methods for X-ray scattering experiments in collaboration with the EMBL Synchrotron Crystallography Team and the ESRF Structural Biology group, as well as contributing to the development of the EMBL@PETRA-III beamlines in Hamburg. An important aspect of our work is to transfer our technologies to industry to make these instruments available to the scientific community worldwide.

In Grenoble, the ESRF's MX beamlines still rely on our MD2/MD2M diffractometers, SC3 sample changers and C3D crystal centering software to process several hundred crystals per day. Continuing the work carried out over the last decade on micro-diffractometers, a collaboration is under way with EMBL Hamburg to build a vertical kappa diffractometer able to process routinely micron-sized crystals. The prototype installed at EMBL@PETRA-III on the MX2 micro-focus beamline exhibits sub-micron stability at sample position. The quality of data obtained from shutterless 4D data collection scans over thin needles has demonstrated the exceptional dynamic performance of the system. This prototype is the base of a new diffractometer generation (MD3) that should include all the features of the MD2 micro-diffractometer.

Concluding a collaboration started in 2008, we have equipped the ESRF BM29 and the EMBL@PETRA-III BioSAXS beamlines with fast, fully integrated sample environments. Users can now automatically process several hundred samples stored in SBS micro-plates within a minute. To characterise difficult proteins and complexes before any degradation occurs, a project has started (part-supported by the EU program BioStruct-X) to couple the Sample Exposure Unit of the system with an online HPLC device.

Our current work mainly focuses on the development of a crystal harvesting system based on CrystalDirect (CD), a concept conceived jointly with the Márquez team (page 104). Crystals grown on a thin film, in specific vapour diffusion crystallisation plates are recovered by photo ablation (see figure). The CD plates can be set up by standard robotics used at HTX facilities. Designed for automated harvesting they are also ideal for X-ray in-plate crystal screening. A first batch of CD plates has been produced and a basic harvester built to develop the harvesting and post harvesting processes. Motivated by the increasing flow of samples at MX beamlines and by the emergence of automated harvesters, studies have started to develop a compact and precise sample holder for frozen crystals. Lead by the EMBL Grenoble, this 'NewPin' collaborative project is part-supported by BioStruct-X.

Future projects and goals

In the context of the ESRF MASSIF upgrade program, we plan to equip the EMBL/ESRF BM14 beamline with industrial robotics to test the 'NewPin' sample-holder standard and to evaluate in situ CD plate crystal screening. Our long-term ambition is to bridge, via CrystalDirect, the Grenoble HTX lab and BM14 in a pilot automated MX facility that can provide full remote service – from crystallisation to data collection.

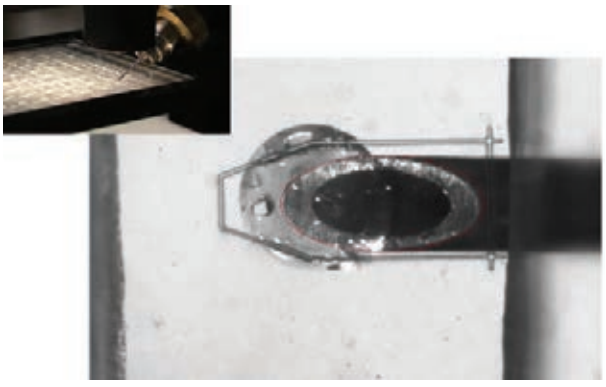
Selected references

Russi, S., Juers, D.H., Sanchez-Weatherby, J., *et al.* (2011). Inducing phase changes in crystals of macromolecules: status and perspectives for controlled crystal dehydration. *J. Struct. Biol.*, 175, 236-43

Pernot, P., Theveneau, P., Giraud, T. *et al*. (2010). New beamline dedicated to solution scattering from biological macromolecules at the ESRF. *Journal of Physics*, Conference Series 247

Cipriani, F., Felisaz, F., Launer, L., *et al.* (2006). Automation of sample mounting for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, 62, 1251-59

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, 1765-70



Harvesting crystals grown on a thin film by photo ablation

High-throughput protein technologies

The Hart team develops new high-throughput molecular biology methods and uses them to study enzymes of biological and medical interest.

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.

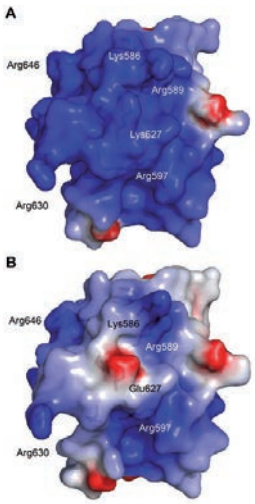
Selected references

An, Y., Yumerefendi, H., Mas, P.J., Chesneau, A. & Hart, D.J. (2011). ORF-selector ESPRIT: A second generation library screen for soluble protein expression employing precise open reading frame selection. *J. Struct. Biol.*, 175, 189-97

Boivin, S. & Hart, D.J. (2011). Interaction of the influenza A virus polymerase PB2 C-terminal region with importin {alpha} isoforms provides insights into host adaptation and polymerase assembly. *J. Biol. Chem.*, 286, 10439-48

An, Y., Meresse, P., Mas, P.J. & Hart, D.J. (2011). CoESPRIT: a library-based construct screening method for identification and expression of soluble protein complexes. *PLoS One*, 6, e16261

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, 28411-7



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

Previous and current research

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. Consequently, we are able to study certain biological questions with advantages over classical approaches.

Influenza RNA polymerase: There is global 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: 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 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.

Intrinsically unstructured proteins and their interactions: A large proportion of the proteome possesses little or no structure. These regions may serve as simple linkers; sometimes however they become structured upon binding partners (proteins, nucleic acids, small molecules). In collaboration with the Blackledge group (IBS, Grenoble), we have recently begun studying several such systems involved in viral replication.

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 determined, disruption of such protein-protein interactions suggests a new route towards specific HDAC inhibition.

Future projects and goals

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 to take advantage of recent 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.

High-throughput crystallisation laboratory

José A. Márquez

PhD 1997, University of Valencia.

Postdoctoral research at EMBL.

Staff scientist at EMBL Grenoble since 2003.
Team leader since 2005.

The Márquez team develops low volume, high-throughput, techniques to optimise protein crystallisation and uses them to study the structure of sensing and signalling molecules.

Previous and current research

The HTX lab is one of the major facilities for high-throughput nanovolume crystallisation screening in Europe, open to scientists working in European academic institutions through the EC-funded Protein Production Platform project P-CUBE. It is also one of the major resources of the Partnership for Structural Biology in Grenoble. Since opening, the HTX lab has offered services to hundreds of scientists, performing several million experiments. The lab is also involved in the development of data management resources, as well as new crystallisation techniques.

Towards the integration of automated crystallisation and data collection: While crystallisation and X-ray data collection is often highly automated, the process of mounting crystals into cryogenic X-ray data collection supports remains difficult and time-consuming. In collaboration with the Cipriani team (page 102), we have developed a new approach (known as Crystal Direct) designed to enable full automation of crystal harvesting process, based on a redesigned vapour diffusion crystallisation plate. Advantages include: direct compatability with X-ray data collection; recovery using laser-induced photo ablation; absence of mechanical stress during crystal mounting; and the possibility to recover micro-crystals. A prototype has been built and is currently being tested.

Crystallisation Information Management System (CRIMS): CRIMS tracks experiments and makes results available to users via the web in real-time. We developed the system to tackle the challenge of capturing the enormous amounts of information generated by modern experiments. CRIMS software is currently in use at six other laboratories in Europe, three of them at synchrotron sites.

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. We have shown how receptor dimerisation modulates ligand binding affinity leading to differential sensitivities towards the hormone (Dupeux *et al.*, 2011). The work provides a novel framework for understanding the ABA signalling pathway and activation of the stress response in plants, and illustrates how receptor oligomerisation can modulate ligand binding affinity by influencing the thermodynamics of the overall activation reaction.

Future projects and goals

Aside from the continuous development of our crystallisation plates and crystal harvesting system, we will focus on the implementation of automated cryo-protection protocols as well as in the connection of CRIMS with synchrotron data management systems, with the aim to fully integrate crystallisation and data collection.

We continue to focus on the structural study of signalling systems. 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. Such molecules should be easier to synthesise and more stable than ABA itself and could be used to improve crop tolerance to drought and other types of environmental stress.

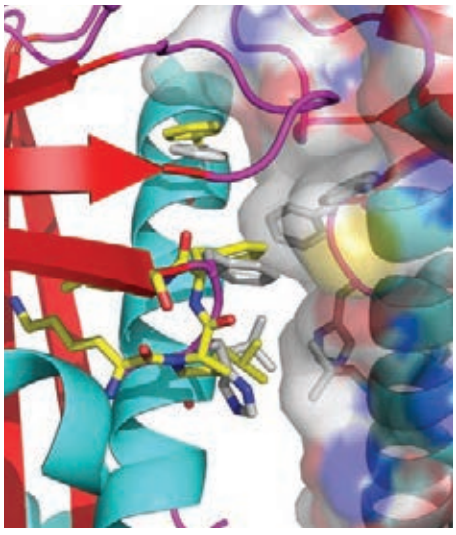
Selected references

Santiago, J., Dupeux, F., Betz, K., Antoni, R., Gonzalez-Guzman, M., Rodriguez, L., Márquez, J.A., Rodriguez, P.L. (2012). Structural insights into PYR/PYL/RCAR ABA receptors and PP2Cs. *Plant Sci.*, 182, 3-11

Dupeux, F., Röwer, M., Seroul, G., Blot, D., Márquez, J.A. (2011). A thermal stability assay can help to estimate the crystallization likelihood of biological samples. *Acta Crystallogr.*, D67, 915-9

Dupeux, F., Santiago, J., Betz, K., *et al.* (2011). A thermodynamic switch modulates abscisic acid receptor sensitivity. *EMBO J.*, 30, 4171-84

Dupeux F, Antoni R, Betz K., *et al*. (2011). Modulation of abscisic acid signaling *in vivo* by an engineered receptor-insensitive protein phosphatase type 2C allele. *Plant Physiol*., 156, 106-16



Detail of the PYR1 Abscisic acid receptor dimerisation interface

Synchrotron crystallography team

The McCarthy team works on the design, construction and operation of MX and BioSAXS beamlines and studies proteins involved in neuron development.

