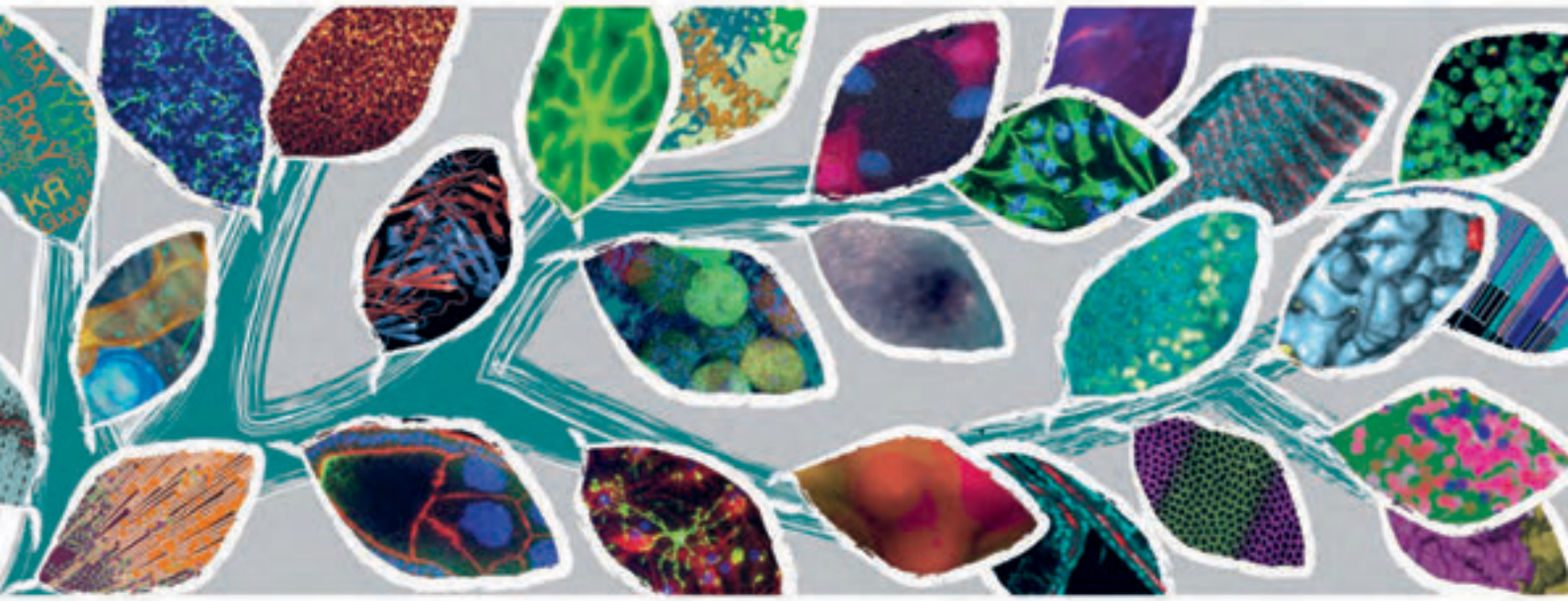
European Molecular Biology Laboratory

Research at a Glance 2013



EMBL

European Molecular Biology Laboratory

Research at a Glance 2013

EMBL

www.embl.org

Contents

The vision of the nations that 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 almost four decades, EMBL has grown and developed substantially, and its member states now number 21, including one 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. Both junior and senior researchers have received some of the most prestigious European grants and awards.

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. EMBL's fixed-term contract system means that when 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. This spirit of working together is also reflected in the numerous external collaborations and partnerships.

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

Research topics


pages 74–91
pages 100–111
pages 112–119
pages 120–129
Janet Thornton
Alex Bateman
Pedro Beltrão
Paul Bertone
Ewan Birney
Alvis Brazma
Anton Enright
Paul Flicek
Nick Goldman
John Marioni
John Overington
Julio Saez-Rodriguez
Oliver Stegle
Christoph Steinbeck
Sarah Teichmann
Stephen Cusack
Imre Berger
Florent Cipriani
Darren Hart
José Márquez
Andrew McCarthy
Daniel Panne
Ramesh Pillai
Christiane Schaffitzel
Matthias Wilmanns
Stefan Fiedler
Victor Lamzin
Rob Meijers
Thomas Schneider
Dmitri Svergun
Philip Avner
Cornelius Gross
Paul Heppenstall
Martin Jechlinger
Christoph Lancrin
Dónal O'Carroll
Rocio Sotillo
Cell signalling and cell differentiation
Cellular organisation and dynamics, and cell division
Chemistry and chemical biology
Computational genomics and metagenomics
Computational modelling of biological systems and processes
Disease mechanisms, pathogens, molecular medicine, stem cells
Evolution
Functional genomics, genetics and gene networks
Gene regulation, transcription, chromatin and epigenetics
Imaging and image analysis
Macromolecular complexes, interaction networks
Neurobiology
Physics and biophysics
Plant biology
Proteomics
RNA metabolism, transport and processing, ncRNAs and miRNAs
Robotics and automation, engineering
Software development and bioinformatics
Tissue morphogenesis, cell polarity and migration
X-ray crystallography, NMR, electron microscopy

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, working across the laboratory's five sites.

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:

- Undertaking 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, EMBL-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 bilathe bilateral links between EMBL scientists and local research institutions.

Hinxton

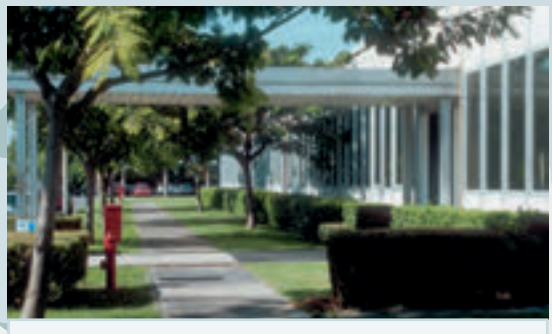
Hamburg

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.

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.

Students have the opportunity to obtain joint PhD degrees between EMBL and one of its partner universities or from a recognised university of their choice.

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**.

An international, interdisciplinary and collaborative workplace

EMBL's staff comprises in excess of 1700 people from more than 60 different countries – this internationality creates an atmosphere that is creative, interdisciplinary and collaborative, with an unparalleled breadth of expertise and complementary skills.

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.

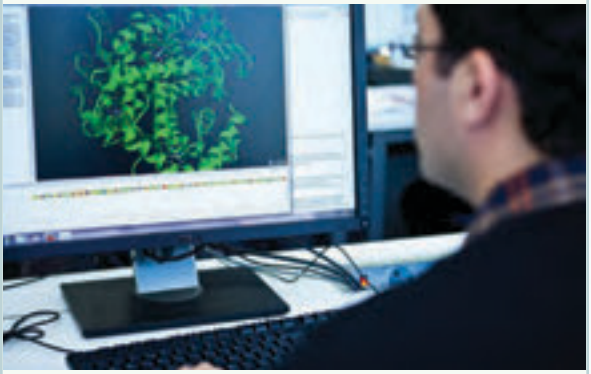
Group and team leaders

EMBL fosters the pursuit of ambitious and long-term 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
KxxR
RG
RxG
KxxD
RxxY
GXxN
DxK
GR
GXxR
AxK
RxxG
YG
SxY
GxxM
DY
KA
DxR
DxY
YxxR
YxxQ
GxxI
YxxY
YS
YRRY

Directors' Research

Directors' Research covers two thematically distinct research groups, headed by the Associate Director of EMBL and the Director of EMBO, an organisation of more than 1500 leading researchers that promotes excellence in the life sciences.

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 rearrange 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.

Cytoplasmic gene regulation and molecular medicine

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.
ERC Advanced Investigator since 2011.

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

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.

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**.

Selected references

Castello, A., *et al*. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, 149, 1393-1406

Galy, B., *et al*. (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, 423-6

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

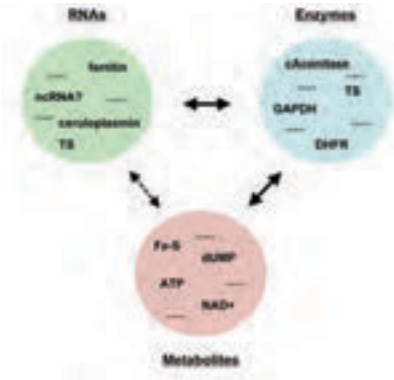


Figure 1: Exploring REM networks

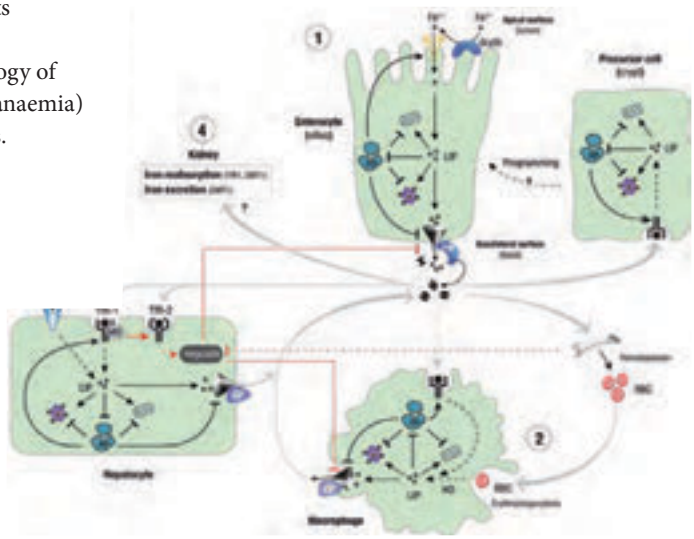


Figure 2: Systems biology of mammalian iron metabolism

Visualising complex cell shapes and signalling pathways

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

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.

Professor, University of Cologne, Institute of Genetics.

Director of EMBO and group leader at EMBL since 2010.

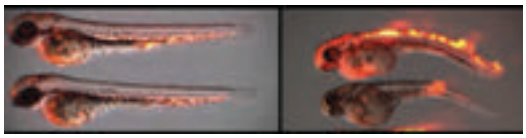
Selected references

Hocevar Brezavscek, A., Rauzi, M., Leptin, M. & Ziherl, P. (2012). A model of epithelial invagination driven by collective mechanics of identical cells. *Biophys. J.*, 103, 1069-77

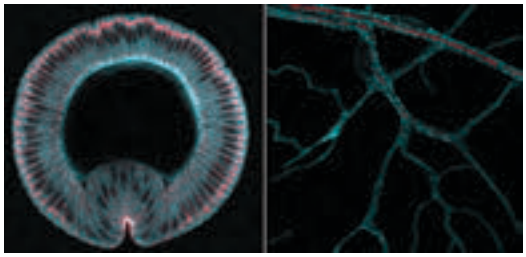
Mathew, S.J., Rembold, M. & Leptin, M. (2011). A role for Traf4 in polarizing adherens junctions as a prerequisite for efficient cell shape changes. *Mol. Cell Biol.*, 31, 4978-93

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.

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

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.

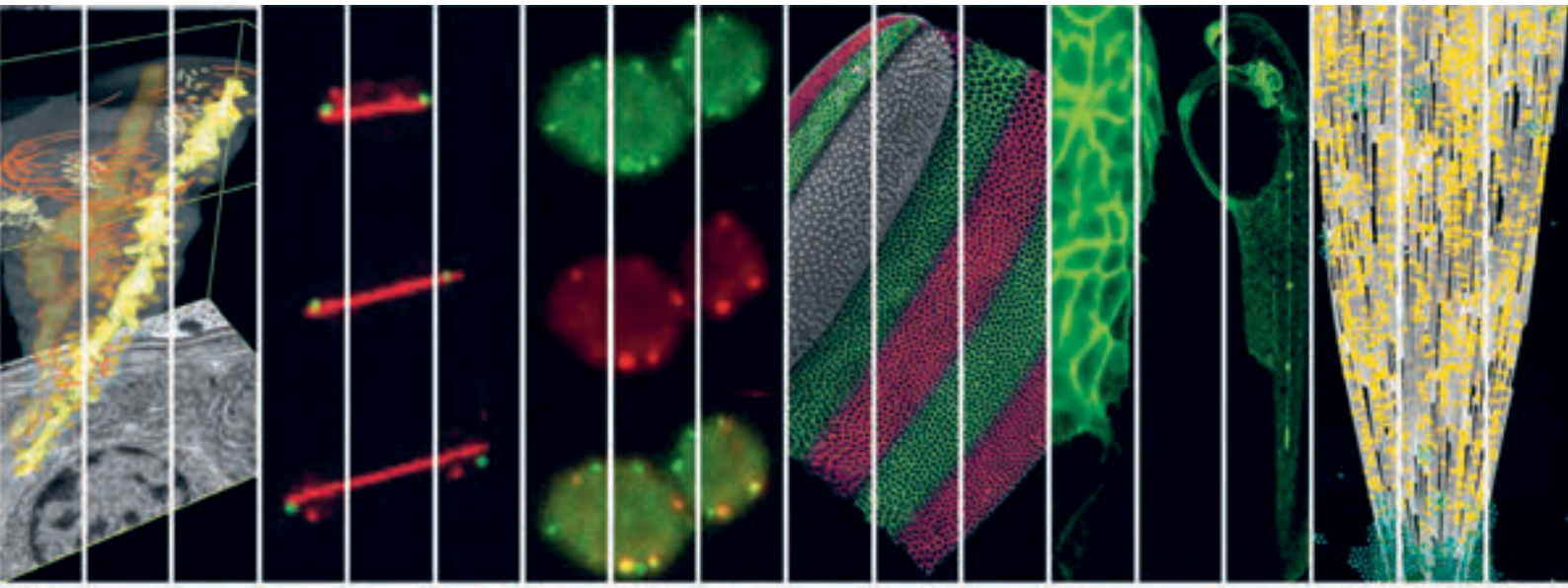
Cell shape determination during development

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 reorganisation 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 through 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 multicolour 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.



Cell Biology and Biophysics

In this 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 short, 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 and employing 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 in 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.

PhD and 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 and early embryos 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 and first mitotic divisions of mammalian oocytes and embryos, 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 and single molecule techniques to study protein function in live cells in high throughput.

For oocyte meiosis and early embryonic mitosis, we are pursuing the molecular mechanism of homologous chromosome segregation and are developing new gene silencing methods and light sheet-based imaging systems to establish a 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., & Ellenberg, J. (2011). Complete kinetochore tracking reveals error-prone homologous chromosome biorientation in mammalian oocytes. *Cell*, 146, 568-81

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

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

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

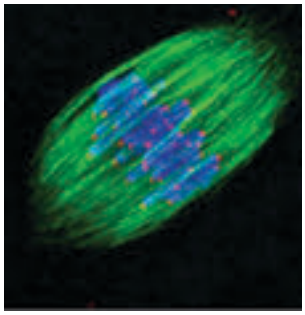


Figure 1: Meiotic spindle of a mouse oocyte. Chromosomes (blue) are bioriented by the meiotic spindle microtubules (green) at their kinetochores (red)

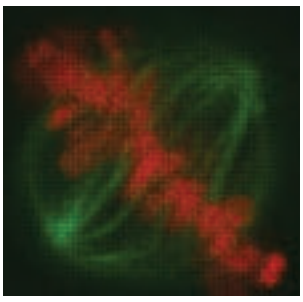


Figure 2: Mitotic spindle composed of thousands of microscopy images of human cells in which individual genes were silenced. Chromosomes (red) are made of images from genes that affect their segregation, while the mitotic spindle (green) is composed of images from genes affecting its assembly

Multicellular morphogenesis

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.

Darren Gilmour

PhD 1996, University of Cambridge.

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

Group leader at EMBL Heidelberg since 2004.

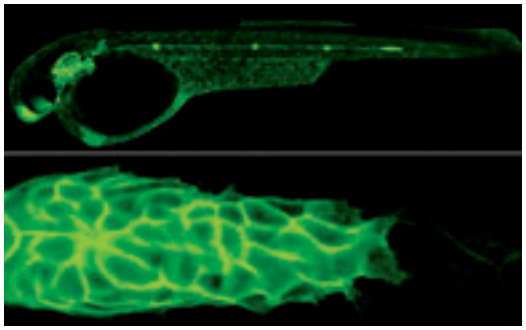
Selected references

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

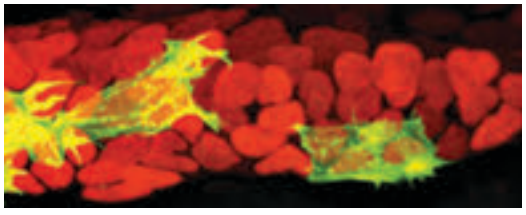
Friedl, P., Gilmour, D. (2009). Collective cell migration in morphogenesis, regeneration and cancer. *Nat. Rev. Mol. Cell Biol.,* 10, 445-57.

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, 2695-705

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, 1026-31



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



Visualising actin dynamics (LifeAct-GFP) within migrating primordium

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 feeds back 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 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.

Chromosome structure and dynamics

Christian Häring

PhD 2003, Institute of Molecular Pathology, Vienna.

Postdoctoral research at the University of Oxford.

Group leader at EMBL Heidelberg since 2007.

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

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 changes is the transformation of interphase chromatin into rod-shaped mitotic chromosomes in preparation for cell division. This process, known as chromosome condensation, is a key step for the successful segregation of chromosomes during mitosis and meiosis. However, the underlying mechanisms are still poorly understood.

The overall aim of our research is to unravel the action of molecular machines that organise the 3D arrangement of chromosome fibres. Insights into the general working principles behind these machines 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 central players in the formation of mitotic chromosomes is a highly conserved multi-subunit protein complex, known as condensin. We have shown that condensin binds to mitotic chromosomes by encircling chromosomal DNA within a large ring structure formed by its 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).

In an independent project, we have started the search for additional players that direct the formation of mitotic and meiotic chromosomes. We have developed a time-resolved light microscopy assay that allows us to quantitatively measure chromosome condensation in live fission yeast cells in high throughput (figure 2).

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, we are taking further advantage of structural and chemical biological as well as biophysical techniques to discover how condensin loads onto chromosomes, how it interacts with other chromosomal components, and how its activity is controlled.

Using the live-cell chromosome condensation assay, we are now screening large pools of fission yeast mutants to identify novel components of the chromosome condensation machinery. With the first candidates in hand, we will investigate the bases of their functions on mitotic chromosomes using the full repertoire of cellular and molecular biology.

Selected references

Petrova, B., Dehler, S., Kruitwagen, T., Heriche, J.K., Miura, K. & Haering, C.H. (2012). Quantitative analysis of chromosome condensation in fission yeast. *Mol. Cell Biol*., e-pub ahead of print

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

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

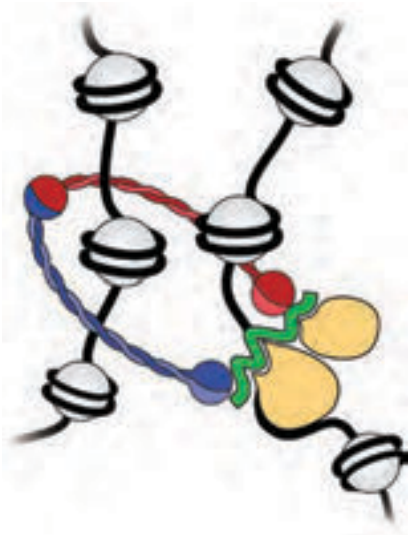
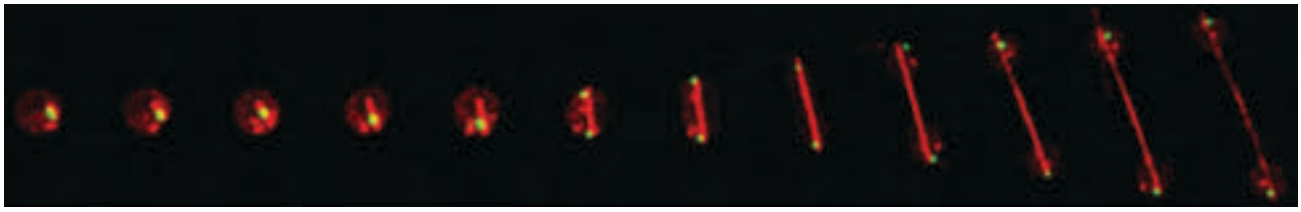


Figure 1 (above): Model for the organisation of mitotic chromosomes by condensin rings

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



Dynamics of cell growth and tissue architecture

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.

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.

Selected references

Krzic, U., Gunther, S., Saunders, T.E., Streichan, S.J. & Hufnagel, L. (2012). Multiview light-sheet microscope for rapid *in toto* imaging. *Nat. Methods*, 9, 730-3

Capoulade, J., Wachsmuth, M., Hufnagel, L. & Knop, M. (2011). Quantitative fluorescence imaging of protein diffusion and interaction in living cells. *Nat. Biotechnol.*, 29, 835-9

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

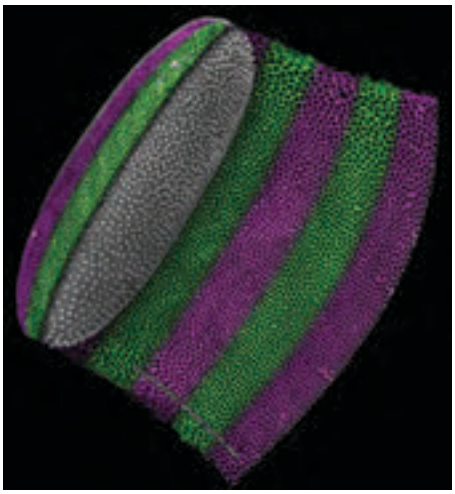
Previous and current research

Biological processes are highly dynamic and span many temporal and spatial scales. During development, cells must integrate and respond to a multitude of biochemical and biophysical signals: for example, changes in intracellular signalling networks, cytoskeleton remodelling, cell shape changes, long-range signalling and tissue remodelling. A whole-embryo view of morphogenesis with subcellular resolution is essential for unravelling the interconnected dynamics at varying scales of development – from interactions within cells to those acting across the whole embryo. Bridging scales from the submicron to the millimeter range with a temporal resolution of several seconds (combined with a total imaging time of several hours) not only poses tremendous challenges for modern microscopy methods but also requires powerful computational approaches for data handling, processing, and image analysis.

The central question that we are interested in is how a complex multi-cellular tissue or organism is formed from individual cells by spatio-temporal regulation of biophysical and intracellular signalling processes. We address all experimental steps, from innovative transgenic lines and microscope development to systematic image processing and biophysical modelling. This requires a multidisciplinary environment of biologists, physicists and computer scientists working closely together.

In order to address these questions we develop novel imaging techniques based on Selective Plane Illumination Microscopy (SPIM). SPIM yields optical sectioning by uncoupling the optical path for sample illumination from emitted photon detection. The illumination branch creates a thin light sheet to illuminate a specimen from the side and the emitted light is collected and imaged onto a high speed and high sensitivity camera by a second objective lens. The unprecedented speed of light sheet-based microscopy poses challenges for data handling and image processing, which we address by developing novel image processing tools.

Currently, we investigate cell shape changes and growth patterns in the *Drosophila* embryo with emphasis on the role of mechanical constraints on organ formation and tissue differentiation, complemented by mammalian cell culture studies investigating cell cycle response of an epithelial tissue to external and internal mechanical perturbations.



MuVi-SPIM image of a Drosophila *embryo. Eight views were fused to yield an* in toto *reconstruction of the embryo (one side membrane unrolled). The high speed of the microscope enables a detailed reconstruction of cell lineage and shape changes over extended periods of development*

Future projects and goals

Our research interests are focused 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. We will continue to not only tailor light-sheet microscopes to match specific biological questions, but also push the boundaries of light-sheet microscopy towards high speed intracellular imaging with extremely thin light sheets, super-resolution techniques, and quantitative *in toto* imaging.

Dynamics of membrane trafficking

Marko Kaksonen

PhD 2002, University of Helsinki.

Postdoctoral research at the University of California, Berkeley.

Group leader at EMBL Heidelberg since 2006.

Using budding yeast as a model, the Kaksonen group wants to understand how complex molecular machineries drive vesicle trafficking.

Previous and current research

Many biological processes at the cellular level are based on complex networks of macromolecular interactions. These networks have a modular organisation, where the different modules form dynamic molecular machines that drive processes such as signalling, cell motility, cytokinesis, and vesicle trafficking. Our group'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 more than 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 (for example particle tracking, FRAP, FCS/FCCS, high-throughput microscopy) in combination with powerful yeast genetics. We also use correlated light and electron microscopy to gain nanometre 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.

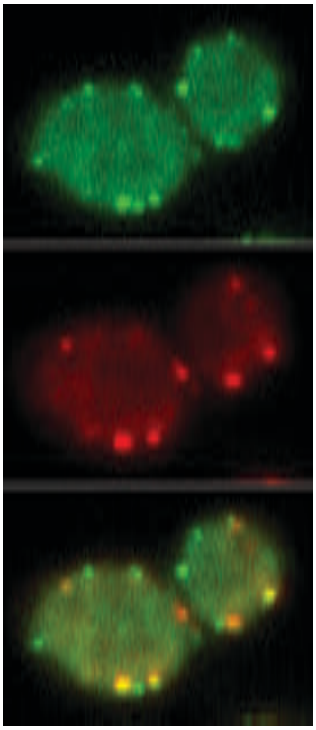
The core membrane trafficking events, such as the clathrin-mediated endocytosis, are elemental cellular processes that are involved in multiple biological phenomena ranging from cell polarisation to neural plasticity. As most of the yeast trafficking proteins are widely conserved in eukaryotes, we believe that mechanisms that we unravel in yeast cells will be applicable to eukaryotes in general.

Selected references

Skruzny, M., Brach, T., Ciuffa, R., Rybina, S., Wachsmuth, M. & Kaksonen, M. (2012). Molecular basis for coupling the plasma membrane to the actin cytoskeleton during clathrin-mediated endocytosis. *Proc. Natl. Acad. Sci. USA*, 109, E2533-42

Kukulski, W., Schorb, M., Kaksonen, M. & Briggs, J.A. (2012). Plasma membrane reshaping during endocytosis is revealed by time-resolved electron tomography. *Cell*, 150, 508-20

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, 328-37



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

Cytoskeletal dynamics and function in oocytes

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.

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.

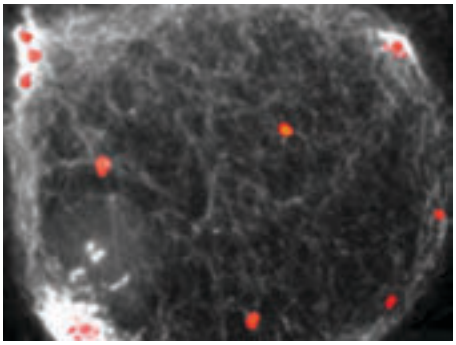
Selected references

Field, C.M. & Lénárt, P. (2011). Bulk cytoplasmic actin and its functions in meiosis and mitosis. *Curr. Biol.*, 21, R825-30

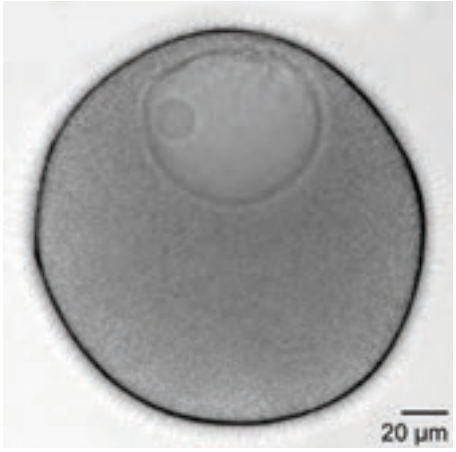
Mori, M., Monnier, N., Daigle, N., Bathe, M., Ellenberg, J. & Lénárt, P. (2011). Intracellular transport by an anchored homogeneously contracting F-actin meshwork. *Curr. Biol.*, 21, 606-11

Lénárt, P., *et al*. (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., *et al.* (2005). A contractile nuclear actin network drives chromosome congression in oocytes. *Nature*, 436, 812-8



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



Transparent starfish oocytes are uniquely suited for imaging meiotic divisions

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 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 mechanisms 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.

Cellular architecture

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.

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

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 have developed an advanced simulation engine – called cytosim – to simulate ensembles of multiple polar fibres 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.

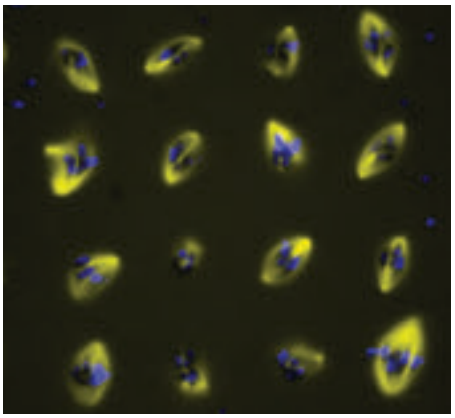
Selected references

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

Dinarina, A., *et al*. (2009). Chromatin shapes the mitotic spindle. *Cell*, 138, 502-13

Jekely, G., *et al*. (2008). Mechanism of phototaxis in marine zooplankton. *Nature*, 456, 395-9

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)*

Systems biology of stem cell differentiation

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

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.

Selected reference

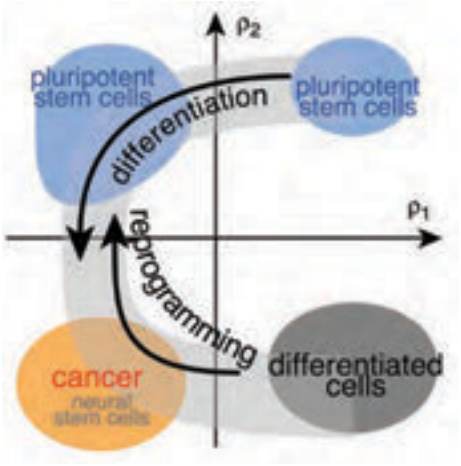
Neveu, P., *et al*. (2010). MicroRNA profiling reveals two distinct p53-related human pluripotent stem cell states. *Cell Stem Cell*, 7, 671-81

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 signalling 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 that 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 embryonic stem cell (ESC) self-renewal using single-cell live imaging.



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

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?

Membrane traffic in the early secretory pathway

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. Senior scientist since 2012.

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

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

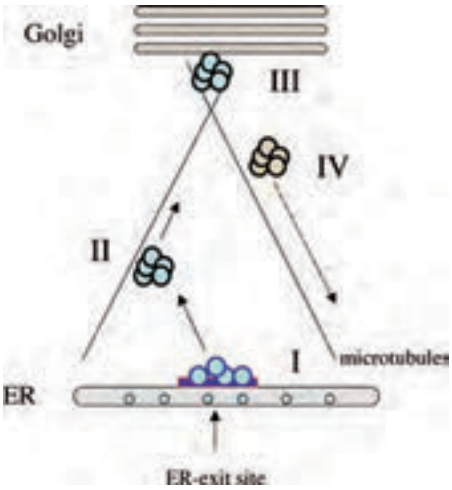
Selected references

Simpson, J.C., *et al*. (2012). Genome-wide RNAi screening identifies human proteins with a regulatory function in the early secretory pathway. *Nat. Cell Biol*., 14, 764-74

Tangemo, C., *et al*. (2011). A novel laser nanosurgery approach supports *de novo* Golgi biogenesis in mammalian cells. *J. Cell Sci.*, 124, 978-87

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

Forster, R., *et al*. (2006). Secretory cargo regulates the turnover of COPII subunits at single ER exit sites. *Curr. Biol.*, 16, 173-9



The four steps involved in ER to Golgi transport in mammalian cells. (I): Biogenesis of COPII coated vesicles occurs at specialised ER exit sites. (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

Cellular nanoscopy

The Ries group develops cutting-edge superresolution microscopy methods, such as automated localisation microscopy for proteome-wide superresolution imaging.

Jonas Ries

PhD 2008, TU Dresden.
Postdoctoral research at the ETH Zurich.
Group leader at EMBL Heidelberg since 2012.

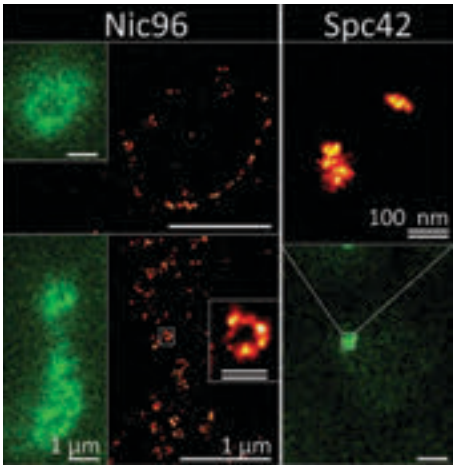
Selected references

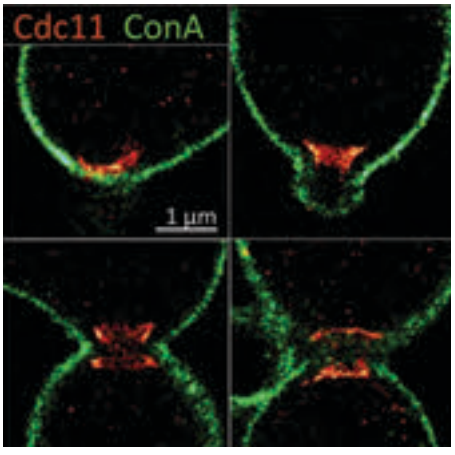
Ries, J., Kaplan, C., Platonova, E., Eghlidi, H. & Ewers, H. (2012). A simple, versatile method for GFP-based super-resolution microscopy via nanobodies. *Nat. Methods*, 9, 582-4

Schoen, I., Ries, J., Klotzsch, E., Ewers, H. & Vogel, V. (2011) Binding-Activated Localization Microscopy of DNA Structures. *Nano Letters*, 11, 4008-11

Ries, J., Yu, S. R., Burkhardt, M., Brand, M. & Schwille, P. (2009) Modular scanning FCS quantifies receptor-ligand interactions in living multicellular organisms. *Nat. Methods*, 6, 643-5

Ries, J., Ruckstuhl, T., Verdes, D. & Schwille, P. (2008). Supercritical angle fluorescence correlation spectroscopy. *Biophys. J.*, 94, 221-9





Previous and current research

The resolution of optical microscopy is limited by diffraction to about 200 nm, which is much larger than the relevant length-scales in cell biology, defined for instance by the size of organelles or supramolecular complexes. Single molecule localisation-based superresolution microscopy methods, such as photoactivated localisation microscopy (PALM), rely on the stochastic activation and subsequent localisation of individual fluorophores. They reach a 10-fold higher resolution, which is optimal for the study of intracellular structures. Until now however, these techniques required special fluorescent proteins to be cloned or high-affinity antibodies to be generated for specific labelling. On the other hand, many laboratories have most of their constructs in green fluorescent protein (GFP) form and entire genomes are available as functional GFP-fusion proteins. We recently developed a labelling scheme to make all these constructs immediately available for superresolution microscopy by targeting them with small antibodies labelled with bright organic dyes. This opens the door to high-throughput localisation analysis of entire genomes at the nanoscopic level in cells.

Our current research efforts are threefold. First, collaborating closely with other groups, we are establishing state-of-the-art superresolution microscopy to answer exciting questions in cell biology, which have only now become accessible due to greatly improved resolution. Second, we are working on automating single-molecule localisation microscopy with the aim of proteome-wide imaging – such superresolution localisation maps of proteins will be an invaluable resource for the life science community. Third, we are developing novel detection schemes for localisation microscopy. In one project we intend to use the principle of surface-generated fluorescence to improve the axial resolution. In another we aim to measure the 3D orientation of single molecules, in addition to their position by polarised detection, in order to resolve the structure of multi-molecular complexes.

Future projects and goals

Our goal is to establish cutting-edge superresolution microscopy and to apply it to biological systems. We will implement the newly developed detection schemes in a powerful microscope and combine this with advanced data analysis and our expertise in sample preparation and labelling. Using automated single-molecule localisation microscopy we are planning to image the whole proteome of budding yeast with a resolution of ~20 nm in dual-color and 3D. The combination of optical superresolution microscopy with dynamic microscopy techniques such as fluorescence correlation spectroscopy (FCS) or single particle tracking (SPT) bears great potential in relating structure, localisation and function, as does the combination with electron microscopy to add molecular specificity to the ultra structure.

Top: Localisation microscopy of proteins from a GFP-fusion construct library in budding yeast. Nic96: component of nuclear pore complex (NPC), individual NPCs are visible as ring-like structures. Spc42: part of the spindle pole body. Diffraction limited images (green) and reconstructed superresolution images (red) are shown. Bottom: Dual-color superresolution images of Cdc11 (red) and the cell-wall marker ConA (green) show the formation and disassembly of the Cdc11 ring

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), and lately peptides (Cobos-Correa *et al*., 2012).

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 develop 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*., 2012). At the cell biology level, our interest shifted to basic signalling networks regulated by G-protein-coupled and 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 60), 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 (see Plass *et al*., 2011, 2012 & Borrmann *et al*., 2012). In collaboration with the Häring group (page 20), 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; 2012).

Future projects and goals

In 2013, 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 26). 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

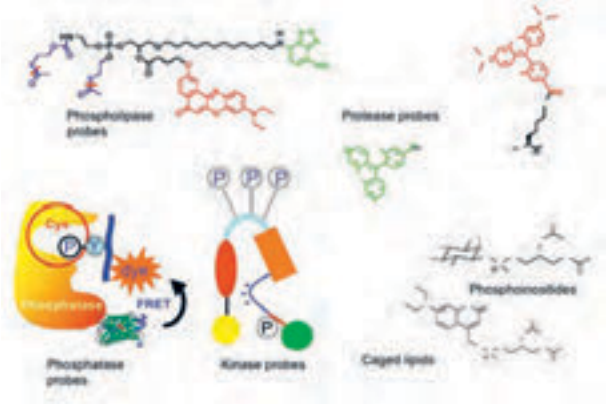
Cobos-Correa, A., Stein, F., & Schultz, C. (2012). Target-activated prodrugs for the auto-regulated inhibition of MMP12. *ACS Med. Chem. Lett.,* 3, 653–7.

Plass, T., *et al*. (2012). Amino acids for Diels-Alder reactions in living cells. *Angew. Chem. Int. Ed*., 51, 4166-70.

Gehrig, S., Mall M. A., & Schultz, C. (2012). Spatially resolved monitoring of neutrophil elastase activity with ratiometric fluorescent reporters. *Angew. Chem. Int. Ed.,* 51, 6258-61.

Rutkowska, A., & Schultz, C. (2012). Protein tango: the toolbox to capture interacting partners. *Angew. Chem. Int. Ed*., 51, 8166–76.

Several reporter and modulator molecules have been 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; and lipid derivatives that can be fluorescently labelled in living cells



Volume correlative light and electron microscopy

The Schwab team is interested in developing tools for the 3D correlation of data generated by fluorescent imaging and by electron microscopy.

Yannick Schwab

PhD 2001, Louis Pasteur University, Strasbourg.

Postdoctoral research, University of Calgary, Canada and IGBMC, Illkirch, France.
Head of Electron Microscopy at the Imaging Center, IGBMC, Illkirch, France.

Facility head and team leader at EMBL since 2012.

Selected references

Kolotuev, I., Hyenne, V., Schwab, Y., Rodriguez, D. & Labouesse, M. (2013). A pathway for unicellular tube extension depending on the lymphatic vessel determinant Prox1 and on osmoregulation. *Nat. Cell Biol*., e-pub ahead of print

Gehart, H., *et al*. (2012). The BAR Domain Protein Arfaptin-1 controls secretory granule biogenesis at the trans-golgi network. *Dev. Cell*, 23, 756-68

Kolotuev, I., Schwab, Y. & Labouesse, M. (2010). A precise and rapid mapping protocol for correlative light and electron microscopy of small invertebrate organisms. *Biol. Cell*, 102, 121-32

Spiegelhalter, C., *et al*. (2010). From dynamic live cell imaging to 3D ultrastructure: novel integrated methods for high pressure freezing and correlative light-electron microscopy. *PLoS One*, 5, e9014

Previous and current research

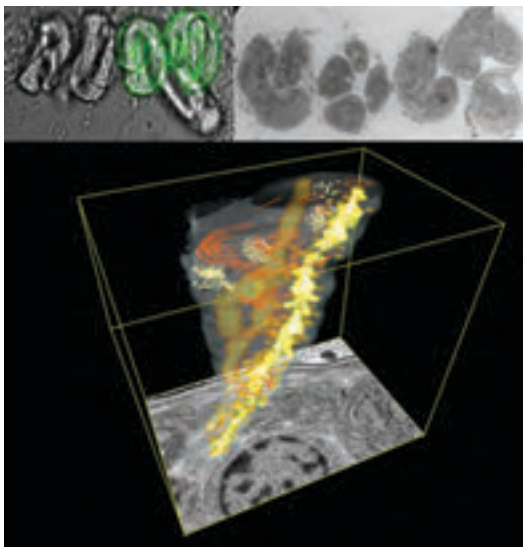
Correlative light and electron microscopy (CLEM) is a set of techniques that allow data acquisition with both imaging modalities on a single object. It is a growing field that now includes a large variety of strategies, and one that reaches a high degree of precision even in complex biological models. Before joining EMBL, we were developing tools and protocols to track rare objects or dynamic phenomena on cultured cells and bulk specimen such as nematodes and murine tissues.