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 since 2007.

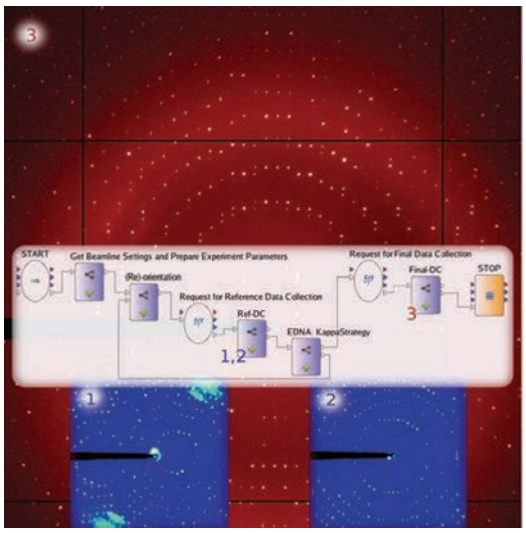
Selected references

Kowalinski, E., Lunardi, T., McCarthy, A.A., Louber, J., Brunel, J., Grigorov, B., Gerlier, D. & Cusack, S. (2011). Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. Cell, 147, 423-35

Brockhauser, S., White, K.I., McCarthy, A.A. & Ravelli, R.B. (2011). Translation calibration of inverse-kappa goniometers in macromolecular crystallography. Acta Crystallogr A, 67, 219-28

Gabadinho, J., Beteva, A., Guijarro, M., *et al.* (2010). MxCuBE: a synchrotron beamline control environment customized for macromolecular crystallography experiments. J Synchrotron Radiat, 17, 700-7

Pernot, P., Theveneau, P., Giraud, T., *et al.* (2010). New beamline dedicated to solution scattering from biological macromolecules at the ESRF. *Journal of Physics*, Conference Series 247



An example of a kappa reorientation workflow with diffraction images captured at different orientations. 1 is the initial crystal orientation, 2 in an aligned orientation and 3 is the first image of the final dataset

Previous and current research

The team works in close collaboration with the Structural Biology Group of the European Synchrotron Radiation Facility (ESRF). We are currently responsible for two macromolecular beamlines, ID14-4 and ID23-2, as well as the biological X-ray scattering (BioSAXS) beamline at BM29. The team also manages the operation of BM14, which is run as a partnership with the ESRF and Indian government. The structural biology beamlines at the ESRF continue to perform, resulting in the deposition of more than 1,005 structures in the PDB last year, the first such milestone for the macromolecular crystallography (MX) beamlines at the ESRF. Another team highlight was the successful upgrade of the BM14 X-ray optics, culminating in a four-fold increase in flux coupled with a similar magnitude contraction in beamsize. We also work in close collaboration with the Cipriani Team (page 102) to develop hardware, software and novel methodologies for sample handling and data collection possibilities. A recent example is the use of workflows (see figure) in the design and implementation of complex experiments in MX.

The team also studies proteins involved in neuronal development. We are particularly interested in the Slit-Robo signalling complex that is essential for the normal development of the central nervous system. This signalling system has also been implicated in heart morphogenesis, angiogenesis and tumour metastasis. Meanwhile the BM14 group is actively involved in structural studies of Toxoplasma gondii epigenetic machinery in collaboration with Mohamed-Ali Hakimi (Grenoble Medical University).

Future projects and goals

This year will be exciting with the completion of the first phase of 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 decade. This year there will see the commissioning and operation of a new BioSAXS beamline on BM29 as well as the first commissioning stages of the new MASSIF suite of MX-beamlines on ID30A. On BM14 we will modify the sample environment to enable the *in situ* screening of crystals in plates. Our team will continue to develop automated data screening, collection and analysis schemes, phasing methods using UV induced damage, as well as a new diffraction image viewer software developed as part of BioStruct-X. We will continue to develop a highly automated BioSAXS beamline in collaboration with the ESRF, the Cipriani team, and EMBL Hamburg, which includes the integration of an online HPLC system. We hope that all our combined efforts will push the boundaries of structural biology to better understand the molecular basis of more complex biological systems.

In the laboratory we will continue our research on the Slit-Robo signalling complex by trying to decipher how exactly Slit activates Robo on the cell surface. The BM14 group will over-express several Toxoplasma proteins in *E. Coli* to be structurally and biochemically characterised using the PSB facilities. Lastly, the initial work on a novel dequibiquitinase and the ESPRIT optimisation of a human kinase in collaboration with the Hart group (page 103) will be continued.

Integrating signals through complex assembly

Daniel Panne

PhD 1999, University of Basel.

Postdoctoral research at Harvard University, Boston.

Group leader at EMBL Grenoble since 2007.

The Panne group looks to understand important signalling processing pathways in the cell, which could help in the discovery of anti-viral drugs.

Previous and current research

Most cellular processes depend on the action of large multi-subunit complexes, many of which are assembled transiently and change 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. Characterisation and visualisation of such cellular structures is one of the most important challenges in molecular biology. Characterisation requires expertise in a number of techniques including molecular biology, biochemistry, biophysics, structural biology and bioinformatics. Visualisation uses low-resolution imaging techniques such as electron microscopy and small angle X-ray scattering , 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. This 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. For the first time, we have detailed insights into the structure of an enhanceosome, and 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. 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 in understanding 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 expand our multiprotein crystallisation strategies to complexes involved in modification of chromatin structure.

Selected references

Reynoird, N., Schwartz, B.E., Delvecchio, M., et al. (2010). Oncogenesis by sequestration of CBP/p300 in transcriptionally inactive hyperacetylated chromatin domains. *EMBO J.*, 29, 2943-52

Shen A., Higgins D.E., and Panne D. (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, 236-42

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

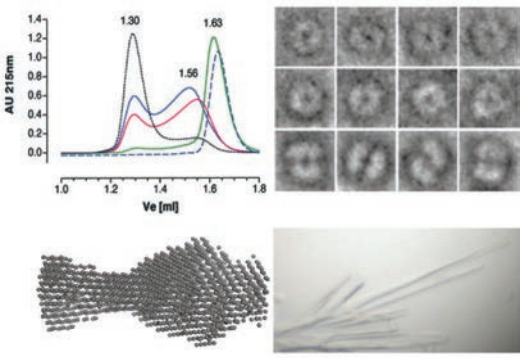


Figure 1: We employ a number of different resolution techniques to visualise cellular structures

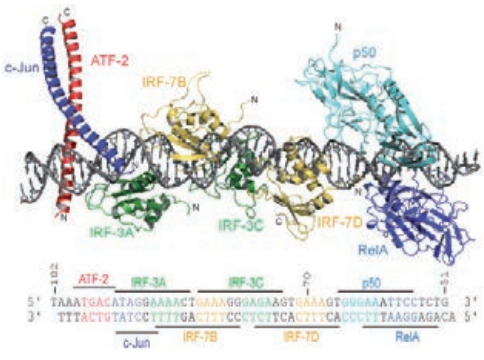


Figure 2: Atomic model of the INF-β enhanceosome

Regulation of gene expression by non-coding RNAs

The Pillai group seeks to understand molecular mechanisms involved in piRNA biogenesis and its function in protecting the genome from instability.

Rameshi Pillai

PhD 2002, University of Bern.

Postdoctoral research at the Friedrich Miescher Institute, Basel.

Group leader at EMBL Grenoble since 2006.

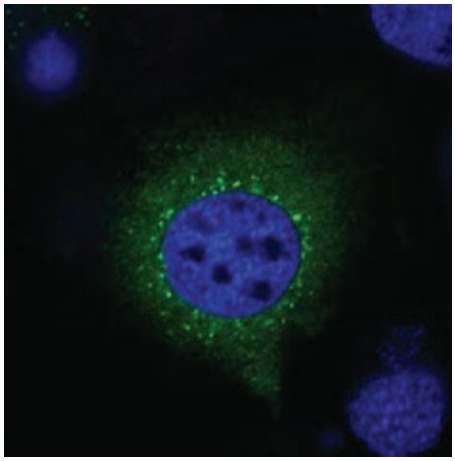
Selected references

Reuter, M., Berninger, P., Chuma, S., Shah, H., Hosokawa, M., Funaya, C., Antony, C., Sachidanandam, R. & Pillai, R.S. (2011). Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature*, 480, 264-67

Simon, B., Kirkpatrick, J.P., Eckhardt, S., Reuter, M., Rocha, E.A., Andrade-Navarro, M.A., Sehr, P., Pillai, R.S. & Carlomagno, T. (2011). Recognition of 2'-O-methylated 3'-end of piRNA by the PAZ domain of a Piwi protein. *Structure*, 19, 172-180

Zheng, K., Xiol, J., Reuter, M., Eckardt, S., Leu, N.A., McLaughlin, K.J., Stark, A., Sachidanandam, R., Pillai, R.S. & Wang, P.J. (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, 11841-46

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



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

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 defence 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.

Ribosomal complexes: targeting, translocation and quality control

Christiane Schaffitzel

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

The Schaffitzel team combines molecular biology, biochemistry, and cryo-electron microscopy, to study the ribosome-protein complexes involved in protein assembly, folding and targeting.

Previous and current research

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-electron microscopy (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 15), 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 101).

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. *Nat. Struct. Mol. Biol.* 18, 88-90.