One common challenge when trying to combine imaging modalities on the same sample is to identify space cues (external or internal) to track single objects through the changes in imaging modalities. On adherent cultured cells, we have developed specific substrates with coordinates to precisely record the position of cells (Spiegelhalter *et al.*, 2009; Gibbings *et al.,* 2011). To make the correlation easy, one key feature of the technique is the preservation of the coordinates in the EM sample. The 3D mapping of the cell of interest by confocal microscopy is also used to focus the 3D analysis to a given subset of serial sections.

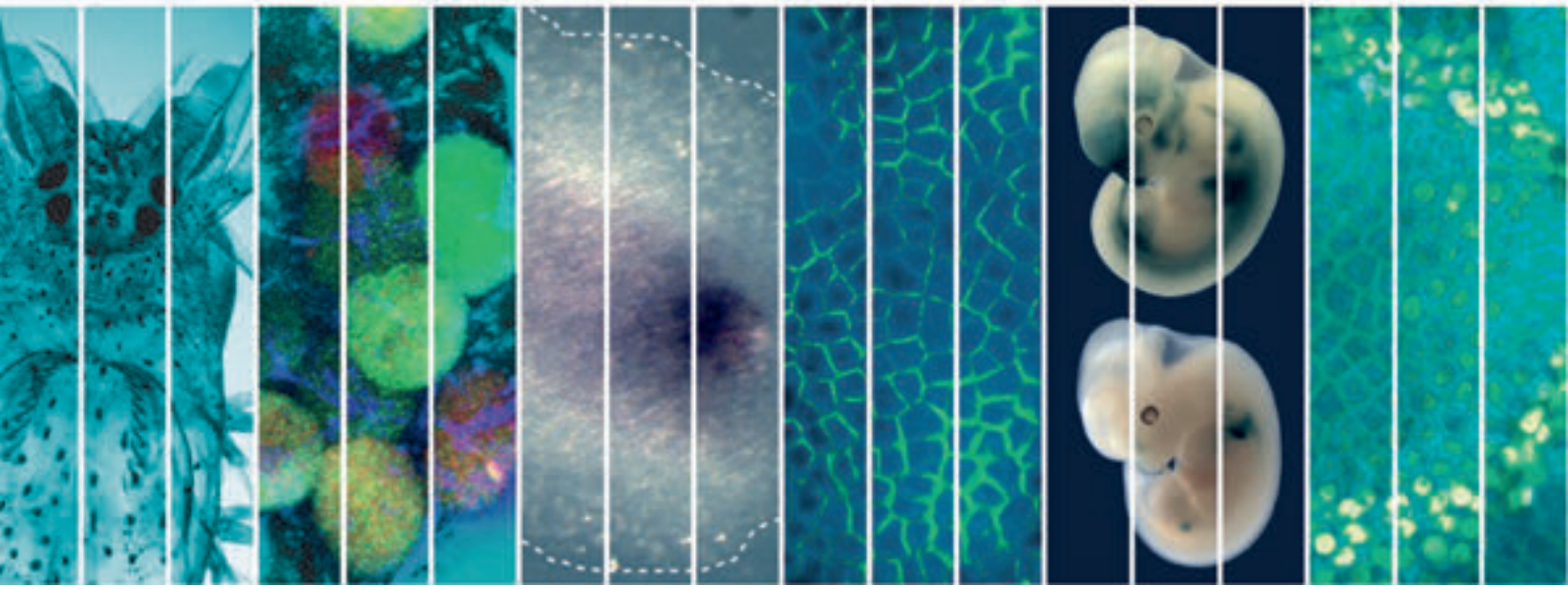
On more complex specimen, such as multicellular organisms, this targeting is even more critical as systematic EM acquisition of their entire volume would otherwise involve tedious and extremely long processes. We took advantage of the 3D mapping by confocal microscopy to characterise the position of an object of interest (cell or organ) that was then tracked at the EM level by targeted ultramicrotomy (Kolotuev *et al* 2009; 2012). With sub-micrometric precision, this technique has greatly improved the yield of data targeted to a given region of interest.

Future projects and goals

In parallel to the fast evolution of CLEM techniques over the past decade, acquisition methods in electron microscopes have significantly evolved with special breakthroughs in the volume analysis of cells by transmission electron microscopy (TEM) and scanning electron microscopy (SEM) tomography. Our team, in collaboration with other research teams at EMBL, will now combine these advanced techniques to perform CLEM in the 3D space of complex model specimen for cell and developmental biology. By automating the targeting and acquisition, we also intend to improve the throughput of data collection.



CLEM techniques, and especially targeted ultramicrotomy, allow the fine selection of objects of interest by light microscopy and a detailed 3D ultrastructural analysis by electron microscopy (taken from Kolotuev et al.*,2013)*



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 – 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 aims 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, another goal 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.

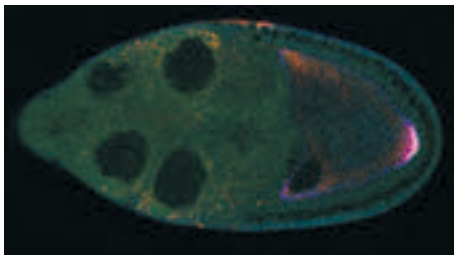
Selected references

Ghosh, S., Marchand, V., Gaspar, I. & Ephrussi, A. (2012). Control of RNP motility and localization by a splicing-dependent structure in *oskar* mRNA. *Nat. Struct. Mol. Biol.*, 19, 441-9

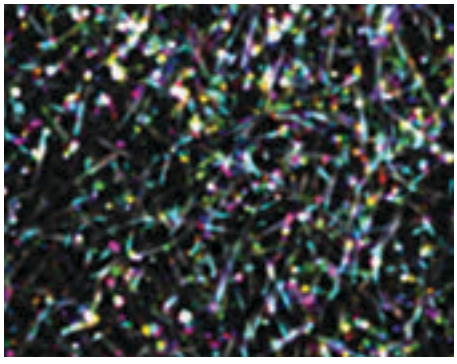
Jambor, H., Brunel, C. & Ephrussi, A. (2011). Dimerization of *oskar* 3' UTRs promotes hitchhiking for RNA localization in the *Drosophila* oocyte. *RNA*, 17, 2049-57

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 Advanced 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., *et al*. (2010). Ancient animal microRNAs and the evolution of tissue identity. *Nature*, 463, 1084-8

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., *et al*. (2008). Mechanism of phototaxis in marine zooplankton. *Nature*, 456, 395-9

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, which lived some 600 million years ago in the ocean – the last common ancestor of humans, flies and most other 'higher' animals that live today.

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. A 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 began 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.

Group leader at EMBL since 2009.

The Aulehla group studies 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, spatially but also in respect to 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 are embryonic oscillators/clocks employed during patterning? What are the dynamics of signalling pathways?

To approach these questions, novel methodologies are required (see video 1). We are generating novel real-time reporter mouse lines using knock-in technology that enables visualisation and quantification of temporal dynamics at different levels in the context of mouse embryonic development. Using *in vivo* imaging, we are focusing on the somite segmentation clock, an oscillatory system that is thought to control the formation of the pre-vertebrae that form periodically in a head-to tail sequence within the paraxial mesoderm. In mouse embryos this clock, with a periodicity of around two hours, drives oscillatory activity of several signalling pathways (Wnt, Notch and Fgf signalling) in the developing mesoderm. We recently developed an *ex vivo* assay that, in combination with real-time imaging reporters, has become instrumental for our approach: the assay recapitulates mesoderm patterning, including segment formation and spatio-temporally controlled oscillatory signalling activities, within the simplified context of a monolayer of primary mesoderm cells put in culture (figure, video 2).

Scaling and phase-shifted oscillators: One fundamental property of vertebrate segment formation is its ability to maintain proportions even when overall embryo size is experimentally altered, a process termed scaling. Intriguingly, scaling behaviour can be observed in the *ex vivo* assay system as well. This enabled us to identify a novel scaling mechanism employing phase-shifted oscillatory activity (Lauschke *et al*, 2013). How temporal devices, or oscillators, mechanistically encode spatial information for patterning constitutes a major interest in the lab. We are particularly interested in studying these problems at an integrated, higher-order level, so as to reveal emergent properties, incorporating mathematical modelling into our approach.

Other research in our group includes a specific focus on oscillatory Wnt-signalling. This signalling pathway serves a multitude of evolutionarily conserved functions during development and has been shown to play an essential role during somite formation. Our novel real-time reporter system is designed to reflect oscillatory Wnt-signalling activity both at gene activity and at protein levels. This will enable us to determine how the striking oscillations of Wnt- signalling activity are generated in the first place and, moreover, to functionally test their role in embryonic patterning. We are particularly interested in identifying the intrinsic and extrinsic factors that are responsible for controlling these oscillations within the segmentation process.

Future projects and goals

- Quantitative (imaging) approach to understand the role of dynamic oscillatory signalling during patterning and scaling.
- Study of emergent properties of coupled oscillator populations.
- Discovery of oscillatory signalling phenomena during embryogenesis.

Selected references

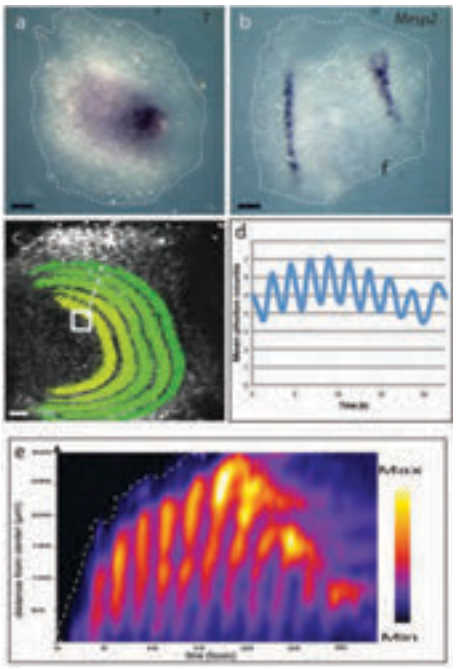
Lauschke, V.M., Tsiairis, C.D., Francois, P. & Aulehla, A. (2013). Scaling of embryonic patterning based on phase-gradient encoding. *Nature*, 493, 101-5

Aulehla, A., *et al*. (2008). A beta-catenin gradient links the clock and wavefront systems in mouse embryo segmentation. *Nat. Cell Biol*., 10, 186-93

Aulehla, A., *et al*. (2003). Wnt3a plays a major role in the segmentation clock controlling somitogenesis. *Dev. Cell*, 4, 395-406

Video links

www.embl.de/research/units/dev_biology/aulehla/



Ex vivo cell culture model for mesoderm patterning and oscillations. a, b) Primary mesoderm cells retain undifferentiated PSM identity in the centre of the culture, before initiating a differentiation program in the periphery. c) Snapshot of time-series using fluorescent lunatic fringe reporter mouse line (LuVeLu), overlaid with time-projection of activity patterns (in green) seen during the time-lapse recordings (see video 2). d) Raw photon counts (measured in quadrant shown in c) demonstrating robust oscillatory activity for extended recording times). e) Time-space kymograph along dashed arrow in c. From this quantification, critical oscillation parameters, wave speed and phase-distributions can be calculated

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., *et al*. (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, 257-64

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, 124-33

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 a model system during cellularisation and early gastrulation stages.

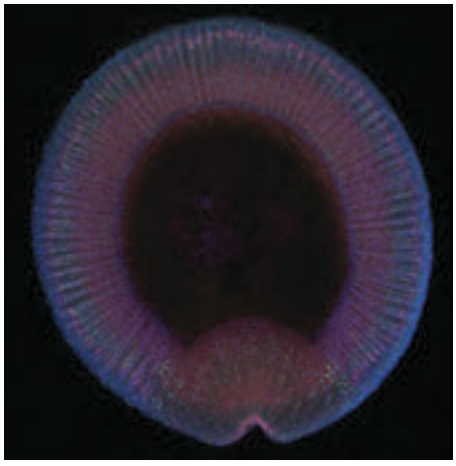
Cellularisation of the *Drosophila* embryo provides an excellent system to study mechanisms linking membrane trafficking and cell/tissue morphogenesis. It takes around an hour for a syncytium of ~6000 nuclei to complete the process of cellularisation, a particular form of cytokinesis involving a massive 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.

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

Group leader at EMBL since 2009.
ERC Investigator.

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

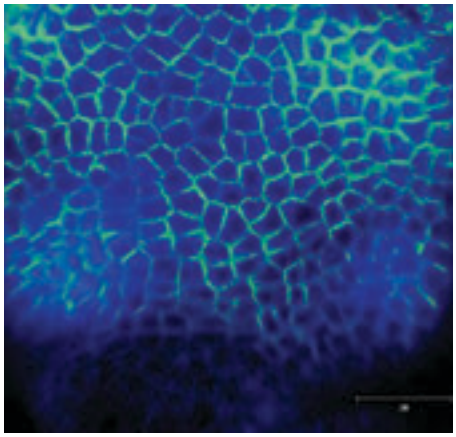
Heisler, M.G., *et al*. (2010). Alignment between PIN1 polarity and microtubule orientation in the shoot apical meristem reveals a tight coupling between morphogenesis and auxin transport. *PLoS Biol*., 8, e1000516

Hamant, O., *et al*. (2008). Developmental patterning by mechanical signals in *Arabidopsis*. *Science*, 322, 1650-5

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, 1633-8

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.

Selected references

Courtois, A., Schuh, M., Ellenberg, J. & Hiiragi, T. (2012). The transition from meiotic to mitotic spindle assembly is gradual during early mammalian development. *J. Cell Biol*., 198, 357-70

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

Motosugi, N., Bauer, T., Polanski, Z., Solter, D. & Hiiragi, T. (2005). Polarity of the mouse embryo is established at blastocyst and is not prepatterned. *Genes Dev.*, 19, 1081-92

Hiiragi, T. & 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

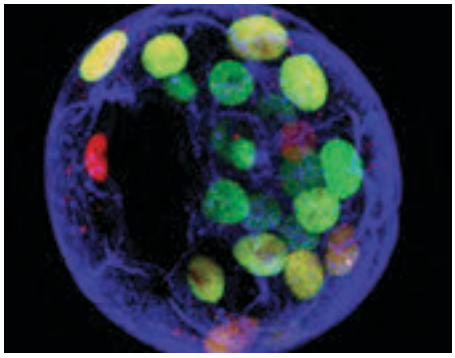


Figure 1: Unprecedented molecular heterogeneity during mouse blastocyst patterning. Cells expressing Nanog (green), Cdx2 (red) or both (yellow)

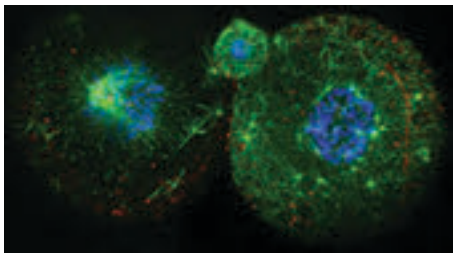


Figure 2: Microtubules (green) self-assemble into a mitotic spindle from multiple microtubule-organising centres (red), in the absence of centriole, during progressive transition from meiosis to mitosis in early mouse development

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, Münster, 2007-2011.

Group leader at EMBL since 2011.
ERC Investigator.

Previous and current research

A fundamental question in biology is the mechanism by which embryonic asymmetry is established during development. In contrast to many organisms in which embryonic development is driven by determinants localised asymmetrically in the egg, mammalian eggs lack polarity and thus symmetry has to be broken during early embryogenesis. This symmetry-breaking process in mammalian embryos results in the formation of the blastocyst, composed of the inner cell mass surrounded by the trophectoderm. Despite its importance, 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, stochasticity and molecular heterogeneity (figure 1) during early embryogenesis. In view of this and its highly regulative capacity, the early mammalian embryo may be a self-organising system patterning through stochastic processes in a particular structural context. 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 to investigate this: fluorescence-based gene-trap mice that allow quantitative characterisation of molecular dynamics; transcriptomics of every single cell in the embryo; computer simulation of the blastocyst morphogenesis.

In the early mouse embryo the spindle is self-assembled from randomly distributed microtubule-organising centres in the absence of centriole. Our recent study demonstrated a surprising gradual transition from meiosis to mitosis over the pre-implantation stage (figure 2), during which centriole is generated *de novo* without template. We use this unique and effective system to study the mechanism of centriole biogenesis *in vivo*.

Overall we aim at understanding principles and robustness underlying early mammalian development.

Future projects and goals

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

- identification of the symmetry breaking cue in the mouse embryo
- understanding the molecular mechanisms leading to the first lineage specification
- investigating the role of molecular heterogeneity in lineage segregation
- identification of the essential molecule and trigger for centriole biogenesis.

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.
ERC Investigator.

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 damages 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

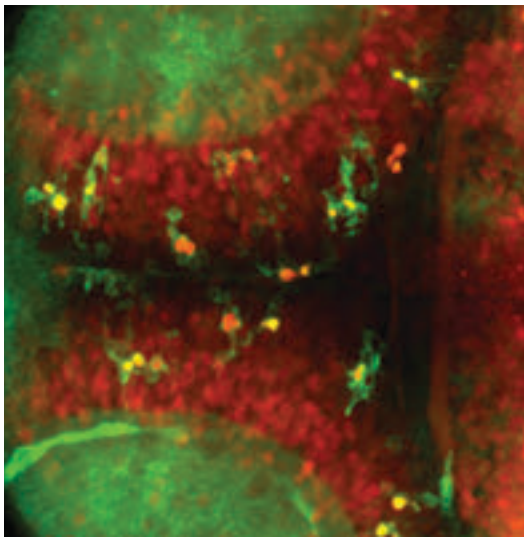
Sieger, D. & Peri, F. (2013). Animal models for studying microglia: The first, the popular, and the new. *Glia*, 61, 3-9

Sieger, D., Moritz, C., Ziegenhals, T., Prykhozhij, S. & Peri, F. (2012). Long-range Ca(2+) waves transmit brain-damage signals to microglia. *Dev. Cell*, 6, 1138-48

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

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 how the intricate distribution of regulatory elements along the genome is transformed into specific gene expression profiles.

François Spitz

PhD 1997, Institut Cochin de Génétique Moléculaire, Paris.

Postdoctoral research at the University of Geneva.

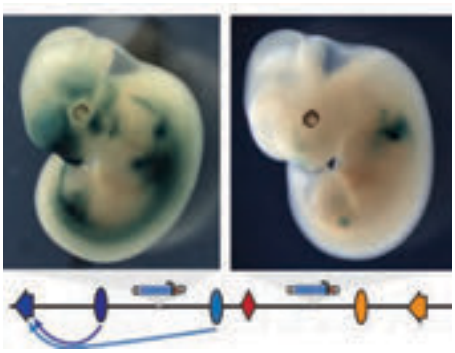
Group leader at EMBL since 2006.