Bingel-Erlenmeyer, R., Kohler, R., Kramer, G., Sandikci, A., Antolic, S., Maier, T., Schaffitzel, C., Wiedmann, B., Bukau, B. & Ban, N. (2008). A peptide deformylase-ribosome complex reveals mechanism of nascent chain processing. *Nature*, 452, 108-11

Schaffitzel, C., Oswald, M., Berger, I., Ishikawa, T., Abrahams, J.P., Koerten, H.K., Koning, R.I. & Ban, N. (2006). Structure of the *E. coli* signal recognition particle bound to a translating ribosome. *Nature*, 444, 503-6

Mitra, K., Schaffitzel, C., Shaikh, T., Tama, F., Jenni, S., Brooks CL, 3rd, Ban, N. & Frank, J. (2005). Structure of the *E. coli* protein-conducting channel bound to a translating ribosome. *Nature*, 438, 318-24

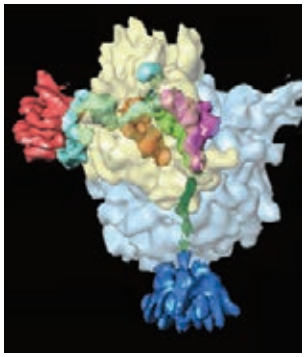


Figure 1: EM reconstruction of the ribosome nascent chain complex and the translocon. The 30S ribosomal subunit is shown in yellow, the 50S subunit in light blue, the A, P and E site tRNAs are in magenta, green and orange respectively. The mRNA is depicted in cyan and the nascent chain in the polypeptide 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

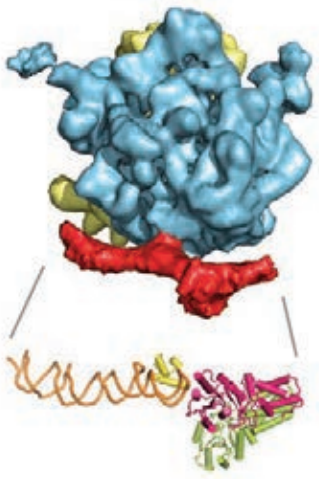
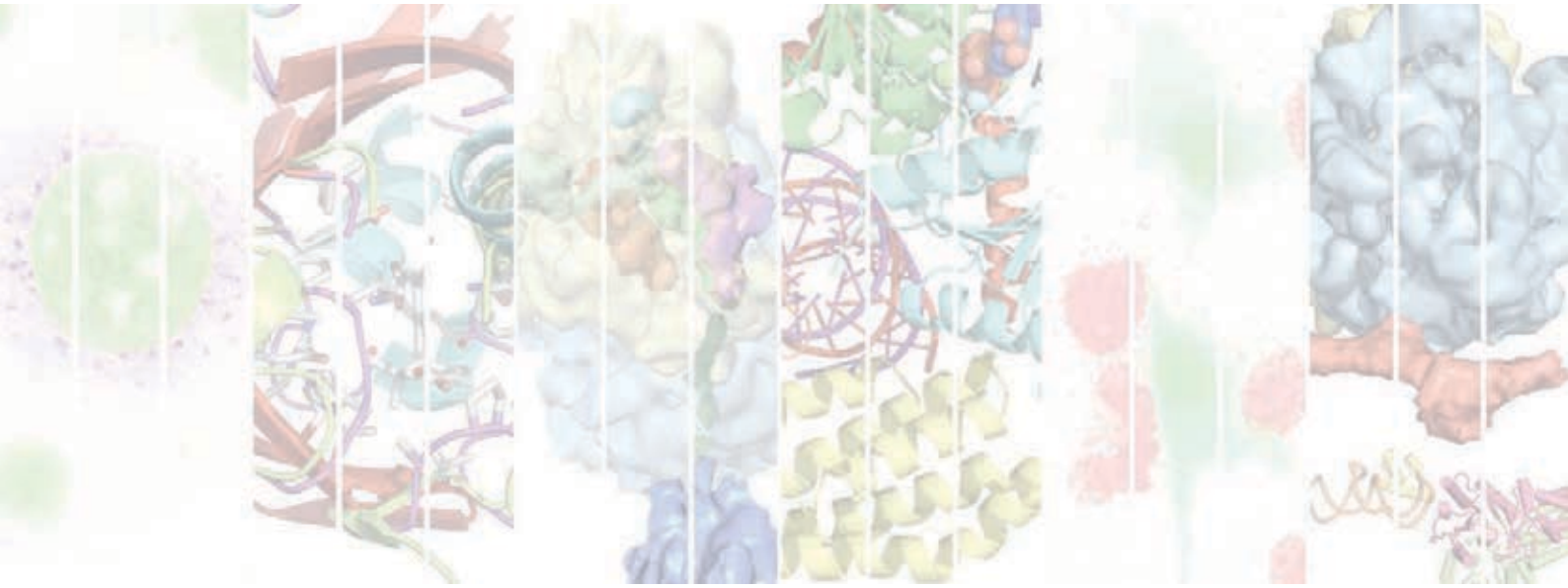
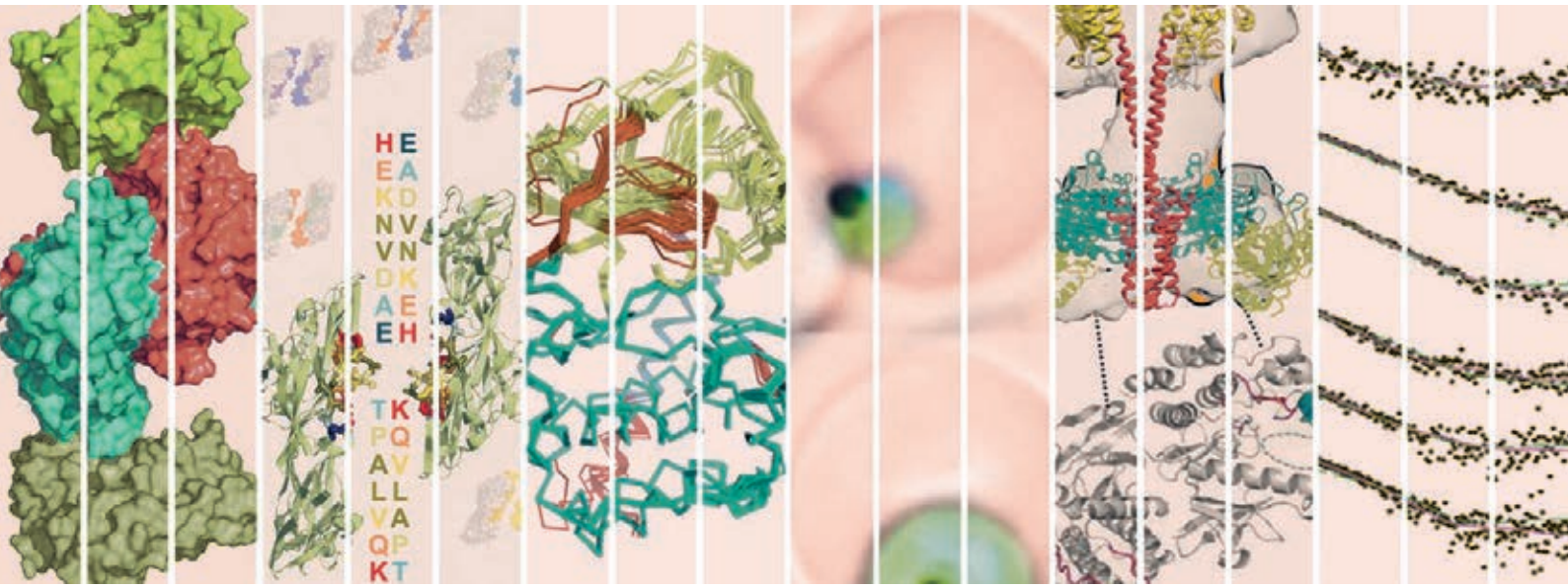


Figure 2: Cryo-EM structure of the ribosome (yellow/blue) bound to the signal recognition particle (SRP) and the SRP receptor (both in red). Below it is an atomic model of SRP (green-yellow/orange) and its receptor (pink)





Structural Biology

Activities at EMBL Hamburg focus on state-of-the-art structural biology methods using synchrotron radiation, combining cutting-edge technology with an ambitious research programme for structures of multifunctional proteins and protein complexes of biomedical relevance.

EMBL Hamburg's laboratories are on the German Synchrotron Research Centre (DESY) campus, with synchrotron radiation (DORIS III, PETRA III) and laser (FLASH) facilities available. In addition, a powerful X-ray Free Electron Laser is under construction. EMBL operates a new integrated facility, called EMBL@PETRA3, for applications in structural biology at the PETRA III ring. It comprises three state-of-the-art beamlines for macromolecular X-ray crystallography and small angle X-ray scattering of biological samples, complemented by facilities for sample preparation and characterisation, and data evaluation. EMBL Hamburg is also 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, protein assemblies in muscle cells, protein kinases, protein translocation into peroxisomes, and several projects relating to tuberculosis. Common to all projects is to make optimum use of on-site high-brilliance synchrotron radiation and to explore novel opportunities of the X-ray Free Electron Laser. 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. 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 example is the ARP/wARP package that allows automatic X-ray structure determination. 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

Structure and function of protein complexes in biological systems

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.

The Wilmanns group investigates a variety of protein-ligand complexes within the context of biological systems, employing a broad range of molecular and structural biology techniques.

The architecture of the protein interactome in sarcomeric muscle cells: Many proteins found in muscle cells, when dysfunctional, are associated with cardiovascular diseases. We investigate how large protein filament systems forming the overall architecture of 'sarcomeric units' in muscle cells – such as actin, myosin, nebulin, titin, myomesin and obscurin – 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) and subsequently unravelled key molecular rules for complex formation of sarcomeric proteins with β-sheet domains (Pinotsis *et al.*, 2009). Our future focus will be on novel protein interactions within the sarcomeric Z-disk and M-line region, and 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). We have also determined the structure of another kinase with apoptotic functions (death associated protein kinase) this time in the presence of CaM (figure 2), providing first insight into how CaM binding leads to kinase activation by withdrawing the autoregulatory domain from the kinase active site. Our goal is to complement 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 and, ultimately, to promote drug discovery against those kinases involved 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 in the cytosol. Proteins 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 (figure 1). Our present focus is on structural/functional studies of other protein components, including the Pex19p receptor, which recognises proteins integrated into the peroxisomal membrane. Our goal is to provide insight into the overall architecture of the peroxisomal translocon.