Selected references

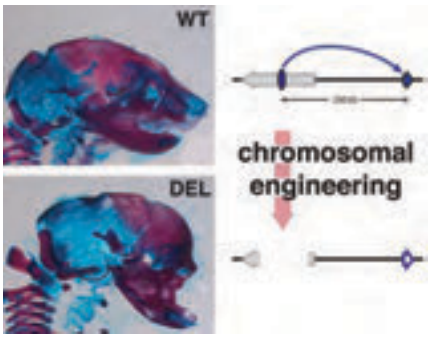
Marinic, M., Aktas, T., Ruf, S. & Spitz, F. (2013). An Integrated Holo-Enhancer Unit Defines Tissue and Gene Specificity of the Fgf8 Regulatory Landscape. *Developmental Cell*, e-pub ahead of print

Montavon, T., *et al*. (2011). A regulatory archipelago controls hox genes transcription in digits. *Cell*, 147, 1132-45

Ruf, S., *et al*. (2011). Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nat. Genet.*, 43, 379-86



Adjacent insertions of a sensor gene showed different activities, highlighting the regulatory architecture of the corresponding locus (see Ruf et al*, 2011)*



Abnormal skull development in mice with a deletion of distal enhancers engineered by in vivo *recombination (see Marinic,* et al. *2013)*

Previous and current research

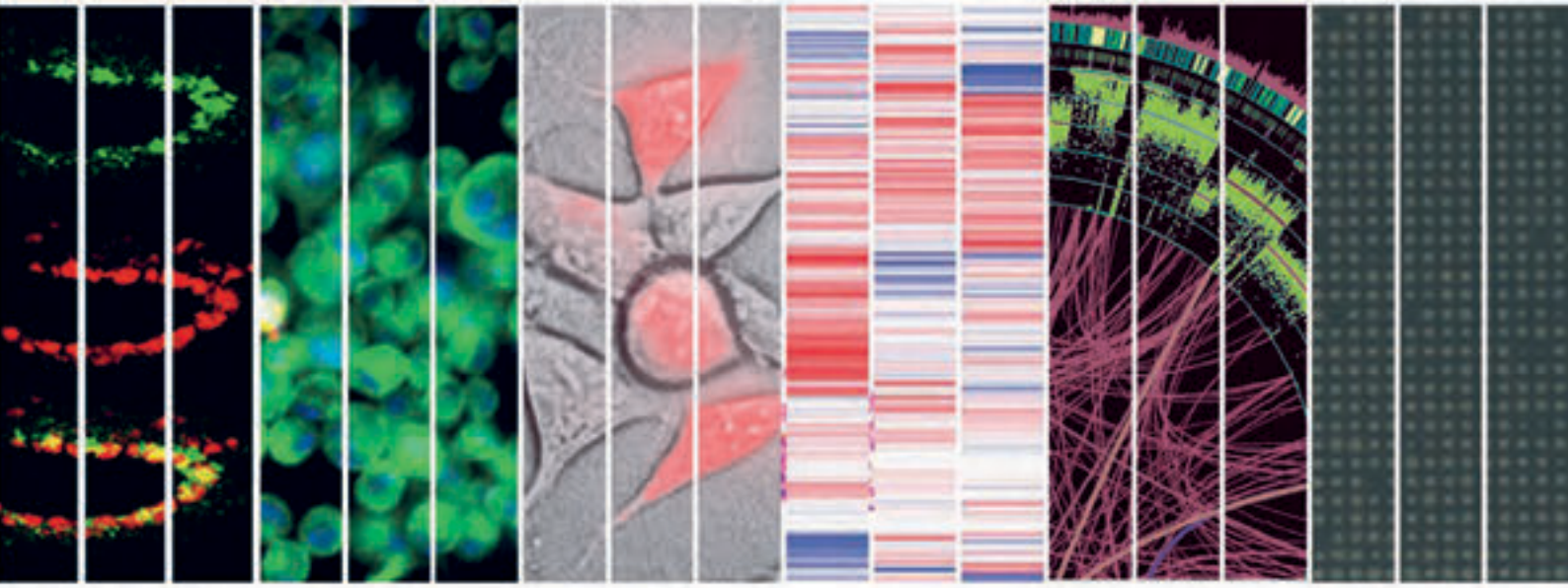
The patterning of the embryo and the specification of its different cell types are driven by the implementation of cell-specific gene expression programs. In vertebrates, the *cis*-acting elements that regulate transcription can be located hundreds of kilobases away from the genes they control, particularly for genes with important functions during development. Because of this, the genome appears to be composed of intermingled arrays of unrelated genes and *cis*-regulatory elements. Therefore, the mechanisms that regulate enhancer-promoter interactions are essential to transform this apparent genomic and regulatory conundrum into gene- and tissue- specific expression programs. Recent data reveal that genomic loci adopt specific chromatin structures and conformations in the nuclei of different cell types, correlating with differential gene activity. Yet, the *cis*-acting genomic elements that determine how a genomic locus folds into specific structural and regulatory architectures, and the precise roles of the chromatin, protein-complexes and non-coding RNAs suggested to contribute to this process are still unclear.

Our lab has developed several experimental approaches to explore the regulatory architecture of the mouse genome and characterise functionally the mechanisms that organise it. Towards this aim, we have established an efficient *in vivo* system that, through the combined used of transposases and recombinases, allows the re-engineering, in a systematic manner, of the mouse genome. With this approach, we generated a unique resource comprising hundreds of mouse strains carrying regulatory sensors throughout the genome, and series of specific chromosomal rearrangements in selected loci. This genomic resource enables us to dissect functionally the genomic information and the mechanisms that organise a linear genome into structurally distinct domains and chromatin loops, so as to implement long-range specific regulatory interactions.

Future projects and goals

Structural conformation and regulatory organisation of the genome: By combining advanced genomic engineering with chromatin profiling (ChIP-Seq) and conformation analyses (4C-chromatin conformation capture, super-high resolution FISH), we aim to learn how the genomic organisation of a locus determines the specific chromatin structures and conformations that it adopts in the nucleus, and determines their functional significance in the context of a developing embryo.

Regulatory architecture, disease and evolution: Our mouse models provide insights into the consequences of structural variations or chromosomal aneuploidies found in humans, both at the phenotypic and molecular level. Comparison of the regulatory architecture of developmental gene loci between different species can reveal how large-scale changes in chromosomal organisation may 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 by 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.
ERC Advanced Investigator since 2013.

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 processes 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 only using 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

Wilczynski, B., Liu, Y.H., Yeo, Z.X. & Furlong, E.E. (2012). Predicting spatial and temporal gene expression using an integrative model of transcription factor occupancy and chromatin state. *PLoS Comput. Biol.*, 8, e1002798

Spitz, F. & Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.*, 13, 613-26

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*, 148, 473-86

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*, 44, 148-56

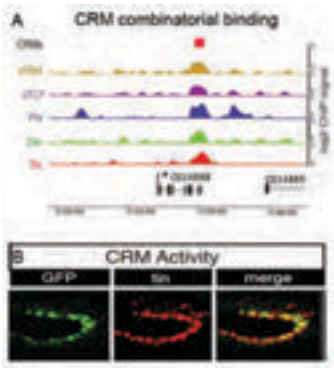


Figure 1: Enhancers can function by highly cooperative transcription factor occupancy using very flexible motif content and organisation (Junion, Spivakov, et al*, 2012)*

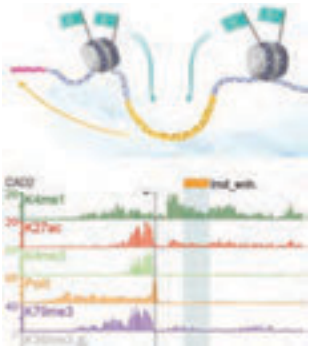


Figure 2: Chromatin state and RNA polymerase II occupancy on enhancers (yellow) is highly predictive of enhancers' activity and is very dynamic, mirroring that of dynamic enhancer usage during development (Bonn, Zinzen, Girardot, et al*, 2012)*

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.

Lars Steinmetz

PhD 1997-2001, Stanford University.

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

Group leader at EMBL since 2003.
Joint Head of Genome Biology Unit and Senior Scientist since 2009.
ERC Advanced Investigator since 2012.

Selected references

Tan-Wong, S.M., *et al*. (2012). Gene loops enhance transcriptional directionality. *Science*, 338, 671-5

Blandin, S.A., *et al*. (2009). Dissecting the genetic basis of resistance to malaria parasites in *Anopheles gambiae*. *Science*, 326, 147-50

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

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

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.

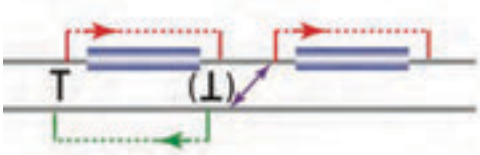


Figure 1: Antisense transcription enables local dispersion of regulatory signals via bidirectional promoters (Xu et al.*, 2011,* Molecular Systems Biology*)*

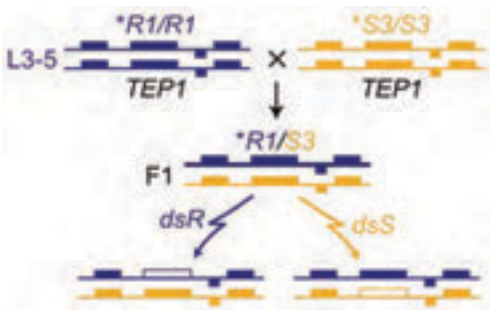


Figure 2: Reciprocal allele-specific RNAi identifies the mosquito gene TEP1 as a major contributor to resistance to malaria parasites (Blandin et al.*, 2009,* Science*)*

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 research at IBM Research, San Jose, California and at DKFZ Heidelberg.

Group leader at EMBL-EBI 2004-2009, 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, and 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.

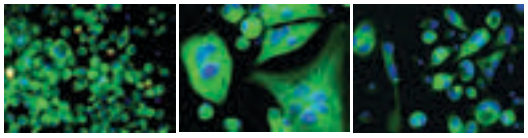
Selected references

Anders, S., Reyes, A. & Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. *Genome Res.*, 22, 2008-17

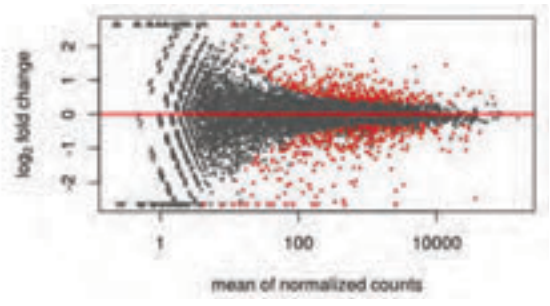
Horn, T., *et al*. (2011). Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. *Nat. Methods*, 8, 341-6

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

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



Large-scale mapping of genetic interactions by combinatorial RNAi, automated image analysis and computational phenotyping. The double knockdown of Rho1 and Dlic (right) shows a phenotype that is different from what is expected from the single-gene knockdowns of Dlic (left) and Rho1 (middle)



Detection of differentially expressed genes in RNA-Seq experiments using the DESeq method

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

Chatterjee, J., *et al*. (2012). Development of a peptide that selectively activates protein phosphatase-1 in living cells. *Angew Chem. Int. Ed. Engl.*, 51, 10054-9

Bru, M., Kotkar, S.P., Kar, N. & Köhn, M. (2012). Development of a solid phase synthesis strategy for soluble phosphoinositide analogues. *Chemical Science*, 3, 1893-902

Rios, P., Li, X. & Köhn, M. (2012). Molecular mechanisms of the PRL phosphatases. *FEBS J.*, 280, 505-24

McParland, V., *et al*. (2011). The metastasis-promoting phosphatase PRL-3 shows activity toward phosphoinositides. *Biochemistry*, 50, 7579-90

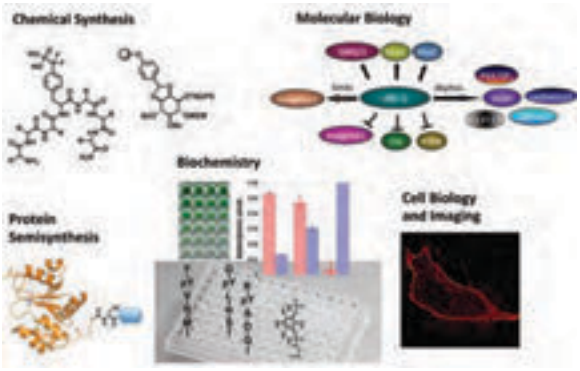
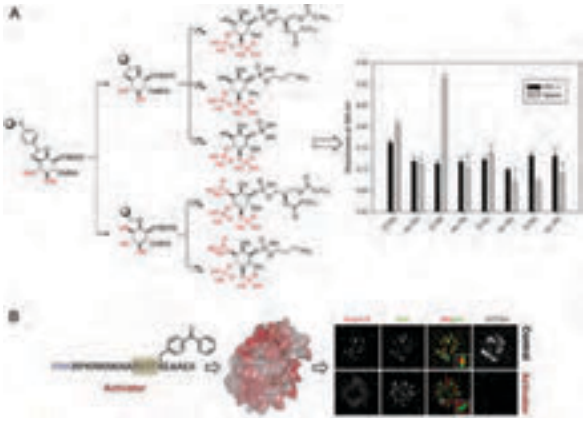


Figure 1 (above): Our group's approaches to the investigation of disease-promoting phosphatases



Previous and current research

Within intracellular signalling networks, phosphatases are counter players of kinases and play crucial roles in health and disease. The investigation of phosphatases is challenging, which is also due to the lack of tools to selectively study particular phosphatases. Understanding of phosphatase function, regulation and substrate interaction is therefore still quite limited. Our main interest is thus to control and investigate phosphatases using interdisciplinary approaches (figure 1). We are focusing on phosphatases that promote diseases.

Specifically, we are interested in the phosphatase of regenerating liver (PRL) enzymes. The PRL family consists of three oncogenes and is barely understood. My group studies biological pathways and the roles of this family using biochemical and molecular cell biology approaches. We aim to design inhibitors for PRL members. We have recently observed phosphoinositide-phosphatase activity *in vitro* for one member, PRL-3 (McParland *et al.*, 2011). In this regard, we developed a solid phase synthesis strategy that accelerates access to phosphoinositides and their analogues (figure 2a, Bru *et al.*, 2012). We aim to obtain a detailed picture of substrate specificities of lipid phosphatases in biochemical structure-activity relationship (SAR) studies using a library of phosphoinositide analogues. Information about specific substrate preferences will help to design specific inhibitors of lipid phosphatases.

Another interest is the tool development for protein phosphatase-1 (PP1), a ubiquitous phosphatase that is responsible for one third of all dephosphorylation reactions on Ser/Thr inside cells and is involved in many processes such as mitosis and cell cycle regulation. We have recently developed a peptide-based PP1 activator that targets protein–protein interactions (figure 2b, Chaterjee *et al.*, 2012). All our chemical modulators are developed to be reliably active inside cells.

The understanding of phosphatase and kinase networks is still incomplete. We use computational, biochemical and structural approaches to view, predict and validate these networks (collaboration with the Thornton (page 76) and Wilmanns (page 114) groups).

Future projects and goals

Studying PRL biology will remain a focus of our lab in the future, with the goal of understanding the underlying mechanisms of oncogenesis caused by these phosphatases. We continue to develop chemical methods to enable us to apply peptides as well as inositides as phosphatase modulators inside cells. In addition to the phosphoinositide-based SAR studies, we have started to employ small molecules and medicinal chemistry for this purpose. Designing modulators for the highly complex serine/threonine phosphatases is another goal, and we continue to develop these for PP1 and have started this for PP2C (collaboration with Márquez team, page 106). The lab consists of an equal number of molecular biologists and organic chemists at pre- and postdoctoral level. The combination of biology and chemistry not only opens up new ways to approach challenging phosphatase research, but also broadens views and skills of every lab member.

Figure 2 (left): Solid phase synthesis of phosphoinositides for the preparation of libraries to enable SAR studies with lipid phosphatases. (B) A selective activator of PP1 in cells (collaboration with the Hart (page 105) and Bollen (KU Leuven) groups)

Origin and function of genetic variation

Jan Korbel

PhD 2005, EMBL Heidelberg/Humboldt University, Berlin.

Postdoctoral research at Yale University, New Haven, Conneticut, USA.

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

Heritable structural variations (SVs) represent a major class of genetic variation in humans and are responsible for more heritable nucleotide sequence differences between individuals than other genetic variant classes, such as single-nucleotide polymorphisms. One of our main focuses has been on developing next-generation DNA sequencing-based approaches for the discovery and genotyping of SVs (see figure). In the 1000 Genomes Project we are contributing to the generation of a fine-resolution SV map across more than 2000 humans: early versions have provided insights into the mechanistic origins of SVs, enabling the delineation of SV formation hotspots in the human genome. Integrating genetic variation maps with cellular (such as transcriptome and protein-DNA binding) and clinical phenotypes will allow us to decipher the functional impact of genetic variants by uncovering genotype-phenotype relationships.

Our group also contributes to the International Cancer Genome Consortium: currently, we are investigating the genomes, epigenomes and transcriptomes of hundreds of patients diagnosed with early-onset prostate cancer, lymphoma, and paediatric brain tumours. Recently, we observed that childhood medulloblastoma patients harbour very few point mutations, raising the question of which molecular events lead to the development of these tumours. Using whole-genome sequencing, we found that medulloblastoma frequently develops in conjunction with a remarkable process termed 'chromothripsis', where localised chromosomal shattering and repair occur in a one-off structural variation catastrophe and leads to massive DNA rearrangement (see figure). Our recent studies have linked chromothripsis in medulloblastoma with germline mutations of the *TP53* gene, encoding the p53 protein. We are currently expanding our panel of completely sequenced cancer genomes to intensify the discovery of links between mutational mechanisms and genotype-phenotype relationships in cancer and are performing complementary molecular and cell biology experiments to identify factors instigating chromothripsis, with the aim of uncovering the mechanistic basis of this catastrophic rearrangement process.

Future projects and goals

- Constructing a near-complete map of human genome variation, including difficult-to-ascertain genomic regions, using novel DNA sequencing approaches.
- Development of wet-lab and *in silico* approaches for deciphering the molecular origin and function of SVs in humans and model systems (including yeast and non-human primates).
- Uncovering genetic determinants for the development and progression of cancer in humans.
- Deciphering the mechanistic basis of 'chromothripsis', a rearrangement process observed in 2-3% of human cancers, which is particularly abundant in some highly aggressive malignancies.
- Dissecting the effect of genetic variation on biological systems in healthy states and in cancer, by integrating 'variomes' with cellular phenotype (such as transcriptome and epigenome) data.

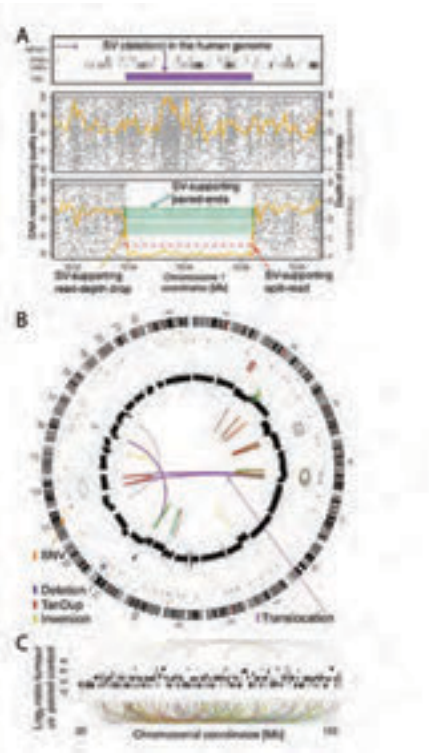
Selected references

Zichner, T., *et al*. (2012). Impact of genomic structural variation in *Drosophila* melanogaster based on population-scale sequencing. *Genome Res.*, epub ahead of print

1000 Genomes Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491, 56-65

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

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



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

Functional 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

Hughes, C., & Krijgsveld, J. (2012). Developments in quantitative mass spectrometry for the analysis of proteome dynamics. *Trends Biotechnol.*, 30, 668-76

Eichelbaum, K., Winter, M., Diaz, M.B., Herzig, S., & Krijgsveld, J. (2012). Selective enrichment of newly synthesized proteins for quantitative secretome analysis. *Nat. Biotechnol.*, 30, 984-90

Klimmeck, D., Hansson, J., Raffel, S., Vakhrushev, S.Y., Trumpp, A., & Krijgsveld, J. (2012). Proteomic cornerstones of hematopoietic stem cell differentiation: distinct signatures of multipotent progenitors and myeloid committed cells. *Mol. Cell Proteomics*, 11, 286-302

Hansson J., *et al*. (2012). Highly coordinated proteome dynamics during reprogramming of somatic cells to pluripotency. *Cell Rep.*, 2, 1579-92

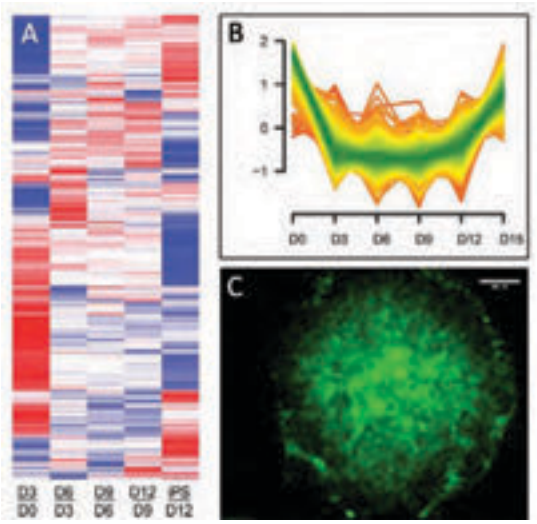
Previous and current research

Proteins fulfil most of the functions that are crucial in establishing cellular phenotypes. 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, or to develop from an embryonic to a mature state. Our interest is to understand cellular properties 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, mass spectrometry, analytical chemistry, and bioinformatics – applied to various biological systems (yeast, *Drosophila*, mammalian cells). Our main interest is to understand how changes in protein expression, localisation and interaction underlie processes in stress-response, differentiation and reprogramming. For instance, large-scale proteomic experiments have enabled us to characterise the proteomes of highly purified mouse hematopoietic stem cells and progenitor populations obtained by FACS sorting, generating novel insights in the initial steps of hematopoiesis *in vivo*. Furthermore, we have performed time course analyses quantifying the proteome changes in fibroblasts during their reprogramming to induced pluripotent stem cells (iPSCs), identifying and functionally validating proteins that are key in gaining pluripotency. Apart from these large-scale analyses of intracellular proteomes, we have developed new tools to study secretory proteins and their role in cell signalling and communication. Furthermore, we are interested in regulatory principles of transcriptional activation and protein turnover in the face of developmental processes or response to stress. We are therefore developing novel techniques to identify proteins that interact with regulatory domains in the genome, both *in vivo* and *in vitro*. In doing so, we aim to identify proteins that drive (or inhibit) transcription in a gene- and condition-specific manner, for instance to understand how transcription of developmentally important genes is controlled. To further explore the link between genome regulation and protein output, we study protein turnover, using yeast as a model system. By determining protein synthesis and degradation proteome-wide and across a range of growth conditions, we aim to construct models of how protein homeostasis is maintained.