Structural systems biology in *M. tuberculosis*: We have determined the X-ray structures of a number of protein targets, some of them of known function and others unknown. 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). Using available structural data, we aim to use systems biology-orientated approaches (including proteomics, metabolomics and modelling) to investigate functional processes in living mycobacteria, with the aim of making data available to promote the development of new drugs, vaccines and diagnostic markers.

Selected references

Williams, C., van den Berg, M., Panjikar, S., *et al*. (2011). Insights into ubiquitin-conjugating enzyme/ co-activator interactions from the structure of the Pex4p:Pex22p complex. *EMBO J*., 31, 391-402

Due, A.V., Kuper, J., Geerlof, A., von Kries, J.P. & Wilmanns, M. (2011). Bisubstrate specificity in histidine/ tryptophan biosynthesis isomerase from *Mycobacterium tuberculosis* by active site metamorphosis. *Proc. Natl. Acad. Sci. USA*, 108, 3554-9

Poulsen, C., Akhter, Y., Jeon, A., *et al*. (2010). Proteome-wide identification of mycobacterial pupylation targets. *Mol. Syst. Biol.*, 6, 386

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

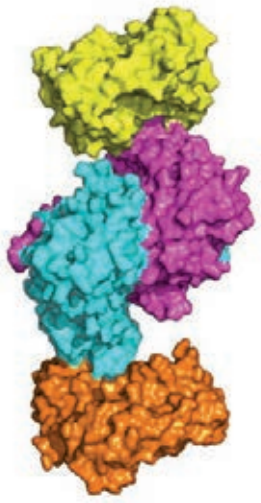


Figure 1: AGT-(Pex5p)2-AGT complex. Colour codes: AGT, yellow, orange; Pex5p, cyan, magenta. (Fodor et al*, unpublished)*

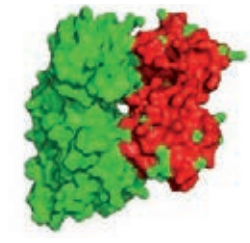


Figure 2: DAPK-CaM complex. Colour codes: DAPK, green; CaM, red. (de Diego et al*, 2010)*

Synchrotron instrumentation for structural biology beamlines at PETRA III

The Fiedler team focuses on the selection, customisation and integration of mechanics, control electronics and control software for X-ray based structural biology research.

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.

Selected references

Schultke, E., Kelly, M.E., Nemoz, C., Fiedler, S., Ogieglo, L., Crawford, P., Paterson, J., Beavis, C., Esteve, F., Brochard, T., Renier, M., Requardt, H., Dallery, D., Le Duc, G. & Meguro, K. (2011). Dual energy CT at the synchrotron: a piglet model for neurovascular research. *Eur. J. Radiol.*, 79, 323-7

Keyrilainen, J., Fernandez, M., Karjalainen-Lindsberg, M.L., Virkkunen, P., Leidenius, M., von Smitten, K., Sipila, P., Fiedler, S., Suhonen, H., Suortti, P. & Bravin, A. (2008). Toward high-contrast breast CT at low radiation dose. *Radiology*, 249, 321-7

Ristau, U., Di Castro, M., Pazos, A. & Fiedler, S. (2008). The concept of EMBL beamline control at PETRA III. *Proceedings of PCaPAC08*, MOZ02, 22-4

Roessle, M.W., Klaering, R., Ristau, U., Robrahn, B., Jahn, D., Gehrmann, T., Konarev, P., Round, A., Fiedler, S., Hermes, C. & Svergun, D. (2007). Upgrade of the small-angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg. *Journal of Applied Crystallography,* 40, S190-S194

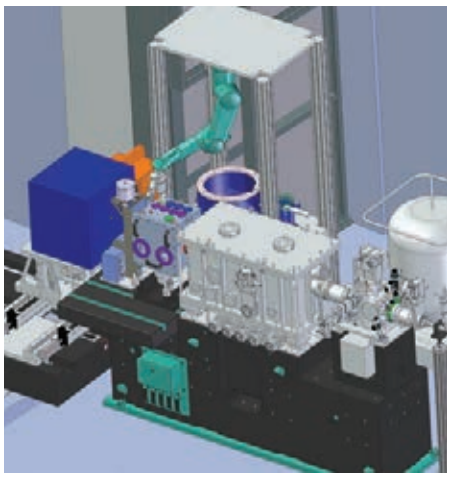
Previous and current research

EMBL has designed, is building, and operates three beamlines for structural biology at the PETRA III synchrotron radiation source on the DESY campus in Hamburg. The facilities will be dedicated to the leading techniques for X-ray-based structural research of biological samples – small angle X-ray scattering (SAXS), and macromolecular crystallography (MX). For the construction and operation of these facilities, our team provides expertise in X-ray optics, precision mechanical engineering, robotics, control software, and electronics. We are in charge of X-ray optical elements, and the experimental endstations, vacuum system, cryogenic system, control system, data acquisition system, technical infrastructure, and parts of the civil engineering.

In 2011, all three beamlines came into operation and first experiments have been successfully carried out. The remaining instruments necessary for the basic optical set-ups have been installed and commissioned. Notably these were the P13 horizontally deflecting mirror pair, several UHV slits and beam position monitor sets, PLC-based vacuum control, an automatic sample changer for SAXS, a high-precision diffractometer for MX, experimental tables for all endstations, and the first of three beam conditioning units. In addition, the cryogenic supply system for the beamline endstations has been prepared and the beamline control network has been designed and mostly implemented.

Examples for complete in-house developments are the construction of a focusing double multilayer monochromator for the BW7a beamline at the DORIS storage ring which serves as test platform for instruments for the PETRA III beamlines (see also Hermes group, page 114) or the development of nanometer resolution slits or the development of white beam monitors in collaboration with DESY. Another on-going project is the development of a robotic sample mounting system for protein crystals named MARVIN (see figure). The system is characterised by its high capacity, sample mounting speed, and flexibility and is integrated into a software-based control system which allows for a heterogeneous control environment and provides distributed access. A prototype is in user operation on the BW7b beamline at DORIS and adapted versions are under installation on the MX beamlines at PETRA.

Design of the P14 experimental endstation with sample mounting system MARVIN.



Future projects and goals

- Establishing stable user operation on all PETRA beamlines in parallel.
- Adaptive focusing mirror optics and control of beam size and shape.
- Integration of detector supports and the MARVIN system.
- Integrating beamline elements into a global instrument protection system.
- Further automation of alignment, sample handling and data acquisition
- Rapid feedback on positional/intensity variations of the incident beam.
- High flux polychromatic optics for P12.
- Crystallisation plate screening with the MX robotic sample changer.

In the longer term, time-resolved structural biology studies down to the microsecond will be performed on the new beamlines. We plan to develop the instruments and the level of synchronisation that are necessary for these kinds of experiments and explore the possibilities for biological imaging research at PETRA III.

Instrumentation for structural research

Christoph Hermes

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

Activities in the Hermes team include designing and constructing equipment for mechanical engineering, vacuum technology, X-ray optics, data acquisition and control electronics.

Previous and current research

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.

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 team, page 113). The opening of the European X-Ray Laser (XFEL) in Hamburg is scheduled for 2014, offering unprecedented research opportunities. Designing experiments that can 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.

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, 913-7

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, 372-7

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, 384-8

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

Integrative modelling for structural biology

The Lamzin group applies and develops cutting-edge computational methods and experimental approaches for structural studies of molecules of biological and medical interest.

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.

Knowledge of the function of biological systems depends on accurate structural information regarding their components - DNA, RNA, proteins, macromolecular complexes and assemblies. We apply and develop computational methods and experimental approaches for sample quality control, measurements and structural modelling using macromolecular crystallography at high and low resolution, electron microscopy and X-ray Free Electron Lasers (FEL). Interpretation of measured data and validation of derived models calls for novel approaches where information, obtained from complementary tools, is used to arrive at an integrated platform for a model of life.

Previous and current research

Methods for biological modelling at high and low resolution: Pattern-recognition methods are the basis of the group's main focus, the ARP/wARP software project (see Langer *et al*, 2008). We continually improve the comprehensive range of algorithms for protein/DNA/ligand X-ray crystal structure determination and develop new procedures for dealing with challenging problems (Hattne & Lamzin, 2011). We exploit inherent properties of macromolecular structures (Wiegels & Lamzin, 2012), and integrate additional information derived from biological databases. We can recognise structural motifs in low-resolution maps (Heuser *et al*, 2009) and contribute to the interpretation of electron microscopy data.

Structure-based drug design: Our recent developments allow for the probing of enzymatic binding sites and the in-silico screening of known ligands to provide new leads for diversity-based drug design. Our interest in this direction is stimulated by our research into the biology of pathogenic species associated with human morbidity and mortality, and is focussed on the probing of bacterial antibiotic resistance.

Biological imaging with Free-Electron Lasers: Breathtaking results from first diffractive imaging experiments using coherent FEL radiation (see Chapman *et al*, 2011 and Seibert *et al*, 2011) show the potential for imaging cellular organelles and understanding dynamics of complex formation at previously unimaginable resolution. In order to exploit many novel and unique opportunities for structural biology that will be provided by the European X-ray FEL source near DESY (www.xfel.eu), we are developing protocols for handling of biological samples for FEL experiments as well as novel computational methods for the interpretation of measured data.

Targets of biomedical interest: We integrate X-ray crystallography, lower resolution imaging, biochemistry and biophysics to study targets of biomedical interest. We research systems of the malarial parasite that are involved in biosynthesis (see Wrenger *et al*, 2011) and also study class I hydrophobins – functional amyloids of fungal origin – aiming to understand the fibril formation pathway of amyloid proteins.

Future projects and goals

The group will continue to focus on crystallographic software developments driven by general academic interest, availability of state-of-the-art beamlines at PETRA-III in Hamburg as well as projects of medical or biotechnological importance. Together with our collaborators from Russia, Sweden, and Germany, as well as from other EMBL groups, our research will expand into collaborative XFEL pilot projects as well as development of methods applicable to structural data obtained from various sources.

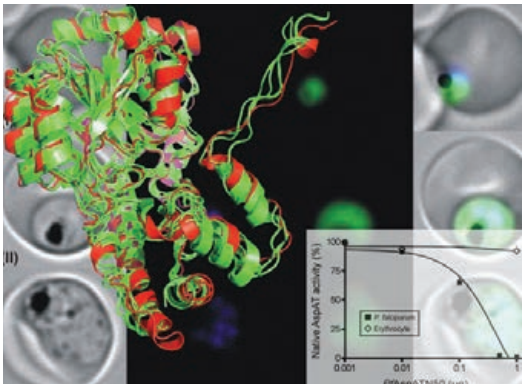
Selected references

Schmidt, A., Teeter, M., Weckert, E. & Lamzin, V.S. (2011). Crystal structure of small protein crambin at 0.48 Å resolution. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, 67, 424-8

Hattne, J. & Lamzin, V.S. (2011). A moment invariant for evaluating the chirality of three-dimensional objects. *J. R. Soc. Interface*, 8, 144-51

Wrenger, C., Muller, 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, 956-71

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, 1171-9



Specific inhibition of the aspartate aminotransferase of Plasmodium falciparum

Structural biology of cell surface receptors

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.

The Meijers group studies how clusters of molecules on a cell's surface rearrange themselves when they encounter another cell.

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 as 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, amplifying the recognition event, resulting in the activation of a signalling cascade within the cell that leads to physiological changes.

Immune receptors: T cell receptors (TCR) 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

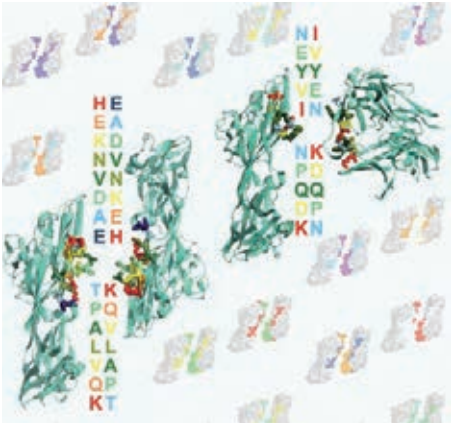
Viruses and bacterial pathogens use cell surface receptors to invade host cells, but they also deregulate the order established in the supramolecular cluster to jam the recognition machinery of the host. The same (de)regulation mechanism is used by the immune system itself which employs internal regulators 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.

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, 10799-804



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*

X-ray crystallography

The Müller-Dieckmann team uses a combination of protein characterisation tools, X-ray crystallography and SAXS to study unusual proteins and signal transduction across membranes.

Jochen Müller-Dieckmann

PhD 1994, Albert-Ludwigs-Universität, Freiburg.

Postdoctoral research at the University of California, Berkeley.

Associate Director, SGX, San Diego, until 2004.

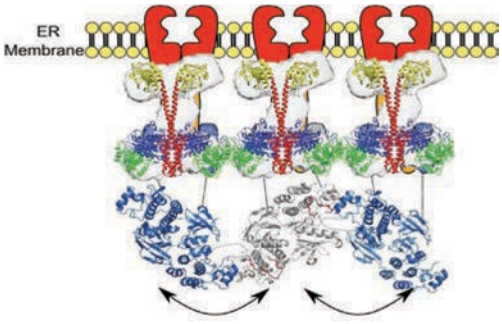
Team leader at EMBL Hamburg since 2004.

Selected references

Mayerhofer, H., Panneerselvam, S. & Müller-Dieckmann, J. (2012). Protein kinase domain of CTR1 from *Arabidospis thaliana* promotes ethylene receptor cross talk. *J. Mol. Biol.,* 415, 768-79

Bijtenhoorn, P., Mayerhofer, H., Müller-Dieckmann, J., Utpatel, C., Schipper, C., Hornung, C., Szesny, M., Grond, S., Thürmer, A., Brzuszkiewicz, E., Daniel, R., Dierking, K., Schulenburg, H., Streit, W.R. (2011). A novel metagenomic short-chain dehydrogenase/reductase attenuates pseudomonas aeruginosa biofilm formation and virulence on *Caenorhabditis elegans*. *PLoS ONE*, 6, e26278.

Meijers, R. & Mueller-Dieckmann, J. (2011). Advances in high-throughput crystallization. In *eLS 2011*, edited by John Wiley & Sons, LTD: Chichester.



Structure based model of ethylene receptor cross talk as mediated by CTR1 (bottom; grey and blue). The model is based on the X-ray structures of the His-kinase domain (red and blue) and the receiver domains (green), which were fit into the SAXS envelope of the cytosolic receptor domain. The unknown GAF domain (yellow) was taken form a homologue in the PDB

Previous and current research

Structural biology: Our goal is to understand the function of metagenome-derived proteins with unusual and highly specific properties, and that 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 understand their interaction with a common cytosolic downstream target, a protein kinase. The system regulates the response to the simple phytohormone ethylene.

Ethylene is a 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 units 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. CTR1 is one of only two confirmed direct recipients of ethylene receptor activity. Thus, this signalling pathway presents the interesting case, wherein a two-component signalling system manipulates a MAPKKK and possibly a MAPKKK signalling cascade. EIN2 constitutes another essential, membrane bound transducer of ethylene signalling.

In the absence of the hormone the receptor and therefore CTR1 are active. Hence, ethylene acts as an inverse agonist of its signalling pathway. The 3D structures of the active, tri-phosphorylated and the unphosphorylated, inactive kinase domain of CTR1 in complex with staurosporine, illustrate the conformational rearrangements that form the basis of activity regulation. Two highly specific CTR1 interfaces engage in interactions that promote CTR1 mediated crosstalk between ethylene receptor clusters. This model provides a structural foundation for the observed high sensitivity of plants to ethylene.

High-throughput crystallisation: EMBL Hamburg operates one of Europe's largest high-throughput crystallisation facilities, which is open to the general user community. Users come from 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.

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 three out of four sub-domains of the cytosolic ethylene receptor. The latter were placed in a SAXS based envelope of the cytosolic part of the receptor. Together with the active and inactive structures of CTR1, a model for the role of CTR1 mediated cross talk of the receptors was constructed.

Tools for structure determination and analysis

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.

The Schneider team works on the commissioning of three beamlines for structural biology at the new PETRA III synchrotron and the development of computational methods to extract the information from structural data.

Previous and current research

The three beamlines under commissioning by our team – in close collaboration with the Fiedler team (page 113) – 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. 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.

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 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 (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.

Future projects and goals

For the integrated facility for structural biology, our goal is to establish a user program on the beamlines during 2012. 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 experimental phase determination for 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 experimentally phased electron density maps obtained on the PETRA III beamlines.

For further information, see: **www.embl-hamburg.de/services/petra**.

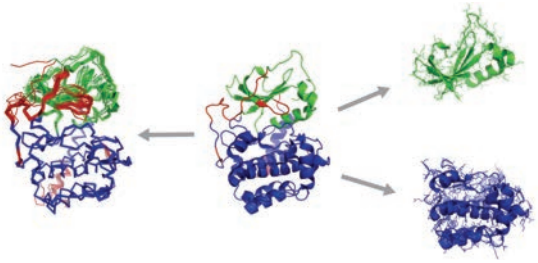
Selected references

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, 2269-75



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)

Small-angle X-ray scattering from macromolecular solutions

The Svergun group places special emphasis on hybrid methods combining SAXS with X-ray crystallography, NMR spectroscopy, and electron microscopy to improve the resolution and cross-validate structural models.

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

Verstraete, K., Vandriessche, G., Januar, M., Elegheert, J., Shkumatov, A.V., Desfosses, A., Van Craenenbroeck, K., Svergun, D.I., Gutsche, I., Vergauwen, B. & Savvides, S.N. (2011). Structural insights into the extracellular assembly of the hematopoietic Flt3 signaling complex. *Blood*, 118, 60-8

Rochel, N., Ciesielski, F., Godet, J., Moman, E., Roessle, M., Peluso-Iltis, C., Moulin, M., Haertlein, M., Callow, P., Mély, Y., Svergun, DI., Moras, D. (2011) Common architecture of nuclear receptor heterodimers on DNA direct repeat elements with different spacings. *Nat. Struct. Mol. Biol.*, 5, 564-70

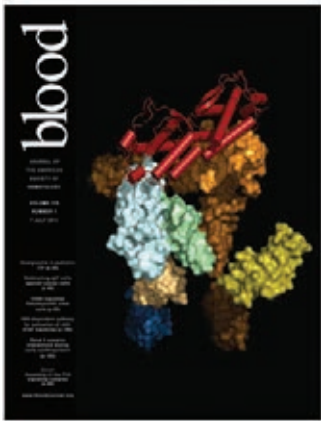


Figure 1: Quaternary structure of the extracellular assembly of the hematopoietic Flt3 signalling complex determined by crystallography, SAXS and electron microscopy (Verstraete et al*, 2011)*

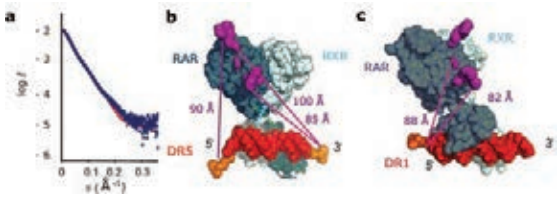


Figure 2: Rigid body models of the human nuclear hormone receptors constructed by SAXS and validated by Foster resonance energy transfer (FRET). (a) Experimental SAXS curve (dots) and the fit by the refined SAXS model (red line). (b,c) Relative positions of the different domains validated using FRET labels (Rochel et al*, 2011)*

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. The versatility and universality makes SAXS an ideal tool for structural systems biology applications.