Future projects and goals

- Develop new tools to study protein-DNA and protein-RNA interactions to identify and functionally characterise proteins that regulate transcription and translation.
- Integrate proteomics and next-generation sequencing to understand the molecular basis of protein homeostasis.
- Study cellular communication via secretory proteins.



Changes in protein expression during reprogramming of fibroblasts (A, B) leading to the formation of induced pluripotent stem cells (C)

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.

During the past 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) enable us to obtain high concentrations of nucleic acids (mRNA, DNA) or proteins (for instance secreted antibodies) from individually encapsulated cells, paving the way for single cell assays. We also 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 the analysis of 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 signalling and mitochondrial activity.

Biomedical applications: Droplet-based microfluidics enables antibody screening at very high throughput (~50 samples/sec). We want to use this technology to identify therapeutic antibodies, starting with primary plasma cell samples from human disease (such as HIV and HCV) survivors. In parallel, we exploit a reversed experimental setup to derive potential HIV vaccine candidates (in close collaboration with the International AIDS Vaccine Initiative).

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

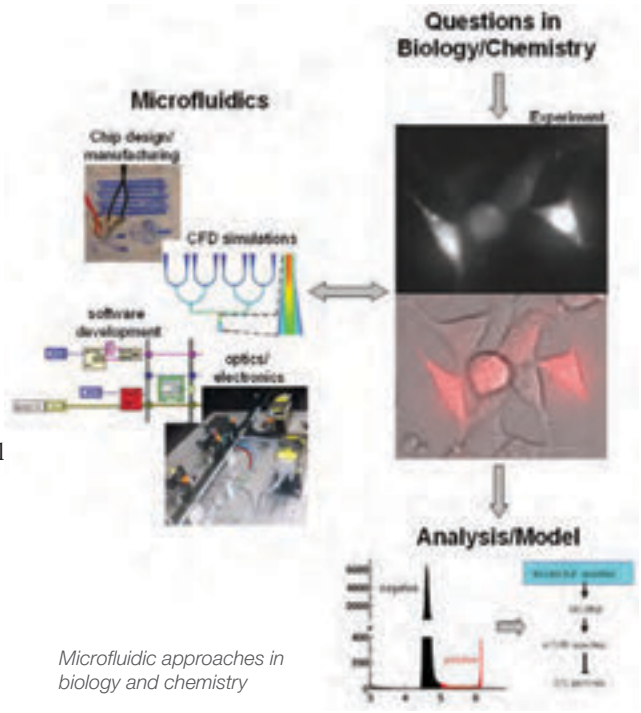
Selected references

El Debs, B., *et al*. (2012). Functional single-cell hybridoma screening using droplet-based microfluidics. *PNAS* 109, 11570-75. Research highlight in *Nature Methods* 9: 866

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.

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., *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, 580-1



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

PhD, 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

Typas, A., Banzhaf, M., Gross, C.A. & Vollmer, W. (2012). From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nat. Rev. Microbiol.*, 10, 123-36

Nichols, R.J., *et al*. (2011). Phenotypic landscape of a bacterial cell. *Cell*, 144, 143-56

Typas, A., *et al*. (2010). Regulation of peptidoglycan synthesis by outer-membrane proteins. *Cell*, 143, 1097-109

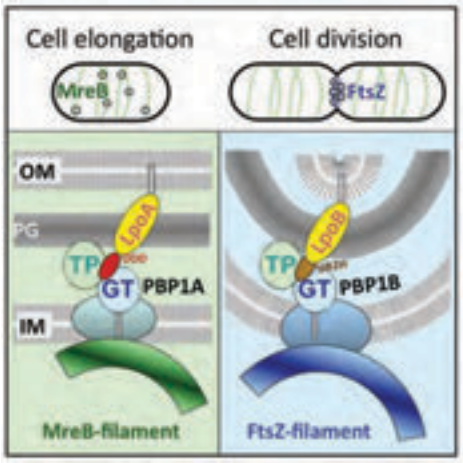
Typas, A., *et al*. (2008). High-throughput, quantitative analyses of genetic interactions in *E. coli*. *Nat. Methods*, 5, 781-7

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.

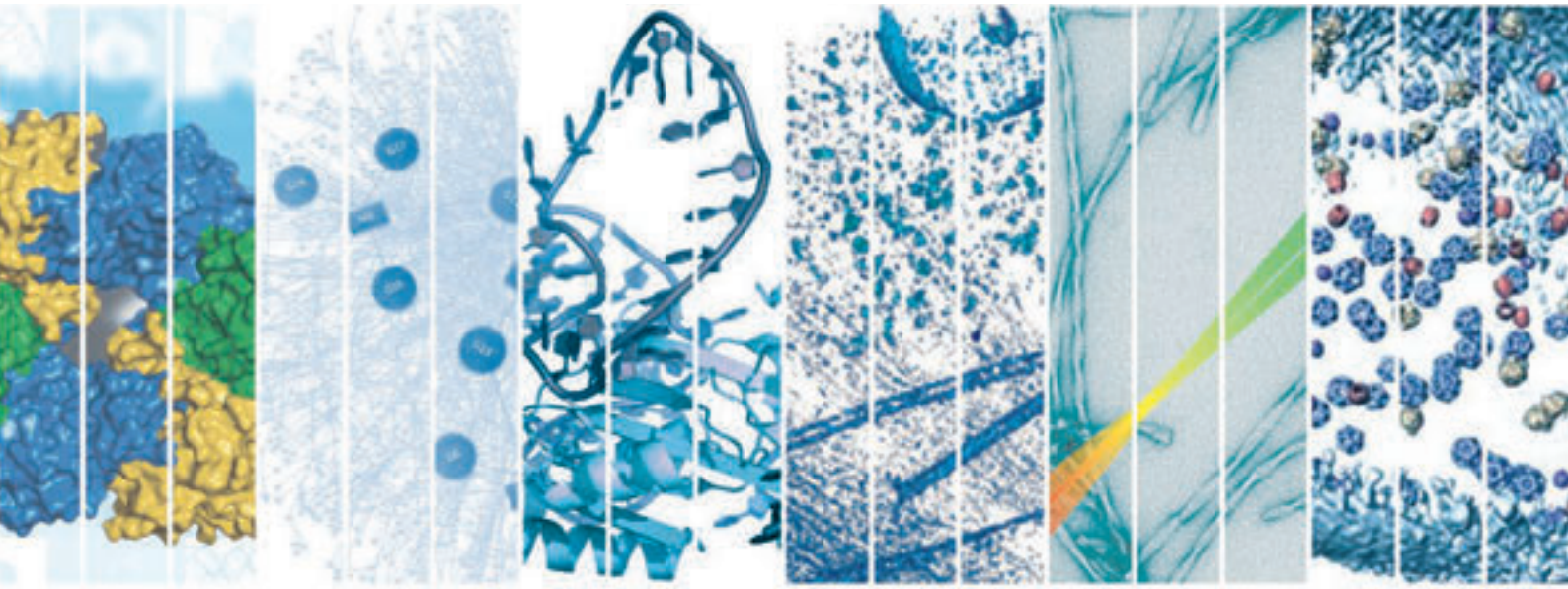
Working at the intersection between systems genomic biology and mechanistic molecular biology, we have discovered key missing players of major envelope components, uncovered niche-specific regulation of conserved envelope processes, identified linking proteins that allow coordination between processes 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.



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

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. This reflects our 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 crystallisation robot and automated crystal visualisation; rotating anode and image plate detector for the collection of X-ray diffraction data; 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 around 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.
ERC Advanced Investigator.

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. We also aim 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

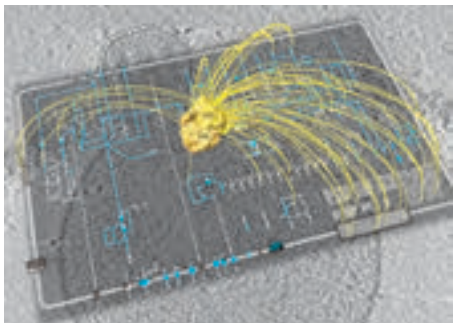
Arumugam, M., *et al.* (2011) Enterotypes of the human gut microbiome, *Nature*, 473, 174-80

Qin, J., *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., *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 connectors) and a metabolic reconstruction in which the correspondence to operon organisation is shown (blue)*



Molecular mechanisms of transcriptional regulation in eukaryotes

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. & Müller, C.W. (2012). The elongator subcomplex Elp456 is a hexameric RecA-like ATPase. *Nat. Struct. Mol. Biol*., 19, 314-20

Nowak, A.J., *et al*. (2011). Chromatin-modifying complex component Nurf55/p55 associates with histones H3 and H4 and polycomb repressive complex 2 subunit Su(z)12 through partially overlapping binding sites. *J. Biol. Chem.*, 286, 23388-96

Fernandez-Tornero, C., *et al*. (2010). Conformational flexibility of RNA polymerase III during transcriptional elongation. *EMBO J.*, 29, 3762-72

Moriniere, J., *et al*. (2009). Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature*, 461, 664-8

Previous and current research

In the context of chromatin, we are interested in how 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 overall structure, architecture and inner-workings of large molecular machines, such as RNA polymerases or chromatin modifying complexes, involved in transcription. We would also like to gain insight into how DNA sequence information and epigenetic modifications work together to regulate gene transcription.

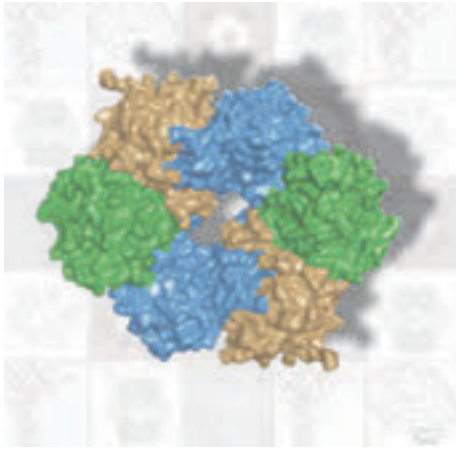
To achieve these goals, we use structural information obtained by X-ray crystallography and electron microscopy combined with other biophysical and biochemical approaches. Systems currently under investigation include transcription factor/DNA complexes, yeast RNA polymerase III, Elongator, and multi-protein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

RNA polymerase III transcription: RNA polymerase III (Pol III) consists of 17 subunits and is responsible for the transcription of small RNAs, such as 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). We aim to understand the overall architecture of the Pol III pre-initiation complex and the interaction between Pol III, TFIIIB and TFIIIC during Pol III recruitment, transcriptional elongation and termination.

Elongator: The 6-subunit Elongator complex was initially identified as a transcriptional regulator associated with elongating RNA polymerase II. However, recent results suggest that Elongator is involved in the specific modification of uridines at the wobble base position of tRNAs. Our group recently solved the Elp456 subcomplex that forms a ring-like heterohexameric structure resembling hexameric RecA-like ATPases. We are now pursuing the structural analysis of the entire Elongator complex to gain further insight into its function.

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.

Elongator subunits Elp4 (green), Elp5 (blue) and Elp6 (brown) form a ring-like structure resembling hexameric RecA-like ATPases that specifically binds tRNA (Glatt et al.*, 2012)*



Future projects and goals

- Gaining molecular insights into the recruitment of transcriptional regulators through DNA sequence-specific recognition and epigenetic modifications.
- Structural and functional analysis of macromolecular machines involved in transcription regulation and chromatin remodelling and modification.
- Contributing to a better mechanistic understanding of eukaryotic transcription and epigenetics using integrated structural biology combined with biochemical and cell biology approaches.

Mechanism of mobile DNA elements and their 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 group uses structural and molecular biology approaches to investigate how controlled DNA rearrangements are carried out and regulated.

Previous and current research

Our research focuses 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 the required 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 and assembly process 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, biophysics 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 gene regulatory landscapes, chromosomal engineering and even gene therapy (Ivics *et al*, 2009). We perform structural and functional studies to obtain a mechanistic understanding of this transposon and, in collaboration with the Gavin group (page 58), we also investigate how it interacts with human cells.

Sequence-specific elements: One of the main obstacles of gene therapy is integration of the therapeutic gene at unwanted locations. Therefore, we seek tools that integrate a genetic cargo to selected specific sequences and can provide a solution. Our recent work revealed the mechanism of the bacterial Insertion Sequence IS608, which uses a short sequence in the transposon DNA to guide its integration to a specific sequence via base pairing (Barabas *et al*, 2008). Consequently, the site of insertion can be altered by making point mutations in the transposon (Guynet *et al*, 2009). We are investigating if this target recognition can be extended to target unique genomic sites, which may provide an easy to customise programmable genetic tool. We are also studying a newly found mobile element – the plasticity zone transposon in *Helicobacter pylori* (Kersulyte *et al*, 2009) – focusing on how it moves and integrates to a 7nt-long specific sequence.

Transposon regulation: To avoid deleterious outcomes, cells must keep their transposons under control. Small RNAs can control transposon activity in various ways. In collaboration with the Carlomagno (page 57) and Pillai (page 109) groups, we investigate these processes in prokaryotes and eukaryotes. Our recent work on the eukaryotic piRNA pathway has revealed the structure and function of a novel component, a piRNA biogenesis factor called Zucchini.

Future projects and goals

- Building on mechanistic insights, we will develop novel genetic tools for genomic screening and transgenesis.
- Exploring the potential of transposon-based genetic tools in synthetic biology.
- Investigating the molecular mechanisms of somatic genome assembly in ciliated protozoa (a large-scale DNA rearrangement process resembling transposition).

Selected references

Voigt, F., Reuter, M., Kasaruho, A., Schulz, E.C., Pillai, R.S. & Barabas, O. (2012). Crystal structure of the primary piRNA biogenesis factor Zucchini reveals similarity to the bacterial PLD endonuclease *Nuc. RNA*, 18, 2128-34

Guynet, C., *et al*. (2009). Resetting the site: redirecting integration of an insertion sequence in a predictable way. *Mol. Cell*, 34, 612-9

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., *et al*. (2008). Mechanism of IS200/IS605 family DNA transposases: activation and transposon-directed target site selection. *Cell*, 132, 208-20

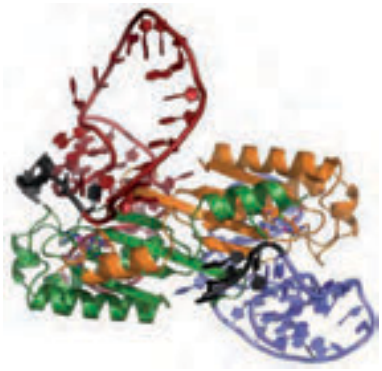


Figure 1: The structure of the IS608 transpososome, modelled based on a series of crystal structures

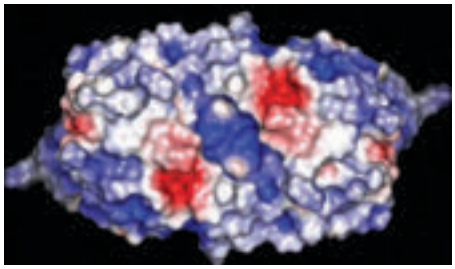


Figure 2: Crystal structure of the primary piRNA biogenesis factor Zucchini reveals its endonuclease function

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.
ERC Investigator.

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, 762-5

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., *et al*. (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 3D 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 (such as NMR and X-ray crystallography).

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

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

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

Future projects and goals

- 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 processes.

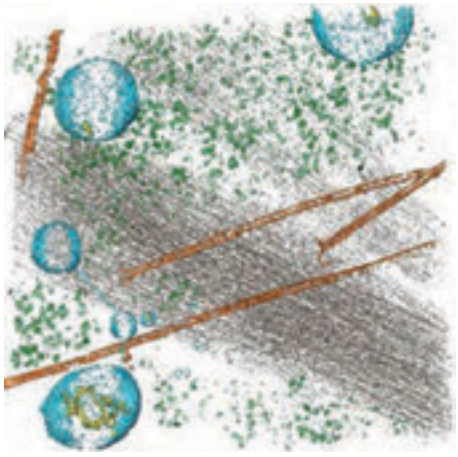


Figure 1: Cryo-electron tomogram of a fraction of the cytoplasm of a human cell. Microtubules are coloured in orange, stress fibres in grey, protein complexes in green, membranes in cyan and vesicular contents in yellow

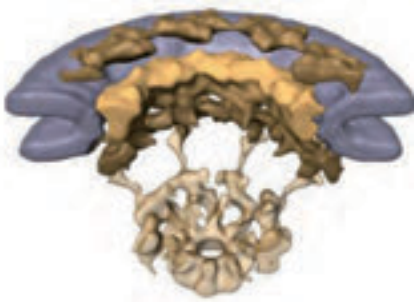


Figure 2: Structure of the nuclear pore complex. Membranes are coloured in grey, the scaffold structure in yellow and the nuclear basket in transparent brown

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 enveloped virus particles (such as HIV and Influenza virus) and coated transport vesicles (such as clathrin or COPI coated vesicles) are assembled. To make a virus or a vesicle, the building blocks and the cargo of the virus or vesicle are gathered together and interact with a lipid membrane, manipulating its shape and curvature to cause budding. To explore how this is achieved we study a range of different cellular and viral specimens. Our core methods are cryo-electron microscopy, cryo-electron tomography, and correlated light and electron microscopy.

Cryo-electron microscopy techniques are particularly appropriate for studying vesicles and viruses because they allow the shape of the membrane to be observed in its native state, while preserving information about the structure and arrangement of associated proteins. Computational image processing and 3D reconstructions are used to extract and interpret this information. As well as applying existing methods, we develop and apply novel microscopy and image processing approaches.

We take a step-by-step approach to derive structural information. Correlated fluorescence and electron microscopy methods can be used to locate and characterise features of interest. 3D 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 single particle or helical reconstruction methods, or by X-ray crystallography.

A particular emphasis of our research is on the structure and life-cycle of asymmetric membrane viruses, such as HIV. The structure and assembly of HIV particles offers insights into general features of membrane budding. Further details on our research into the structure and inhibition of HIV are available on our Molecular Medicine Partnership Unit website: **www.embl.de/mmpu/mmpu/research_groups/hiv/index.html**.

Future projects and goals

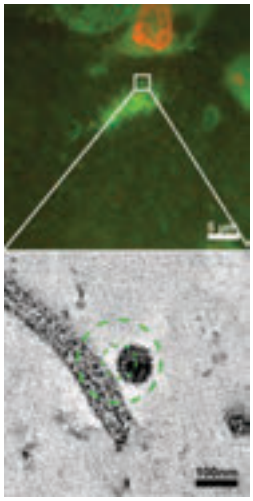
Our goal is to understand the interplay between protein assemblies, membrane shape and virus/vesicle structure. How do proteins induce the distortion of cellular membranes into vesicles of different dimensions? What are the similarities and differences between the variety of cellular budding events? 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? What kind of protein-protein interactions can drive virus assembly while maintaining structural flexibility? We are developing and applying novel microscopy and image processing approaches to address these questions.