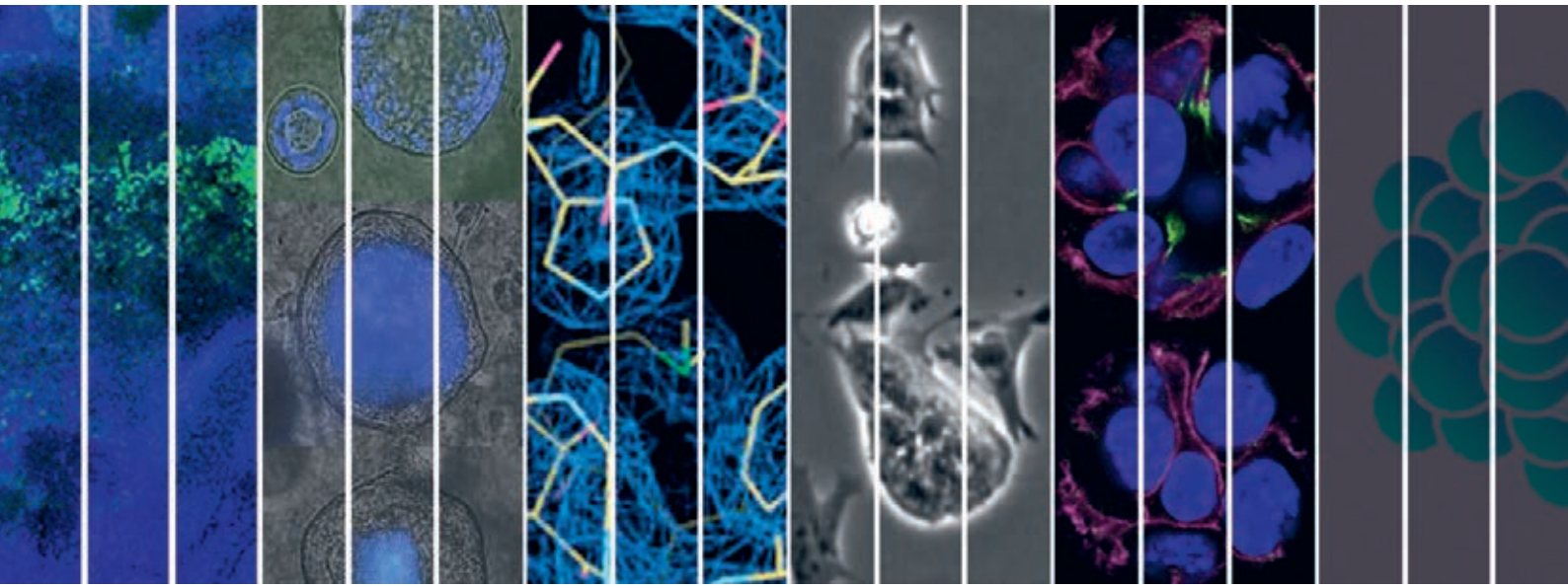
Our group leads the development of novel computational methods for constructing 3D structural models from the scattering data. Special attention is given to the joint use of SAXS with other structural, biophysical, biochemical and computational techniques including macromolecular crystallography, NMR, electron microscopy, neutron scattering and bioinformatics. We developed the world's most used program package, ATSAS, employed in over 1500 laboratories, and we continue to develop novel approaches for use by the scientific community.

We run a synchrotron beamline X33, dedicated to biological solution SAXS, at DESY's storage ring, DORIS-3. The rapidly-growing demand for SAXS has led to a sixfold increase in the user demand at X33 during the last decade. X33 is equipped with a robotic sample changer, a data analysis pipeline for building structural models online, and offers a remote user access. All these developments have already been integrated into the new high brilliance BioSAXS beamline P12 recently constructed at the PETRA-3 storage ring at DESY. P12 provides novel and exciting possibilities including microfluidic, high throughput, time resolved and anomalous biological SAXS.

Most of the external users of our beamlines are seeking collaborative projects where the SAXS group members help not only with data collection but also with modelling. In numerous exciting applications, SAXS is employed to study quaternary and domain structure of proteins (figure 1), nucleic acids and complexes (figure 2), oligomeric mixtures, conformational transitions upon ligand binding, flexible systems and intrinsically unfolded proteins, hierarchical systems, processes of amyloid fibrillation and many other objects of high biological and medical importance.

Future projects and goals

- Further development of novel methods and approaches for the reconstruction of tertiary and quaternary macromolecular structures from X-ray and neutron scattering data.
- Novel biological SAXS experiments on high-brilliance synchrotron beamlines including microfluidics, time resolved and cryoSAXS.
- Use of bioinformatics to construct and validate SAXS-based models and joint applications of SAXS with crystallography, NMR, FRET and others.
- Participation in collaborative projects with external users employing SAXS to study the structure of a wide range of biological systems in solution.
- Complete automation of biological SAXS experiment and data analysis, biophysical sample characterisation, and modern sample tracking databases.



Mouse Biology

Dynamic partnerships and exchanges with other international academic research and clinical centres, and participation in multiple EU-wide mouse research and informatics initiatives, are integral parts of our discovery process into genetics and genomics, cell biology and pathology.

The continued refinement of genetic and epigenetic perturbations of cellular and physiological function through the use of conditional and other increasingly finely engineered mouse mutations at the outstation is generating ever more accurate models of human disease and multigenic disorders.

Research groups at EMBL Monterotondo use these powerful tools to investigate wide-ranging aspects of mammalian biology, including gene expression, development and differentiation, cancer and regeneration, behaviour and sensory perception.

A state-of-art animal facility provides a full range of mouse transgenic and gene knock-out production, embryo rederivation and cryopreservation services, and a fully phenotyping suite. Other centralised facilities include histology, confocal microscopy and flow cytometry. 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 with the Centre for Genomic Regulation (CRG) in Barcelona add additional opportunities for translational research.

Philip Avner
Head of EMBL Monterotondo

Dynamics of epigenetic regulation

Philip Avner

PhD in yeast genetics, University of Warwick.

CNRS staff scientist

Head of the Mouse Molecular Genetics Unit at the Institut Pasteur since 1990

Head of the Developmental Biology Department at the Institut Pasteur since 2006

Head of EMBL Monterotondo since 2012.

The Avner group combines genetics, genomics, biochemistry, and cell biology to study the nature of the key complex on the X chromosome responsible for X-inactivation.

Previous and current research

The genetic material of the cell is not all equally available for transcription and this availability, which varies with cell type and developmental stage, is mediated largely by epigenetic modifications to the genome playing out mainly at the level of the chromatin. The double focus of our research has been on mouse genetics and epigenetics, with a particular focus on the interface between genetics and epigenetics, as revealed using X-inactivation as an experimental paradigm. X-inactivation, which occurs early during development in female mammalian embryos, ensures the dosage compensation between females carrying two copies of the X chromosome and males with a single copy of the X. Parts of the process in the mouse can be modelled *ex vivo* using female ES cells. We have been at the forefront of research into the characterisation and functional analysis of the different components of the X-inactivation centre, the key complex on the X chromosome for the initiation of X-inactivation. Ongoing research involves, for instance, the study of the Xce locus (X-controlling element), a genetic locus existing in different forms, which appears to influence which of the two X chromosomes will be chosen to undergo X-inactivation, and studies on the basis of the differences in the stability of the X-inactivation process occurring in the different extra-embryonic lineages of the mouse.

By carrying out this research we are not only discovering the multiplicity and plasticity of mechanisms that feed into this process of epigenetic control but also providing insights into the links between epigenetic processes and development. Our approach involves a combination of genetics, genomics, biochemistry and cell biology and both *ex vivo* and *in vivo* experimental approaches.

Future projects and goals

Objectives of our future work will be to gain further mechanistic insights into variations in the X-inactivation process that occur in different cell lineages and the sensitivity of the process to trans acting genetic factors. We will use our knowledge of the X-chromosome to explore more widely the contribution of epigenetic regulation to allele specific epistasis, the process where the effects of one gene are modified in an allele specific manner by one or several other genes.

Selected references

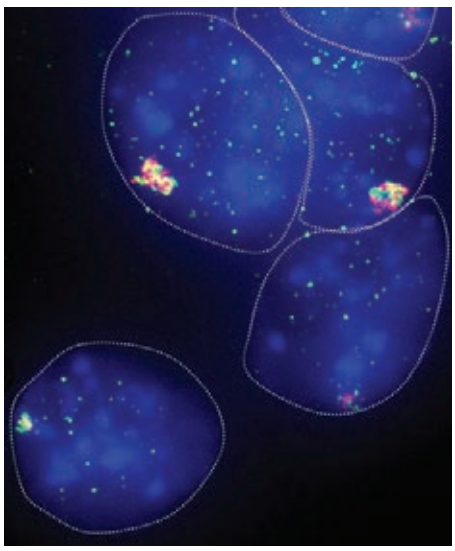
Chureau, C., Chantalat, S., Romito, A., Galvani, A., Duret, L., Avner, P. and Rougeulle, C. (2011) Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. *Human Molecular Genetics* 20(4), 705-18

Navarro, P., Oldfield, A., Legoupi, J., *et al*. (2010). Molecular coupling of Tsix regulation and pluripotency. *Nature* 468, 457-60

Navarro, P., Chambers, I., Karwacki-Neisius, V., Chureau, C., Morey, C., Rougeulle, C., Avner, P. (2008) Molecular coupling of Xist regulation and pluripotency. *Science,* 321, 1693-5

Duret, L., Chureau, C., Samain, S., Weissenbach, J., Avner, P. (2006) The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science*, 312, 1653-5

Mouse female trophoectoderm stem cells: The XIST non-coding RNA (green) shows partial overlap with the repressive H3K27Me3 histone mark on the inactive X chromosome (Morey et al.*)*



Developmental programming of anxiety

The Gross group uses pharmacological, histochemical, electrophysiological and behavioural genetic approaches to study the neural circuits underlying anxiety behaviour in mice.

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.

Selected references

Jain, A., Dvorkin, A., Fonio, E., Golani, I., Gross, C.T. (2012). Validation of the dimensionality emergence assay for the measurement of innate anxiety in laboratory mice. *Eur. Neuropsychopharmacol.*, 22, 153-63

Paolicelli, R.C., Bolasco, G., Pagani, F., Maggi, L., Scianni, M., Panzanelli, P., Giustetto, M., Ferreira, T.A., Guiducci, E., Dumas, L., Ragozzino, D., Gross, C.T. (2011). Synaptic Pruning by Microglia Is Necessary for Normal Brain Development. *Science*, 333, 1456-8. (Reviewed in: R. Ranshoff and B. Stevens. (2011) How many cell types does it take to wire a brain? *Science* 333, 1391-2)

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, 656-66

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, 130-3

Previous and current research

Anxiety disorders are debilitating mental illnesses characterised by excessive worry and rumination and exaggerated responses to threatening stimuli. Studies suggest that both genetic and environmental factors contribute to the prevalence of these disorders. For example, exposure to adverse events such as trauma during childhood is known to result in an increased risk of anxiety disorders in adulthood, 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. Several ongoing projects in the lab are addressing this question from different angles.