Selected references

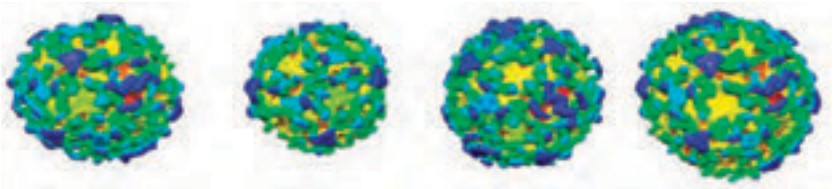
Kukulski, W., Schorb, M., Kaksonen, M. & Briggs, J.A. (2012). Plasma membrane reshaping during endocytosis is revealed by time-resolved electron tomography. *Cell*, 150, 508-20

Bharat, T.A., *et al*. (2012). Structure of the immature retroviral capsid at 8 A resolution by cryo-electron microscopy. *Nature*, 487, 385-9

Faini, M., *et al*. (2012). The structures of COPI-coated vesicles reveal alternate coatomer conformations and interactions. *Science*, 336, 1451-4



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



3D reconstructions of COPI coated vesicles obtained using cryo-electron tomography and sub-tomogram averaging (Faini et al. *2012)*

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

Carlomagno, T., *et al*. (2013). Structural principles of RNA catalysis in a 2'-5' lariat-forming ribozyme. *J. Am. Chem. Soc.*, e-pub ahead of print

Asami, S., Rakwalska-Bange, M., Carlomagno, T. & Reif, B. (2013). Protein-RNA interfaces probed by 1H-detected MAS solid-state NMR spectroscopy. *Angew Chem. Int. Ed. Engl.*, 52, 2345-9

Ballare, C., *et al*. (2012). Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nat. Struct. Mol. Biol.*, 19, 1257-65

Simon, B., *et al*. (2011). Recognition of 2'-O-methylated 3'-end of piRNA by the PAZ domain of a Piwi protein. *Structure*, 19, 172-80

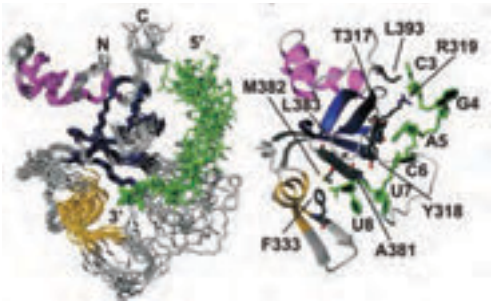


Figure 1: Overview of the structure of the complex between the Miwi-PAZ domain and the 3'-end 2'-O-methylated piRNA

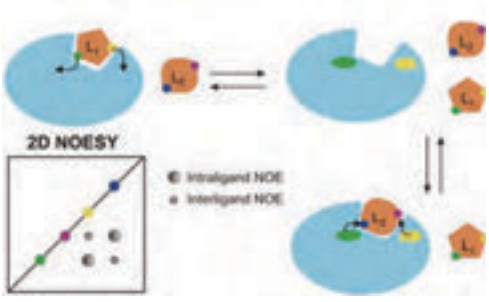


Figure 2: Schematic representation of the principle of the INPHARMA NOEs

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: i) structure-activity and dynamics-activity relationships of RNP complexes and catalytic RNAs involved in RNA processing; and ii) 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 109) 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, in order 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

The rules that govern the behaviour of biological systems are the focus of intense research in the field of systems biology. The resulting models are expected to be predictive of different healthy and pathological conditions and might provide the general principles for the (re)engineering of biological systems. Our group has pioneered biochemical methods, coupled to quantitative mass-spectrometry, with the aim of systematically linking dynamic protein interaction networks to various phenotypes in model organisms, human cells and human pathogens. Long term, we aim to advance network biology and medicine through the integration of quantitative biochemistry, proteomics and structural biology, and define system-wide hypotheses explaining complex phenotypes and human diseases. We will contribute new strategies for the targeting of human pathologies and provide insight into fundamental principles and rules guiding biomolecular recognition.

Charting biological networks: The way biological systems organise themselves in dynamic, functional assemblies with varying levels of complexity remains largely elusive. One of our main focuses is on deciphering the molecular mechanisms of cell function or dysfunction, which relies to a large extent on tracing the multitude of physical interactions between the cell's many components. We apply a range of biochemical and quantitative mass spectrometry approaches to organisms including yeast, a human pathogen and human somatic stem cells. The datasets guide the identification of drug targets and help us understand the mechanisms and side-effects of therapeutic compounds. Incorporation of structural models, single-particle electron microscopy, and cellular electron tomograms (collaboration with structural groups at EMBL) provide supporting details for the proteome organisation.

Development of new methods for charting new types of biological networks: While protein–protein and protein–DNA networks currently produce spectacular results, other critically important cellular components – metabolites – have rarely been studied in systematic interaction screens and remain best known for their housekeeping, metabolic functions. We currently focus on lipids and have developed new technologies with the capacity to produce systematic datasets measuring protein-lipid interactions. We designed miniaturised arrays of artificial membranes on a small footprint, coupled to microfluidic systems. We have also combined protein fractionation and lipidomics to characterise soluble protein-lipid complexes. We aim to extend the analyses to the entire proteome and lipidome and develop more generic approaches measuring all protein-metabolite interactions.

Future projects and goals

- Development of chemical biology methods based on affinity purification to monitor protein-metabolite interactions.
- Global screen aiming at the systematic charting of the interactions taking place between the proteome and the metabolome in the model organism *Saccharomyces cerevisiae* and in human.
- Development of new and existing collaborations to tackle the structural and functional aspects of biomolecular recognition.

Selected references

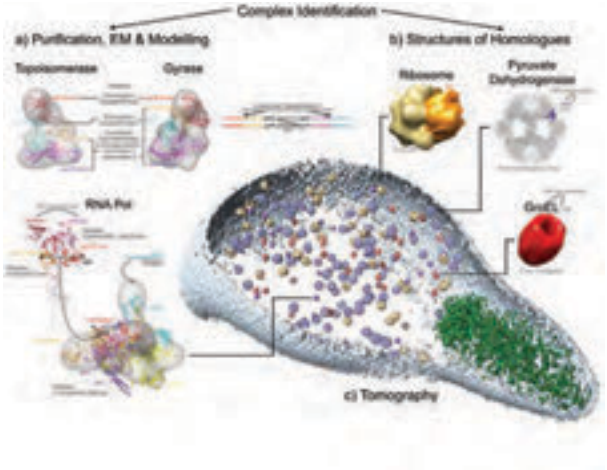
van Noort, V., *et al*. (2012). Cross-talk between phosphorylation and lysine acetylation in a genome-reduced bacterium. *Mol. Syst. Biol.*, 8, 571

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

Gehlenborg, N., *et al*. (2010). Visualization of omics data for systems biology. *Nat. Methods*, 7, S56-68

Kühner S., *et al*. (2009). Proteome organization in a genome-reduced bacterium. *Science*, 326, 1235-40

The group studies diverse organisms: the yeasts Saccharomyces cerevisiae, Chaetomium thermophilum *(thermophilic eukaryote), the human pathogen* Mycoplasma pneumoniae, *and human somatic stem cells (MMPU group and EU-funded SyStemAge), with datasets contributing detailed cartographies of biological processes relevant to human health or disease. Another major goal is the generation of organism-wide, systematic datasets of protein-metabolite regulatory circuits, and hypotheses or models concerning the consequences of dysfunction in human diseases*



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

Jiang, K., *et al*. (2012). A proteome-wide screen for mammalian SxIP motif-containing microtubule plus-end tracking proteins. *Curr. Biol.*, 22, 1800-7

Weatheritt, R.J. & Gibson, T.J. (2012). Linear motifs: lost in (pre)translation. *Trends Biochem. Sci.*, 37, 333-41

Van Roey, K., Gibson, T.J. & Davey, N.E. (2012). Motif switches: decision-making in cell regulation. *Curr. Opin. Struct. Biol.*, 22, 378-85

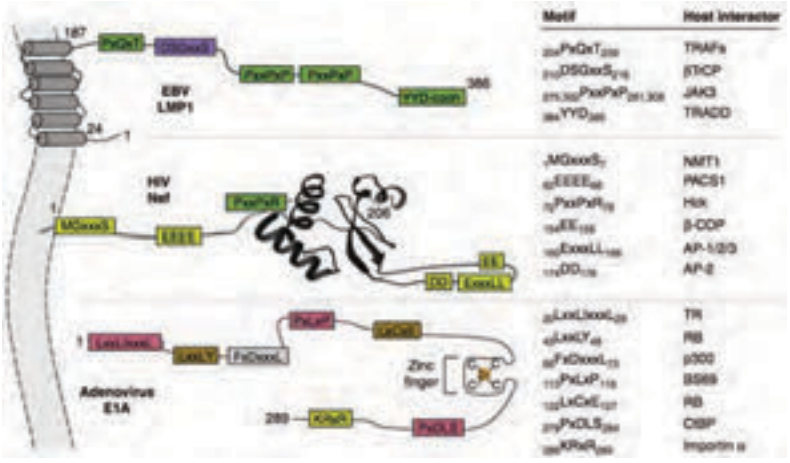
Dinkel, H., *et al*. (2012). ELM – the database of eukaryotic linear motifs. *Nucleic Acids Res.*, 40, D242-51

Previous and current research

Regulatory decisions during eukaryotic cell signalling are made within large dynamic protein complexes. Cell regulation is networked, redundant and, above all, cooperative. Decisions are made by in-complex molecular switching (see Van Roey, *et al*, 2012). 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. We look to collaborate with groups undertaking validation experiments. For example, in a recent interdisciplinary collaboration we performed bioinformatics analyses of the SxIP motif that is critical for the regulation of microtubule ends.

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 assist in targeting epitopes for antibody selection.



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

Future projects and goals

We will continue to hunt for regulatory motifs and may survey individual protein 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 these tools to investigate modular protein function and may deploy them in proteome and protein network analysis pipelines. We are 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.
Emmy Noether Group leader since 2010.
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. However, even if all 3D protein structures were available, our view of the molecular building blocks of cellular function would still be incomplete as we now know that many proteins are intrinsically disordered, which means that they are unfolded in their native state. Interestingly, the estimated percentage of intrinsically disordered proteins (IDPs) grows with the complexity of the organism (eukaryotes ≈ 50%). In a modern view of systems biology, these disordered proteins are believed to be multi-functional signalling hubs and their ability to adopt multiple conformations is considered a major driving force behind their evolution and enrichment in eukaryotes.

While the importance of IDPs in biology is now well established, many common strategies for probing protein structure are incompatible with molecular disorder and the highly dynamic nature of those systems. In contrast, single molecule techniques, which directly probe the distribution of molecular events, can reveal important mechanisms that otherwise remain obscured. In particular, single molecule fluorescence studies allow probing of molecular structures and dynamics at near atomic scale with exceptional time resolution. While such experiments are even possible in the natural environment of the entire cell, single molecule fluorescence studies require labelling with special fluorescent dyes, which still hampers the broad application of this technique. In our laboratory we are utilising a large spectrum of chemical biology and state-of-the-art protein engineering tools to overcome this limitation, with genetically encoding unnatural amino acids as one of our primary strategies. With a focus on studying biological questions, we also continue to develop new methods and recruit techniques from other disciplines (such as microfluidics), whenever they promise to assist our overall goal of improving our biological understanding.

Future projects and goals

Recent studies have shown that even the building blocks of some of the most complex and precise machines with an absolute critical role to survival of the cell are largely built from IDPs. For example, many nucleoporins are known to have central roles in the nuclear pore complex, but also in chromatin organisation, epigenetic mechanisms, transcription and oncogenesis. How this multifunctionality can be encoded into protein disorder is a central question in biology. We aim to explore the physical and molecular rationale behind the fundamental role of IDPs by combining molecular biology and protein engineering tools with single molecule biophysics.

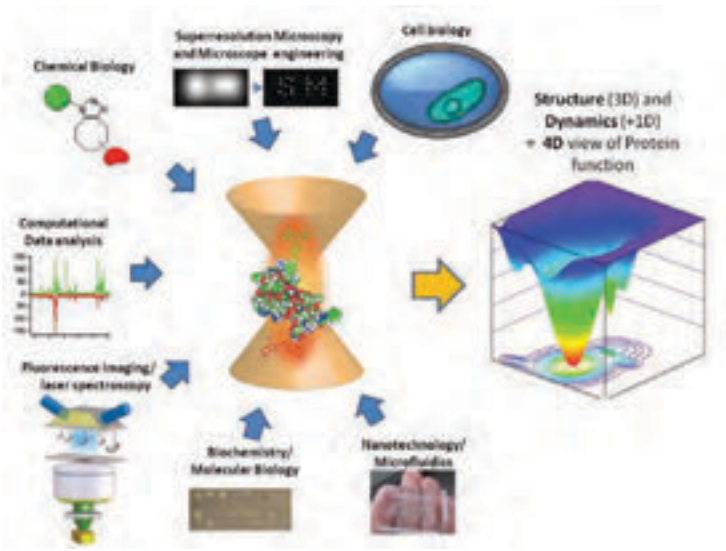
Selected references

Milles, S., Huy Bui, K., Koehler, C., Eltsov, M., Beck, M., & Lemke, E.A. (2012). Facilitated aggregation of FG nucleoporins under molecular crowding conditions. *EMBO Rep*. e-pub ahead of print

Milles, S., Koehler, C., Gambin, Y., Deniz, A.A. & Lemke, E.A. (2012). Intramolecular three-colour single pair FRET of intrinsically disordered proteins with increased dynamic range. *Mol. Biosyst.*, 8, 2531-4

Plass, T., *et al*. (2012). Amino acids for diels-alder reactions in living cells. *Angew Chem. Int. Ed. Engl.*, 51, 4166-70

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



Our long-term goal is to interface a large set of tools with our home-built, highly sensitive equipment to study structure and dynamics, even of heterogeneous biological systems such as nuclear pore complexes, chromatin, and transcription, 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

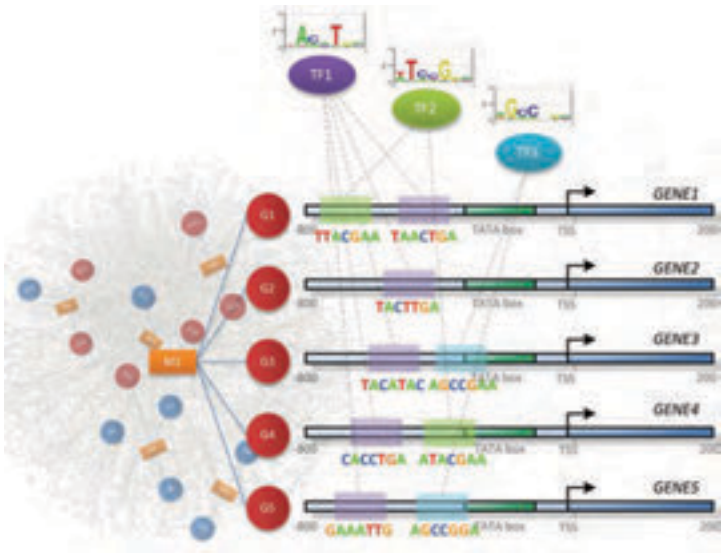
Cakir, T., Patil, K.R., Onsan, Z., Ulgen, K.O., Kirdar, B. & Nielsen, J. (2006). Integration of metabolome data with metabolic networks reveals reporter reactions. *Mol. Syst. Biol.*, 2, 50

Patil, K.R. & Nielsen, J. (2005). Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc. Natl. Acad. Sci. USA*, 102, 2685-9

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 hotspots. 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 cryo-electron 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 cryo-electron microscopy to study protein aggregates, typical of neurodegerative diseases such as Alzheimer's, and the mechanisms cells normally use to eliminate them.

Previous and current research

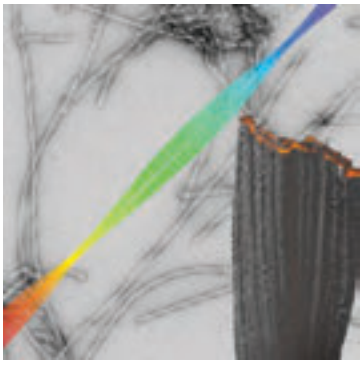
The molecular hallmark of neurodegenerative diseases, such as Alzheimer's and Parkinson's, is the formation of large protein aggregates called amyloid fibrils. The net build-up of these fibrillar aggregates is a result of an imbalance in the cellular production and clearance of misfolded polypeptides.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

We are visualising the molecules by cryo-electron 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 cryo-electron 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

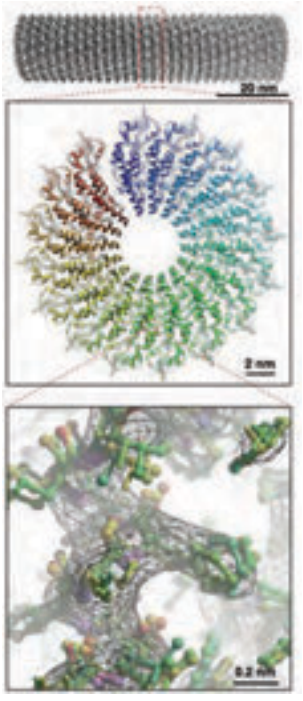
Selected references

Guichard, P., *et al*. (2012). Cartwheel architecture of *Trichonympha* basal body. *Science*, 337, 553

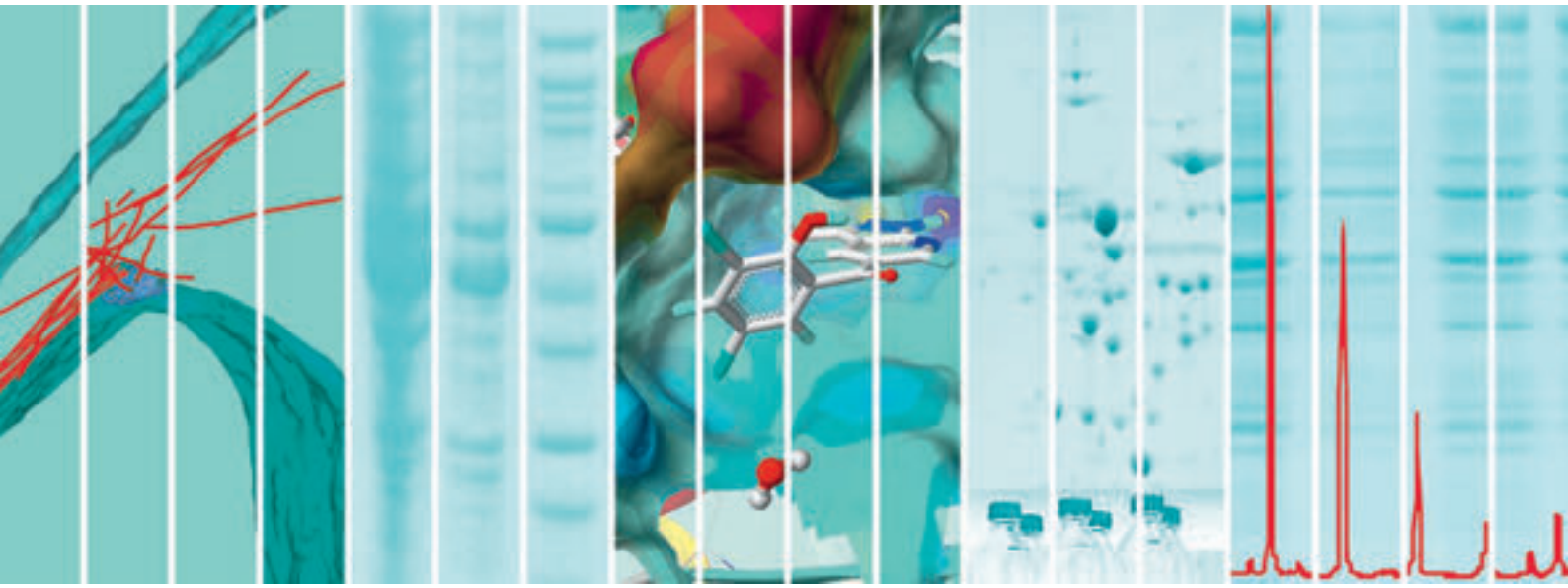
Low, H.H., Sachse, C., Amos, L.A. & Lowe, J. (2009). Structure of a bacterial dynamin-like protein lipid tube provides a mechanism for assembly and membrane curving. *Cell*, 139, 1342-52

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



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.

In line with EMBL staff turnover rules, a number of our technical employees as well as Heads of Core Facilities have left for pastures new and in the past year we have welcomed a number of new staff. Yannick Schwab has replaced Claude Antony as Head of the Electron Microscopy Core Facility; Alexis Perez has taken over as Head of the Flow Cytometry Core Facility from Andrew Riddell; and Alan Sawyer, former Head of the Monoclonal Antibody Core Facility has left EMBL to start his own business and we have now closed this in-house activity. We wish all great success in their new ventures. Core Facilities will be externally reviewed (SAC and external experts) for the third time in 2014.

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.
Senior scientist since 2012.

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

Major projects and accomplishments

- The ALMF presently manages 18 top-of-the-line microscope systems and eight high-content screening microscopes from leading industrial companies, as well as four image analysis workstations.
- More than 50 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 50 000 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.
- Supporting all aspects of automated microscopy and high content screening projects, including RNAi technology.

Technology partners

The ALMF presently has collaborations with the following companies:

- Applied Precision*
- Bitplane*
- Carl Zeiss*
- Eppendorf*
- Lambert Instruments*
- Leica Microsystems*
- Olympus*
- Perkin Elmer*
- PicoQuant
- Scientific Volume Imaging (SVI)

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

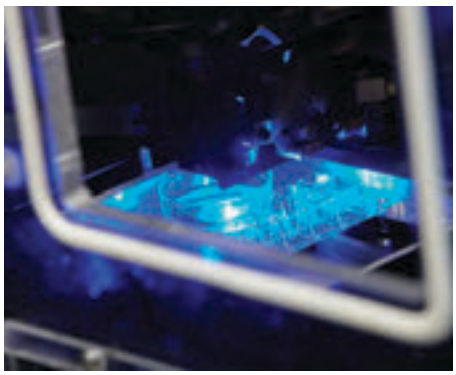
Selected references

Simpson, J.C., *et al*. (2012). Genome-wide RNAi screening identifies human proteins with a regulatory function in the early secretory pathway. *Nat. Cell Biol.*, 14, 764-74

Conrad, C., *et al.* (2011). Micropilot: automation of fluorescence microscopy-based imaging for systems biology. *Nat. Methods*, 8, 246–9

Neumann, B., *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



The ALMF manages 18 advanced microscope systems and eight high-content screening microscopes

Chemical Biology Core Facility

The facility assists 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. Our aim is to enable research groups to address biological questions by identifying and developing 'biotool' compounds against novel targets. We can assist groups in the development of primary and secondary assays for screening against our in-house compound library and guide them through the process of developing tool compounds for their specific target. Chemical optimisation projects can be done in collaboration with our chemistry partners. The facility is a collaboration between EMBL, the German Cancer Research Center (DKFZ), and the University of Heidelberg (since February 2012) to provide the infrastructure and expertise to open up small molecule development to research groups at these 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 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 identified in the facility, targeting specific cancer cell signalling pathways and the influenza virus respectively.