Early gene-by-environment risk factors: We are 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 such tools to examine the contribution of hippocampal and amygdala cell-types to anxiety and fear behaviour and are further developing these tools in combination with *in vivo* electrophysiology in awake behaving mice to dissect the circuits involved.

Future projects and goals

- Identification of the molecular mechanisms that mediate the long-term programming of behaviour by early environmental experiences in mice and humans;
- 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 tissue and cell-type specific suppression of neural activity in behaving mice;
- identification and validation of the neurophysiological correlates of anxiety with particular focus on hippocampus, amygdala, and medial hypothalamus;
- study of copy number variations as predisposing factors for disease in mice.

A better understanding of the molecular signals that trigger these long-term plastic mechanisms 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.

Molecular physiology of somatosensation

Paul Heppenstall

PhD 1997, University of Edinburgh.

Postdoctoral work at the Max Delbrück Centrum, Berlin.

Junior Professor at the Charité, Berlin.

Group leader at EMBL Monterotondo since 2008.

The Heppenstall group combines molecular, imaging and electrophysiological techniques to examine how sensory neurons turn information about touch and pain into electrical signals.

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.

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 calcium ions directly activate TRPA1 via an EF-hand domain in the N-terminus of the protein and that calcium is essential for normal activation of the channel by noxious chemicals. We are now interested in how TRPA channels have evolved to sense diverse stimuli across different phyla; for example, in snakes and insects TRPA1 orthologues are activated by warm temperatures. Using a combination of molecular and electrophysiological techniques we have mapped the regions in *Drosophila* TRPA1 that are responsible for sensing temperature and described how single TRPA1 channels are activated by heat.

Future projects and goals

- 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 de fined subpopulations of sensory neuron.
- Development of new techniques to measure functional properties of sensory neurons at their terminals.

The major focus of the laboratory is to correlate cellular studies on somatosensation with observations made at the physiological level. To this end we are developing genetic approaches that combined with electrophysiological and molecular imaging techniques, enable us to characterise sensory neurons *in situ*. A better understanding of sensory neuron function may ultimately lead to improved therapies for the treatment of chronic pain.

Selected references

Zurborg, S., Piszczek, A., Martinez, C., Hublitz, P., Al Banchaabouchi, M., Moreira, P., Perlas, E. & Heppenstall, P.A. (2011). Generation and characterization of an Advillin-Cre driver mouse line. *Mol. Pain*, 7, 66

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, 277-9

Wetzel, C., Hu, J., Riethmacher, D., Benckendorff, A., Harder, L., Eilers, A., Moshourab, R., Kozlenkov, A., Labuz, D., Caspani, O., Erdmann, B., Machelska, H., Heppenstall, P.A. & Lewin, G.R. (2007). A stomatin-domain protein essential for touch sensation in the mouse. *Nature*, 445, 206-9

Heppenstall, P.A. & Lewin, G.R. (2001). BDNF but not NT-4 is required for normal flexion reflex plasticity and function. *Proc. Natl. Acad. Sci. USA*, 98, 8107-12

Mechanisms of oncogene dependence and tumour relapse

The Jechlinger group uses a 3D culture system of primary mouse mammary epithelial cells to study cancer-initiating oncogenes.

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.

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, 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, 1841-44

Jechlinger, M., Sommer, A., Moriggl, R., Seither, P., Kraut, N., Capodiecci, P., Donovan, M., Cordon-Cardo, C., Beug, H. & Grunert, S. (2006). Autocrine PDGFR signaling promotes mammary cancer metastasis. *J. Clin. Invest*, 116, 1561-70

Jechlinger, M., Grunert, S., Tamir, I.H., Janda, E., Lüdemann, S., Waerner, T., Seither, P., Weith, A., Beug, H., Kraut, N. (2003). Expression profiling of epithelial plasticity in tumor progression. *Oncogene*, 16;22, 7155-69

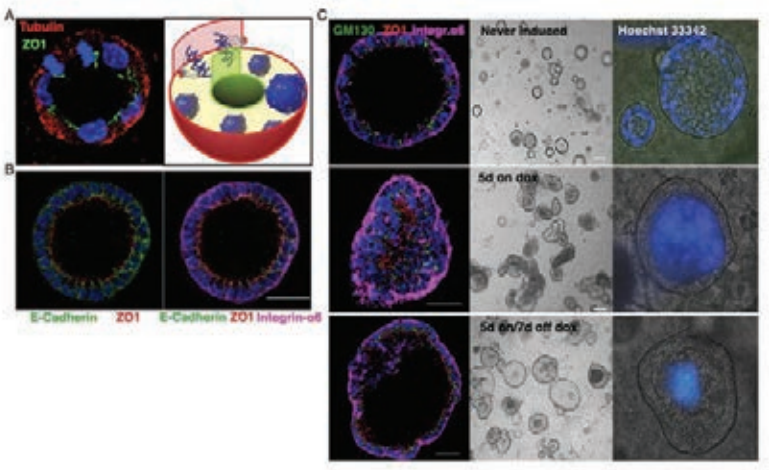
Previous and current research

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 tumour 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.



(A) One cell in anaphase divides with the sets of chromosomes perpendicular to the apical surface, while keeping ZO1 localized to the luminal membrane. (B) Confocal microscopy (5mm projection through the middle) shows epithelial cell polarity. E-Cadherin, (adherens junctions, lateral); ZO1, (tight junctions, apical); Integrin a6 (basolateral). (C) Doxycycline exposure causes loss of epithelial polarity and filling of the lumen; Removal of doxycycline results in survival of a re-polarised cell layer, that acquires the ability to exclude Hoechst 33342. Left panels: Confocal microscopy (5mm projection through the middle) shows Dapi, GM130 (apical), ZO1 (tight junction), Integrina6 (basal) at indicated times. Middle panels: Bright-field pictures show: (top) small, hollow acini; (middle) filled, irregular shaped spheres (bottom) hollow, irregular shaped spheres that show debris of internal cells. Right panels: Exclusion of Hoechst 33342 (1 hour incubation) at indicated times

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

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 2011.

The Lancrin group studies the haematopoietic system and looks to develop strategies to improve methods for generating blood cells from stem 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 the 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

Ferreras, C., Lancrin, C., Lie-A-Ling, M., Kouskoff, V. & Lacaud, G. (2011). Identification and characterization of a novel transcriptional target of RUNX1/AML1 at the onset of hematopoietic development. *Blood*, 118, 594-7

Lancrin, C., Sroczynska, P., Serrano, A.G., *et al*. (2010). Blood cell generation from the hemangioblast. *J. Mol. Med*., 88, 167-72

Hoogenkamp, M., Lichtinger, M., Krysinska, H., *et al.* (2009). Early chromatin unfolding by RUNX1: a molecular explanation for differential requirements during specification versus maintenance of the hematopoietic gene expression program. *Blood*, 114, 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, 892-5

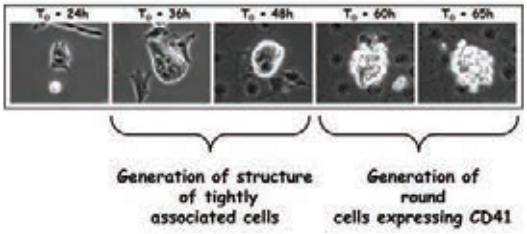


Figure 1: Time-lapse microscopy analysis of haemangioblast differentiation. The formation of a blast colony from the haemangioblast can be retrospectively divided in 2 consecutive phases: a generation of a structure of tightly associated endothelial cells and the production of round non adherent cells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

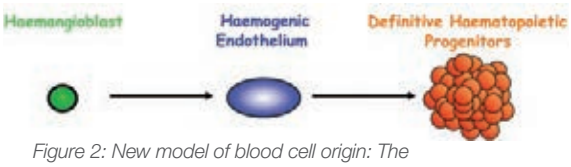


Figure 2: New model of blood cell origin: The haemangioblast and the haemogenic endothelium are part of the same developmental pathway to generate blood cell progenitors during embryonic life

Non-coding RNA function in mammalian development and physiology

The O'Carroll group studies mouse blood cell formation, embryology and germ cell development using state-of-the-art genetic strategies and high-throughput sequencing approaches.

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.

Selected references

De Fazio, S., Bartonicek, N., Di Giacomo, M., *et al.* (2011). The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature*, 480, 259–63

Rasmussen, K.D. & O'Carroll, D. (2011). The miR-144/451eGFP allele, a novel tool for resolving the erythroid potential of hematopoietic precursors. *Blood*, 118, 2988-92

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, 15163-68

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, 1351-8

Previous and current research

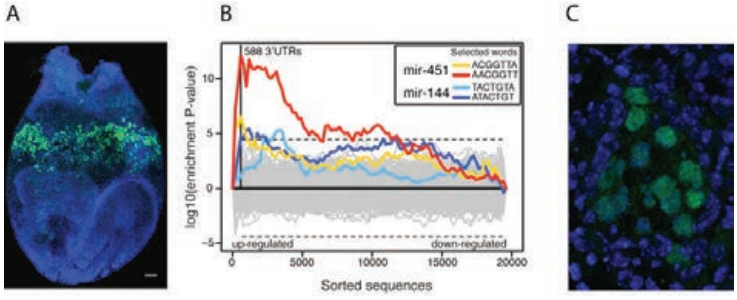
The aim of the group is to understand the mechanisms by which non-coding RNAs contribute to mammalian development and physiology with an emphasis on mouse hematopoiesis, embryology and germ cell development. We address our challenges using state-of-the-art mouse genetic strategies coupled with high throughput 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, and transposon silencing to the establishment of heterochromatin domains. 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. Our current RISC interests now focus on the physiological importance of Ago2 post-translational regulation *in vivo*.

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.

Transcriptome sequencing has established long non-coding RNAs (lncRNAs) as a distinct class of genes that encompass abundant and diverse classes of cellular RNAs, mostly of unknown function. The defining feature of lncRNAs is simply that their length exceeds 200 nucleotides. These RNAs are derived from both intergenic or genic regions of the genome. Similar to their coding counterparts lncRNAs can undergo extensive processing, such as capping, splicing, trans-splicing and polyadenylation. LncRNAs have been shown to affect gene expression via epigenetic, transcriptional or post-transcriptional mechanisms. We currently investigate the role of these newly identified genomic outputs in germ cell development and homeostasis.

Figure: A. Single cell visualisation of the miR-144/451 locus activity in E8.5 mouse embryos. A miR-144/451 EGFP BAC transgenic E8.5 mouse embryo is shown, DAPI in blue and EGFP in green. EGFP expression is found in the nascent blood islands.
B. Sylamer analysis of transcriptomes from wildtype and miR-144/451 -/- erythroblasts reveals miR-451 targets and the importance of miR-451 within the miR-144/451 locus (Collaboration with Enright Group).
C. Nuclear Miwi2 localisation in gonadocytes from fetal testis of Slicer-inactive Mili ADH mice. DAPI in blue and Miwi2 in green



Future projects and goals

- Determine the *in vivo* significance of post-translational modifications of Ago2.
- Explore the function of Miwi2 and Mili during spermatogenesis beyond transposon silencing.
- Understand the contribution of lncRNAs to germ cell development and homeostasis.

Mitotic chromosomal instability and oncogene dependence

Rocio Sotillo

PhD 2002, Autónoma University (Madrid) and Spanish National Cancer Center (CNIO), Spain.

Post doctoral work at Memorial Sloan Kettering Cancer Center, New York, USA.

At EMBL since September 2010.

ERC investigator since 2011.

Sotillo's research uses genetics and 3D cell culture systems to study the molecular mechanisms that lead to chromosomal instability and its consequences in tumour initiation and suppression.

Previous and current research

Chromosomal instability (CIN), the inability to correctly segregate sister chromatids during mitosis, is a hallmark of cancer cells. Overexpression of the mitotic checkpoint protein Mad2, commonly found in human tumours, leads to CIN and the development of aneuploid tumours in mouse models. Moreover, CIN can facilitate escape from oncogene addiction (the dependence of tumour cells on their initiating lesion for survival) and may be responsible for tumour relapse after targeted therapies. Very little is known about the mechanism of how and when CIN promotes tumour relapse. Our lab will focus on understanding the molecular mechanisms that lead to CIN and the consequences it may have in tumour initiation, suppression and relapse, hoping that the genes or proteins identified could be targeted therapeutically. We will use a combination of mouse genetics and highly innovative 3D *in vitro* culture systems.

Future projects and goals

- To study the dependence of tumour cells on the mitotic checkpoint *in vivo* and evaluate the potential for therapeutic interference with mitotic checkpoint genes.
- To study the effects of tumour regression and recurrence in chromosomally unstable tissues.
- To analyse the dual role of chromosome instability (tumour initiating and tumour suppressive) dependent on levels of aneuploidy, tissue type and molecular nature of the cooperating lesion(s).

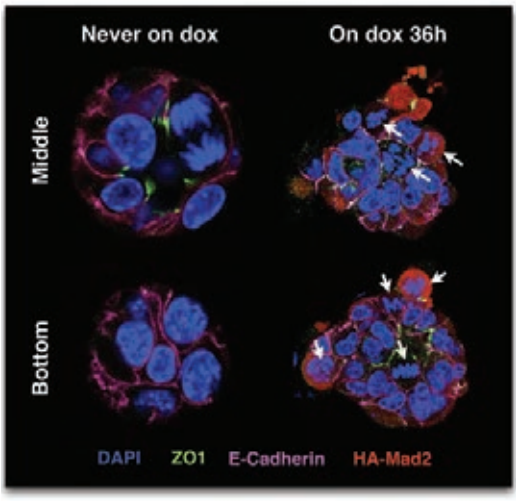
Selected references

Schvartzman, J.M., Duijf, P.H., Sotillo, R., Coker, C. & Benezra, R. (2011). Mad2 is a critical mediator of the chromosome instability observed upon Rb and p53 pathway inhibition. *Cancer Cell*, 19, 701-714

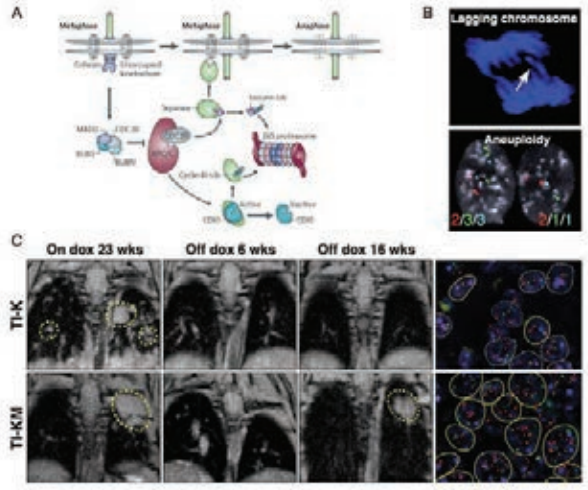
Sotillo, R., Schvartzman, J.M., Socci, N.D. & Benezra, R. (2010). Mad2-induced chromosome instability leads to lung tumour relapse after oncogene withdrawal. *Nature*, 464, 436-440

Schvartzman, J.M., Sotillo, R. & Benezra, R. (2010). Mitotic chromosomal instability and cancer: mouse modelling of the human disease. *Nat. Rev. Cancer*, 10, 102-115

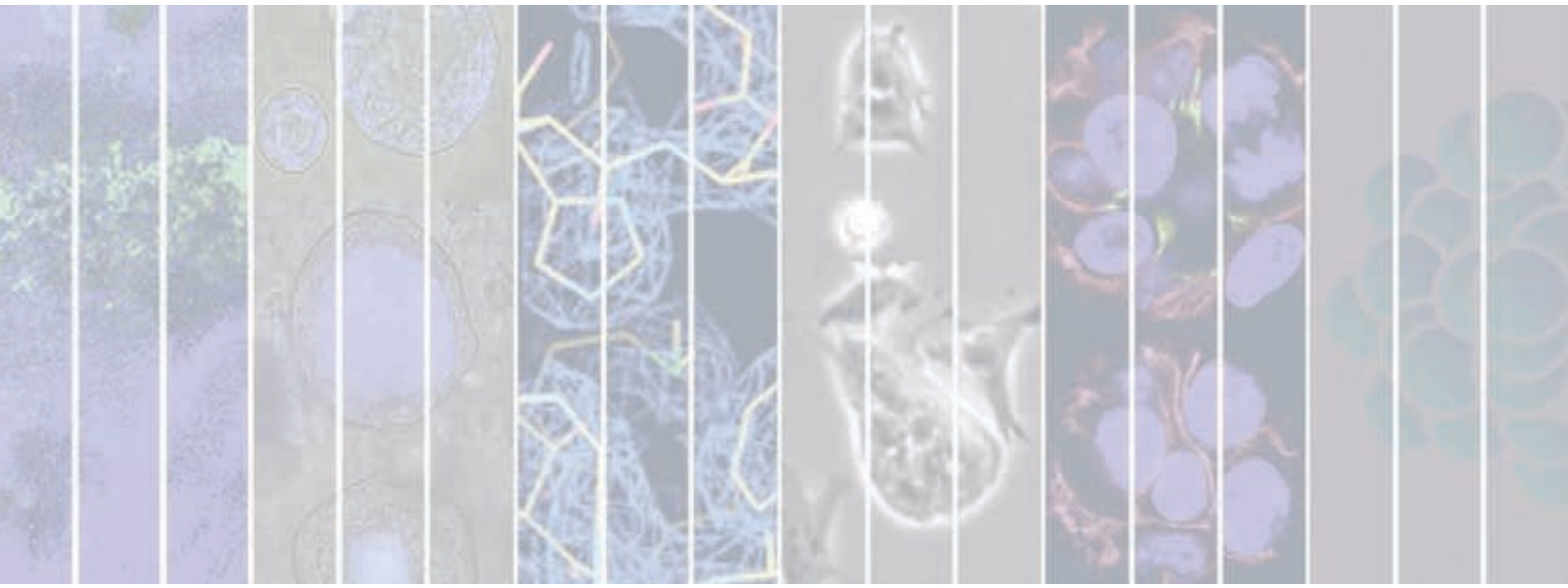
Sotillo, R., Hernando, E., Diaz-Rodriguez, E., Teruya-Feldstein, J., Cordon-Cardo, C., Lowe, S.W. & Benezra, R. (2007). Mad2 overexpression promotes aneuploidy and tumorigenesis in mice. *Cancer Cell*, 11, 9-23



Primary mammary cells from TetO-Mad2/TetO-Myc/MMTV-rtTA mice grown in 3D culture. Left panel shows never induced cells that begging to form a polarized acinus. Right panel: loss of epithelial cell polarity in an acinus grown from tritransgenic cells and exposed to doxycycline for 36h. White arrows show abundant mitotic cells and lagging chromosomes after Mad2 overexpression



A) The mitotic checkpoint complex (Mad2, Cdc20, BubR1 and Bub3) is shown to inhibit the anaphase promoting complex/cyclosome (APC/C) until the last kinetochore is correctly attached to microtubules. B) Evidence of lagging chromosomes and aneuploidy measured by FISH analysis on cells overexpressing Mad2. C) MR images of Kras transgenic mice (TI-K) and Kras+Mad2 (TI-KM) mice on doxycycline at indicated times showing lung tumours (yellow circles) (left panel), after 2-6 weeks of doxycycline withdrawal showing complete regression (middle panel) and after 14-16 weeks off doxycycline (right panel) showing recurrent tumours in mice that overexpress Mad2. Moreover recurrent tumours from these mice are highly aneuploid



Index

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