Services provided

Our screening library is composed of around 80 000 compounds. The selection focused 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. 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, Certara, GE, TTP Labtech.
- **Chemistry partners:** ChemDiv, Chembridge and Enamine.

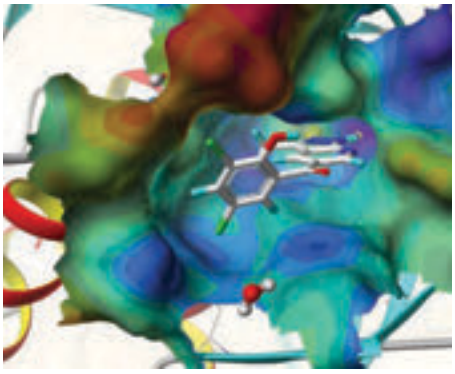
Selected references

Bartonova, V., *et al.* (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., *et al*. (2008). The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat. Struct. Mol. Biol*, 15, 500-6

Polycarpou-Schwarz, M., *et al*. (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. *J. Biomol. Screen.*, 12, 560-7



Ligand docked into target protein

Electron Microscopy Core Facility

Yannick Schwab

PhD 2001, Louis Pasteur University, Strasbourg.

Postdoctoral research at the University of Calgary, Canada and at the IGBMC, Illkirch, France.

Head of Electron Microscopy at the Imaging Center, IGBMC, Illkirch, France.

Facility head and team leader at EMBL since 2012.

The facility provides training in 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 and 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 welcome using this technology, and can be carried out in combination with methods developed at EMBL, such as by the Briggs group (page 56).

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, University of Heidelberg).

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 to 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).

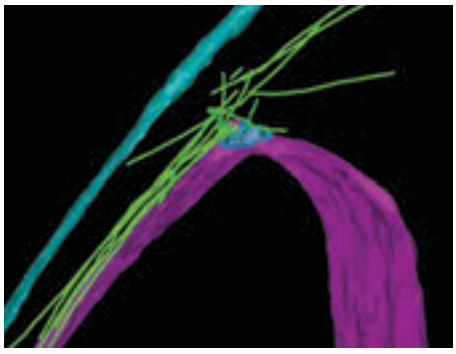
Selected references

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

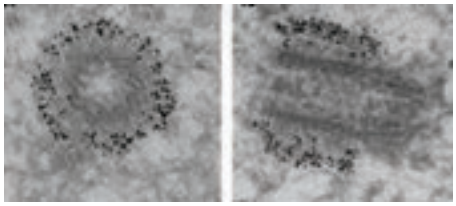
Tängemo C., *et al*. (2011). A novel laser nanosurgery approach provides support for *de novo* Golgi biogenesis in mammalian cells. *J. Cell Sci.*, 124, 978-87

Cizmecioglu, O., *et al*. (2010). Cep152 acts as a scaffold for recruitment of Plk4 and CPAP to the centrosome. *J. Cell Biol.*, 191, 731-9

Reuter, M., *et al*. (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, German Cancer Research Centre (see Cizmecioglu et al.*, 2010))*

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.

Alexis Pérez González

PhD 2003, Center of Molecular Immunology & University of Havana, Cuba.

Postdoctoral research and cytometry lab manager at Gulbenkian Institute of Science, Oeiras, Portugal

At EMBL since 2006. Facility Manager since 2012.

Selected references

Bonn, S., Zinzen, R.P., Perez-Gonzalez, A., Riddell, A., Gavin, A.C. & Furlong, E.E. (2012). Cell type-specific chromatin immunoprecipitation from multicellular complex samples using BiTS-ChIP. *Nat. Protoc.*, 7, 978-94.

Bonn, S., *et al*. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat. Genet.*, 44, 148-56.

Caramalho, I., Rodrigues-Duarte, L., Perez, A., Zelenay, S., Penha-Gonçalves, C. & Demengeot, J. (2011). Regulatory T cells contribute to diabetes protection in lipopolysaccharide-treated non-obese diabetic mice. *Scand J. Immunol.*, 74, 585-95

Prakash, H., *et al*. (2010). Sphingosine kinase-1 (SphK-1) regulates *Mycobacterium smegmatis* infection in macrophages. *PLoS One*, 5, e10657

We offer a wide range of flow cytometric techniques. Our equipment adds flexibility in the preparation and execution of experiments, allowing different approaches to addressing scientific problems. Our facility strives to meet researchers' needs and enable the highest possible resolution in terms of analysis and product.

Major projects and accomplishments

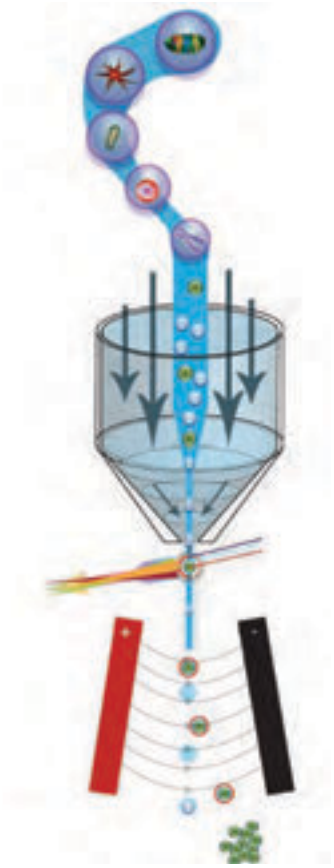
- Cell cloning by Darwinian selection, requiring several rounds of single cell sorting of target populations into 96-well plates, agar plates or other culture vessels. Examples of how this technique is being used include the generation of homogeneous cell lines with stably genome-integrated transfected genes and the selection of clones carrying improved fluorescent protein variants out of genetically engineered libraries.
- High-resolution analysis through photo saturation of dimly fluorescent bi-stable states in reworked bacterial signalling cascades.
- Evaluation of FRET probes in a novel assay for the evaluation of apoptosis.
- High throughput sorting of low frequency tissue-specific fly embryo nuclei carrying dim fluorescent signals as a preparative step in the analysis of chromatin modifications with regulatory function.
- Chromosome karyotyping and sorting for DNA sequencing and proteomics studies.

Services provided

- Complex multi-colour analysis of cell populations based on light scatter, fluorescent probes content and light intensities (including polarisation).
- Sorting of rare populations out of a heterogeneous particle mix. Cell cloning, particle enrichment and high purity bulk sorts.
- Providing EMBL scientific staff with expertise in flow cytometric techniques required in their research projects.
- Providing our researchers with advice and training in the use of flow cytometry, instrument operation and post-acquisition data analysis.
- Developing novel flow cytometric techniques to meet EMBL's diverse scientific needs.

Technology partners

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



The facility provides key services, such as sorting heterogeneous cell populations into homogeneous populations based on their fluorescence

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). The implementation 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 MiSeq and Ion Torrent sequencers. Preparation of MPS libraries for various applications is supported by a robust instrumentation infrastructure (e.g. Covaris, Bioanalyzer, AAT Fragment Analyzer, Qubit, and more). To deal with increasing numbers of incoming samples, we recently reinforced our instrumentation infrastructure through acquisition of Beckman FX liquid handling robots.

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 acid-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 & re-sequencing of genomic DNA.
- Targeted enrichment (sequence capture) in solution coupled with MPS.

GeneCore continues to establish new protocols enabling the processing of challenging samples such as low input or metagenomics samples. For analysis of MPS data, we work intensively with EMBL's bioinformatics community on the development of in-house and freely accessible tools. To date, GeneCore has generated around 3 terabases of MPS sequence data for its users.

Training is another crucial aspect of our work – GeneCore staff tutor individual researchers and also organise practical courses on corresponding subjects.

Services provided

- MPS sequencing, microarrays (homemade, commercial).
- miRNA qPCR profiling, Bioanalyzer, liquid handling robotics.
- Access to instruments and complete support: qPCR, NanoDrop, PCR cyclers.

We offer processing of samples for a range of microarray applications (mRNA, miRNA and other ncRNA expression profiling, comparative genome hybridisation) available from Affymetrix and Agilent platforms and, upon demand, spotting of customised arrays. In addition to three qPCR instruments managed by GeneCore, our qPCR capacity has been considerably enhanced by a Fluidigm Biomark HD instrument – a device capable of quantitation of transcripts on a single cell level.

Technology partners

MPS is still a very dynamic and rapidly evolving technology. We collaborate with several companies involved in developing MPS-related products, for instance testing them in our workflows. GeneCore is a member of the early-access program of Illumina, Agilent, NuGEN and Beckman Coulter.

Selected references

Xiol, J., *et al*. (2012). A role for Fkbp6 and the chaperone machinery in piRNA amplification and transposon silencing. *Mol. Cell*, 47, 970-9

Simpson, J.C., *et al*. (2012). Genome-wide RNAi screening identifies human proteins with a regulatory function in the early secretory pathway. *Nat. Cell Biol.*, 14, 764-74

Kahramanoglou, C., *et al*. (2012). Genomics of DNA cytosine methylation in *Escherichia coli* reveals its role in stationary phase transcription. *Nat. Commun.*, 3, 886

Sanchez, M., *et al*. (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, e168-79

GeneCore processes users' samples with the help of top-end instruments

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.

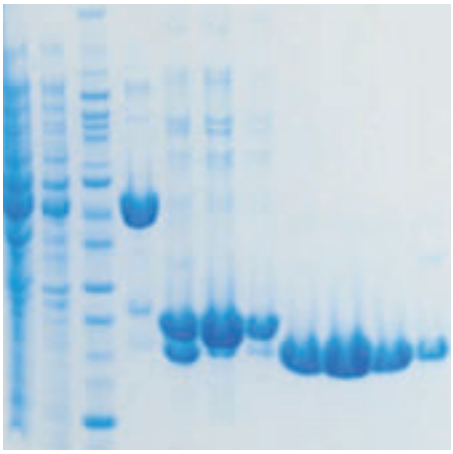
Selected references

Costa, S.J., Almeida, A., Castro, A., Domingues, L., & Besir, H. (2012) The novel Fh8 and H fusion partners for soluble protein expression in *Escherichia coli*: a comparison with the traditional gene fusion technology. *Appl. Microbiol. Biotechnol*. Epub ahead of print

Mackereth, C.D., *et al*. (2011). Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature*, 475, 408-11

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

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 facility 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, it is possible to integrate the insert into all of the vectors due to the universal overlaps that are present in the linearised vectors and the PCR product. A lethal gene insert in the original template vectors inhibits the growth of false positive colonies, which reduces the number of clones to test for the correct insert. With this new vector set, one can test the expression of a gene in different expression systems in parallel and avoid the 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 these 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 markers 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 (analytical ultracentrifugation, for example).

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 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.

Proteomics:

- Protein identification from gel or in solution.
- High-resolution and high mass accuracy MS, MSMS, and LC-MSMS (Thermo Orbitrap Velos Pro) for identification and quantification of proteins in complex mixtures.
- Ion trap (Bruker HCT) LC-MSMS for routine identification of proteins from coomassie and silver-stained gels.
- Protein quantification by stable-isotope labelling (SILAC and dimethyl labelling).
- Identification of post-translational modifications.
- Enrichment of phosphopeptides (TiO2 and IMAC).
- Multi-dimensional peptide separation (isoelectric focusing and liquid chromatography)

Technology partner

- BIO-RAD

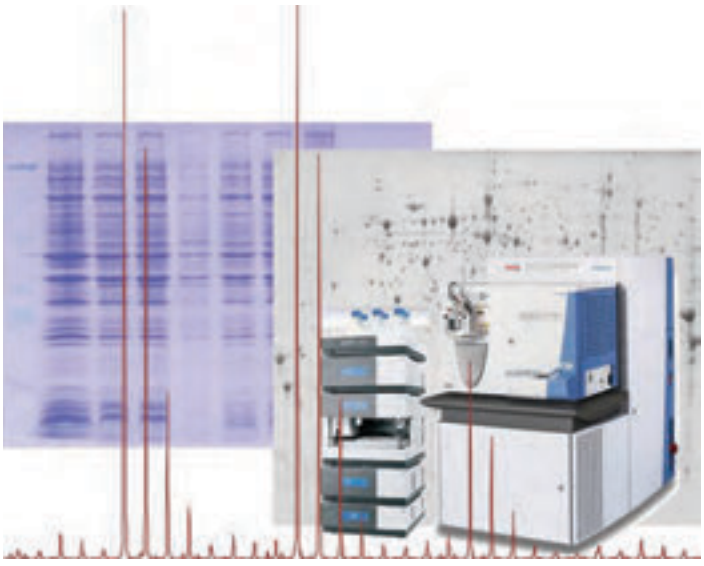
Selected references

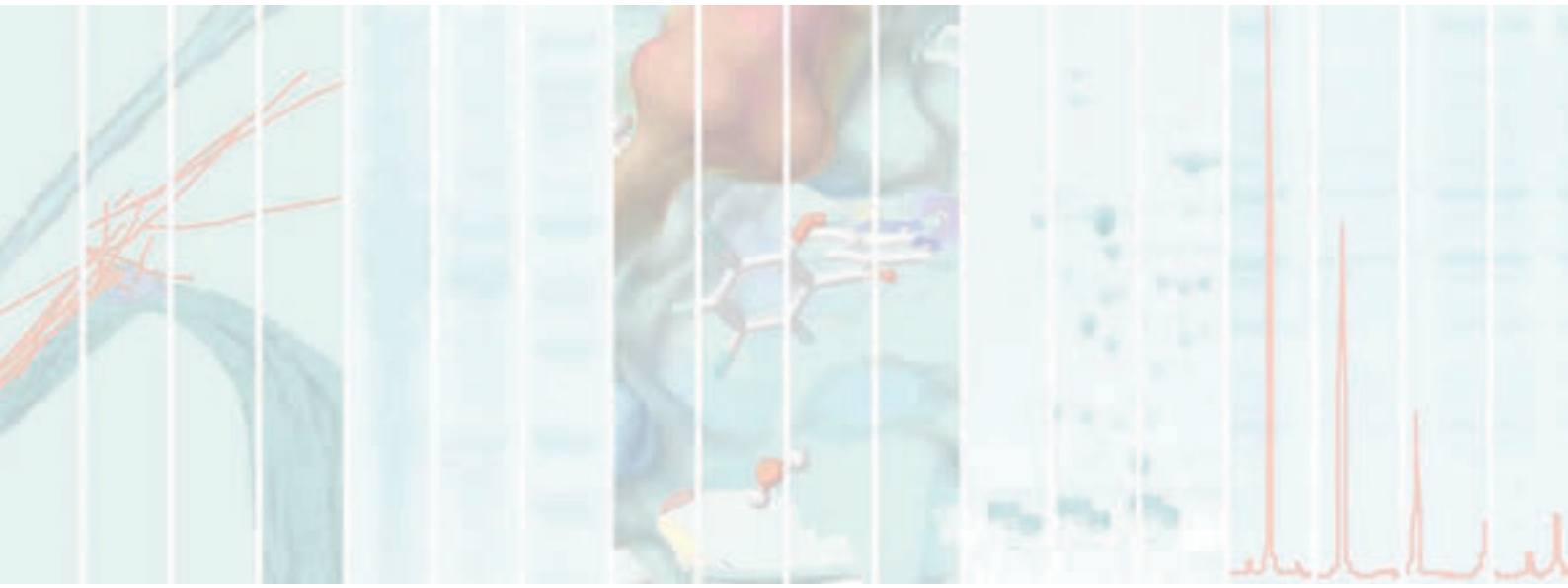
Dastidar, E.G., *et al*. (2013). Comprehensive histone phosphorylation analysis and identification of pf14-3-3 protein as a histone h3 phosphorylation reader in malaria parasites. *PLoS One*, 8, e53179

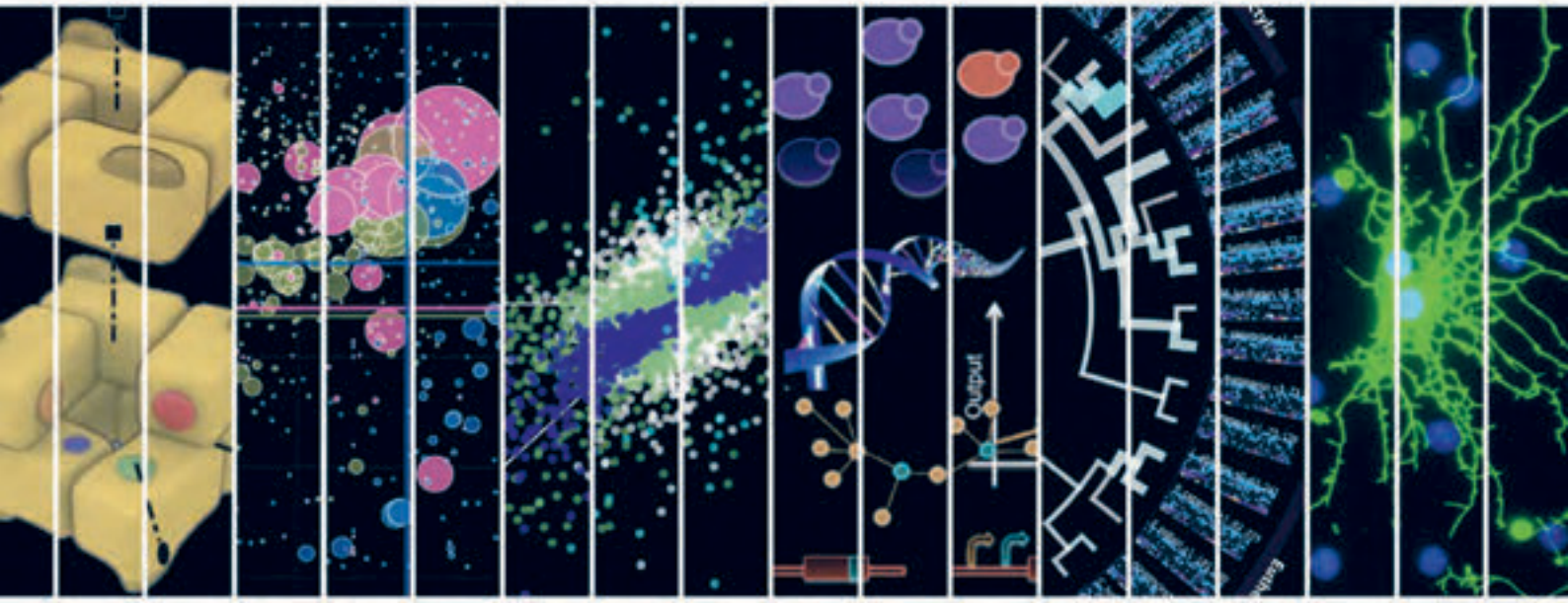
Klimmeck, D., *et al*. (2012). Proteomic cornerstones of hematopoietic stem cell differentiation: distinct signatures of multipotent progenitors and myeloid committed cells. *Mol. Cell Proteomics*, 11, 286-302

Castello, A., *et al*, (2012). Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell*, 149, 1393-406

Moriniere, J., *et al*. (2009). Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature*, 461, 664-8






Output

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 curiosity-driven 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 EMBL-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 and 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 this 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

Protein 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.

Professor of Biomolecular Structure, UCL since 1990.

Bernal Professor 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 learn more about the 3D structure and evolution of proteins, for example by studying how enzymes perform catalysis, and how the insulin signalling pathway affects 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. We explore how enzymes perform catalysis by gathering relevant data from the literature and developing novel software tools that allow us to characterise enzyme mechanisms and navigate the catalytic and substrate space. In parallel, we investigate the evolution of these enzymes to discover how they can evolve new mechanisms and specificities. This involves integrating heterogeneous data with phylogenetic relationships within protein families, which are based on protein structure classification data derived by colleagues at University College London (UCL). 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.

We also explore sequence variation between individuals in different contexts and for different species. To understand more about the molecular basis of ageing in different organisms, we participate in a strong collaboration with experimental biologists at UCL. Our role is to analyse functional genomics data from flies, worms and mice and, by developing new software tools, relate these observations to effects on lifespan.

Selected references

Furnham, N., *et al*. (2012). Exploring the evolution of novel enzyme functions within structurally defined protein superfamilies. *PLoS Comp. Biol.* 8, e1002403

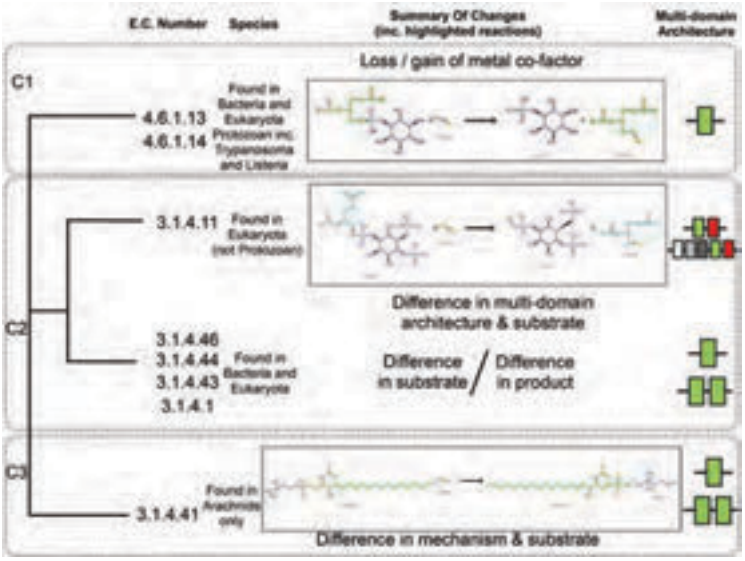
Furnham, N., *et al*. (2012). FunTree: a resource for exploring the functional evolution of structurally defined enzyme superfamilies. *Nucleic Acids Res*., 40, D776-82

Papatheodorou, I., *et al*. (2012). Using answer set programming to integrate RNA expression with signalling pathway information to infer how mutations affect ageing. *PLoS One*, 7, e50881

Future projects and goals

Our work on understanding enzymes and their mechanisms using structural and chemical information will include a study of how enzymes, their families and pathways have evolved. We will study sequence variation in different individuals, including humans, flies and bacteria, and explore how genetic variations impact on the structure and function of a protein and sometimes cause disease. We will seek to gain a better understanding of reaction space and its impact on pathways, and to use this new knowledge to improve chemistry queries across our databases. Using evolutionary approaches, we hope to improve our prediction of protein function from sequence and structure. We will also improve our analyses of survival curves and combine data with network analysis for flies, worms and mice in order to compare the different pathways and ultimately explore effects related to human variation and age.

The evolution of novel enzyme functions: The phosphatidylinositol (PI) phosphodiesterase domain superfamily divides into three clades (C1–C3) by phylogenetic analysis. Significant changes in function occur between clades (as well as within clades) as defined by the Enzyme Commission classification number with underlying changes in reaction chemistry as captured by atom-atom mapping of reaction small molecules (see reaction diagrams) as well as changes in multi-domain architecture



Analysis of protein and RNA sequence

The Bateman group endeavors to classify proteins and certain RNAs into functional families with a view to producing a 'periodic table' of these molecules.

Alex Bateman

PhD 1997, University of Cambridge.

Postdoctoral work at the Sanger Centre.

Group leader at Wellcome Trust Sanger Institute 2001-2012.

Head of Protein Sequence Resources at EMBL-EBI since 2012.

Selected references

Buljan, M., *et al*. (2012). Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol. Cell,* 46, 871-83

Buljan, M., Frankish, A., & Bateman, A. (2010). Quantifying the mechanisms of domain gain in animal proteins. *Genome Biology*, 11, R74

Bateman, A., *et al*. (2009). Phospholipid scramblases and Tubby-like proteins belong to a new superfamily of membrane tethered transcription factors. *Bioinformatics*, 25, 159-62

Schuster-Böckler, B. & Bateman, A. (2008). Protein interactions in human genetic diseases. *Genome Biology*, 9, R9

Previous and current research

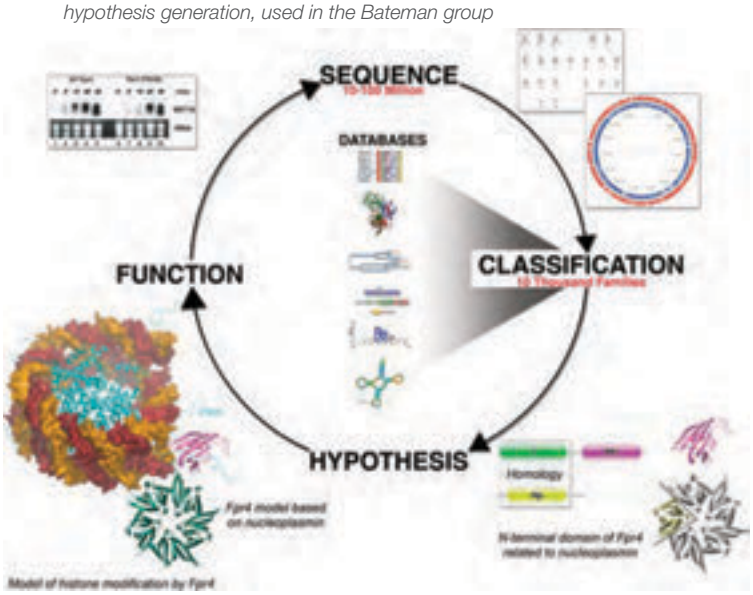
Our work has centred around the idea that there are a finite number of families of protein and RNA genes. We wish to enumerate all of these families to gain an understanding of how complex biological processes have evolved from a relatively small number of components. We have produced a number of widely used biological database resources such as Pfam, Rfam, TreeFam and MEROPS to collect and analyse these families of molecules. Over the years we have published a large number of novel protein domains and families of particularly high interest. For example, we discovered the Paz and Piwi domains, which allowed us to identify the Dicer proteins as having an important role in RNAi, several months before this was experimentally verified. More recently, we showed that the scramblase genes may act as membrane-tethered transcription factors.

Our research interests focus on how proteins and non-coding RNAs interact with each other and how these interaction networks can be rewired due to disease mutations or natural variation. We are interested in how proteins have evolved through the gain and loss of new protein domains. Recently we have been involved in using Wikipedia for collecting community annotation and other biological information for biological databases. Wikipedia provides an enormous opportunity for public engagement in science and we have been encouraging scientists in a number of ways to edit Wikipedia. Current research is looking at identification of non-coding RNAs and understanding the function through computational analysis.

Future projects and goals

We will continue to develop tools and databases to understand the function and evolution of RNA and proteins. Using this data and computational analyses we aim to investigate interaction networks in two directions. Firstly, we will investigate the plasticity of the protein interaction network between individuals. To do this we will identify natural human variation such as SNPs and CNVs that rewire the protein interaction network. The second direction we will take is to explore the large and growing set of important molecular interactions involving RNA that are currently dispersed among diverse databases and experimental studies. By bringing this data together we wish to uncover the extent and evolution of the RNA interaction network compared to the protein interaction network. In another strand of our research we will develop automated techniques to identify spurious protein predictions that are polluting sequence databases. We have collected thousands of examples of proteins that are unlikely to be translated. These examples will form a good training set for machine learning techniques to identify further suspicious proteins.

The flow of information, from sequence to knowledge of function through classification and hypothesis generation, used in the Bateman group



Evolution of Cellular Networks

Pedro Beltrao

PhD 2007, University of Aveiro (research conducted at EMBL Heidelberg).

Postdoctoral research at the University of California San Francisco, USA.

Group leader at EMBL-EBI since 2013.

The Beltrao group studies the molecular impacts of genetic variability on phenotypic variability in order to understand better the function and evolution of cellular networks.

Previous and current research

Our group is interested in understanding how novel cellular functions arise and diverge during evolution. We study the molecular sources of phenotypic novelties, exploring how genetic variability that is introduced at the DNA level is propagated through protein structures and interaction networks to give rise to phenotypic variability. Within the broad scope of this evolutionary problem, we focus on two areas: the function and evolution of post-translational regulatory networks; and the evolution of genetic and chemical-genetic interactions. Looking beyond evolutionary processes, we also seek to understand the genomic differences between individuals and improve our capacity to devise therapeutic strategies.

In collaboration with mass-spectrometry groups, we develop a resource of experimentally derived, post-translational modifications (PTMs) for different species in order to study the evolutionary dynamics and functional importance of post-translational regulatory networks. We use these data to create novel computational methods to predict PTM function and regulatory interactions. Our goal is to gain insights into the relationship between genetic variation and changes in PTM interactions and function.

Future projects and goals

Changes in cellular interaction networks underpin variation in cellular responses and sensitivity to environmental perturbations or small molecules. As we model and study the evolution of cellular interaction networks, we begin to see how different individuals or species diverge in their response to drugs. Understanding this relationship will enable us to develop methods to predict how genetic changes result in specific sensitivity to drug combinations.

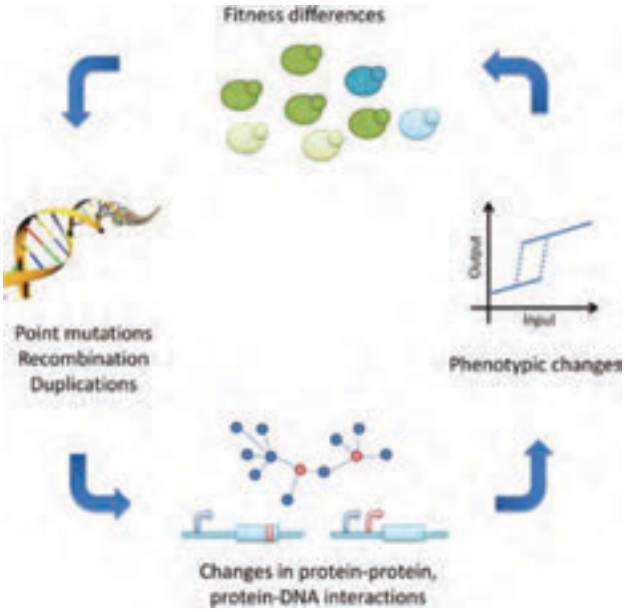
Selected references

Beltrao, P., *et al*. (2012). Systematic functional prioritization of protein post-translational modifications. *Cell,* 150, 413-25

Ryan, C.J., *et al*. (2012). Hierarchical modularity and the evolution of genetic interactomes across species. *Mol. Cell,* 46, 691-704

Charles, G., *et al.* (2011). A site-specific acetylation mark on the essential RSC chromatin remodeling complex promotes resistance to replication stress. *PNAS,* 108, 10620-5

Nichols, R.J., *et al*. (2011). Phenotypic landscape of a bacterial cell. *Cell*, 144, 143-56



Pluripotency, reprogramming and differentiation

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

Paul Bertone

PhD 2005, Yale University.

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

Associate Investigator, Wellcome Trust-Medical Research Council Stem Cell Institute, University of Cambridge.

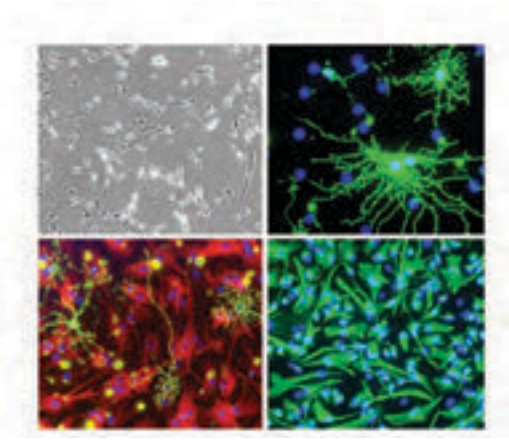
Selected references

Goldman, N., *et al*. (2013). Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature*, 494, 77-80

Reynolds, N., *et al*. (2012). NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. *Cell Stem Cell,* 10, 583-94

Engström, P.G., *et al*. (2012). Digital transcriptome profiling of normal and glioblastoma-derived neural stem cells identifies genes associated with patient survival. *Genome Med*., 4, 76

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



GNS cells propagate indefinitely in culture and can differentiate into the major cell types of the central nervous system, such as astrocytes and oligodendrocytes

Previous and current research

We investigate the cellular and molecular attributes of embryonic and tissue-specific stem cells using a combination of experimental and computational methods. We develop and apply genomic technologies to the analysis of stem cell function to address fundamental aspects of development and disease. Properties of proliferation, differentiation and lineage specialisation, for instance, are fundamental to cellular diversification and growth patterning during organismal development, as well as the initiation of cellular repair processes throughout life.

The fundamental processes that regulate cell differentiation are not well understood and are likely to be misregulated in cancer: another focus is the study of neural cancer stem cells derived from human glioblastoma multiforme tumours. Using neural stem cell derivation protocols it is possible to expand tumour-initiating, glioblastoma-derived neural stem (GNS) cells continuously *in vitro*. Although the normal and disease-related counterparts are highly similar in morphology and lineage marker expression, GNS cells harbour genetic mutations typical of gliomas and give rise to authentic tumours following orthotopic xenotransplantation. We apply genomic technologies to determine transcriptional changes and chromosomal architecture of patient-derived GNS cell lines and their individual genetic variants. These data provide a unified framework for the genomic analysis of stem cell populations that drive cancer progression, and contribute to the understanding of tumourigenesis.

Future projects and goals

We have in place the most robust and stable systems for stem cell derivation and propagation. To realise the potential of ES cells in species other than mouse, however, precise knowledge is needed of the biological state of these cells, and particularly of the molecular processes that maintain pluripotency and direct differentiation. Translating knowledge and methods to other mammalian species involves the characterisation of germline-competent ES cells from the laboratory rat, along with the production of pluripotent human iPS cells using alternative reprogramming strategies. Using deep transcriptome sequencing we have shown a broad equivalence in self-renewal capacity and cellular state, albeit with intriguing species-specific differences. Thus, while ground-state pluripotency can be captured and maintained in several species, the mechanisms used to repress lineage differentiation may be fundamentally different.

Cancer stem cells constitute a renewable resource of homogeneous cells that can be studied in a wide range of experimental contexts and provide key insights toward new therapeutic opportunities. To this end we are carrying out in-depth analyses of our GNS cell bank using comprehensive genetic and transcriptomic profiling. With these data we are developing methods to stratify glioblastoma classes based on the molecular attributes and differentiation capacities of tumour-initiating stem cells. Existing tumour subtypes are associated with diverse clinical outcomes and are important for prognostic value but, as with any analysis of complex tissues, previous results have suffered from sample heterogeneity. With access to the underlying stem cell populations derived from parental tumours, we are refining existing subtype classification to improve the diagnostic utility of this approach and performing functional experiments to identify alterations in GNS cells that impart tumourigenic potential.

Sequence algorithms and intra-species variation

Ewan Birney

PhD 2000, Wellcome Trust Sanger Institute.
At EMBL-EBI since 2000.
Joint Associate Director since 2012.

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

Previous and current research

Our group has a long-standing interest in developing sequencing algorithms. During the past four years a considerable focus has been on compression, with theoretical and now practical implementations of compression techniques. Our 'blue skies' research includes collaborating with the Goldman group (page 84) on a method to store digital data in DNA molecules. We continue to be involved in this area as new opportunities arise – including the application of new sequencing technologies.

We are also interested in the interplay of natural DNA sequence variation with cellular assays and basic biology. During 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 very general and need not be restricted to the human disease arena. Association analysis can be applied to nearly any measureable phenotype in a cellular or organismal system where an accessible, outbred population is available. We are pursuing association analysis for a number of molecular (e.g. RNA expression levels and chromatin levels) and basic biology traits in species where favourable populations are available including human and *Drosophila*. In the future we hope to expand this to a variety of other basic biological phenotypes in other species, including establishing the first vertebrate near-isogenic wild panel in Japanese rice paddy fish (Medaka, *Oryzias latipes*).

Future projects and goals

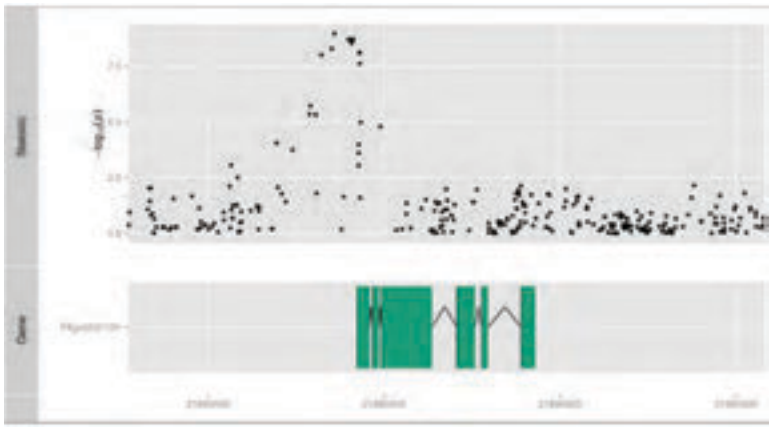
We will continue to work on sequence algorithms and intra-species variation. In humans there will be work on molecular phenotypes in an iPSC panel generated as part of the HipSci consortium where favourable populations are available including human, *Drosophila* and in the future, Medaka. In *Drosophila* we will look at multi-time-point developmental biology measures and will characterise the first vertebrate near isogenic panel, which we are developing with collaborators studying Medaka fish.

Selected references

Goldman, N., *et al*. (2013). Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature*, 494, 77-80

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489, 57-74

Spivakov, M., *et al*. (2012). Analysis of variation at transcription factor binding sites in *Drosophila* and humans. *Genome Biology*, 13, R49



The association of Drosophila *SNPs across a 20 kb genomic region with RNA expression levels of the FBgn0031191 gene contained in the region. The level of association is shown as –log10 of the P value on the Y-axis with genomic position on the X axis in the top panel. The genomic structure of the FBgn0031191 gene is shown in the bottom panel, with the direction of transcription from left to right. The most strongly associated SNPs in this case are clustered at the start of the gene, consistent with an effect through modification of promoter efficiency*

Functional genomics research

Research in the Brazma team focuses on analysis of gene, transcript and protein expression, on cancer genomics and proteomics, and on integrative analysis of functional genomics data.

Alvis Brazma

MSc in Mathematics 1982, University of Latvia.

PhD in Computer Science 1987, Moscow State University.

At EMBL-EBI since 1997.

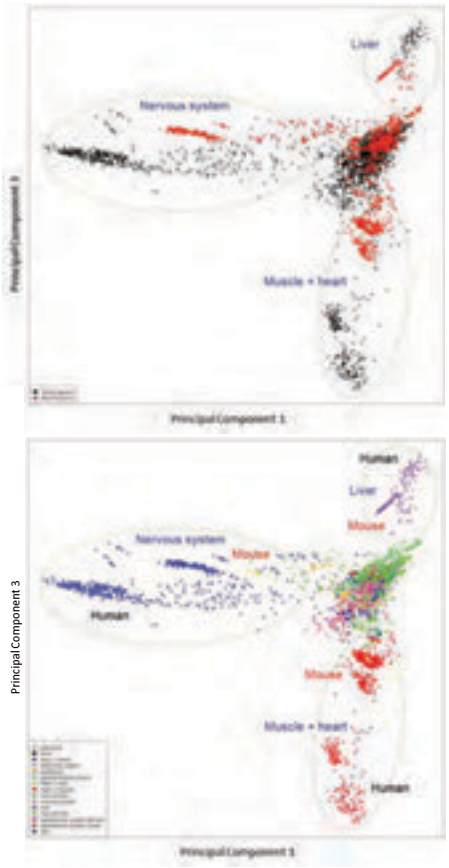
Selected references

Rung, J. & Brazma A. (2012). Reuse of public genome-wide gene expression data. *Nat. Rev. Genet.*, 14, 89-99

Fonseca, N.A., *et al*. (2012). Tools for mapping high-throughput sequencing data. *Bioinformatics*, 28, 3169-77

Goncalves, A., *et al*. (2012). Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. *Genome Res*., 22, 2376-84

Kapushesky M, *et al*. (2012). Gene Expression Atlas update – a value-added database of microarray and sequencing-based functional genomics experiments. *Nucleic Acids Res*., 40, D1077-81



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

Previous and current research

Signals that are too weak in an individual dataset may become clearer when data from many experiments are integrated. Combined datasets can also cover a larger variety of conditions, thus enabling exploration of problems of a different kind and enable the construction of 'global maps' of gene expression and other data. New analysis methods need to be developed; however, the main difficulty is usually in the systematic annotation and in understanding how datasets can be made comparable for meaningful integration (Rung & Brazma, 2012). Towards these goals, we work closely with EMBL-EBI's Functional Genomics and Expression Atlas teams (see Kapushesky *et al*, 2012). Recently, our research efforts have centred on analysing RNA sequencing-based gene expression data and integrating different types of data across multiple platforms (Fonseca, 2012). We collaborate closely with the Marioni group (page 85).

As a part of our participation in the European Commission-funded GEUVADIS project, we analysed mRNA and small RNA from lymphoblastoid cell lines of 465 individuals who participated in the 1000 Genomes Project. Our group led the analysis of transcript isoform use and fusion gene discovery. By integrating RNA and DNA sequencing data we were able to link gene expression and genetic variation and to characterise mRNA and miRNA variation in several human populations. All of the data generated in the project are available through the ArrayExpress database, which is run by our service team.

A particular area of interest for us is cancer genomics, where we lead gene expression analysis in the International Cancer Genomics Consortium project on renal cancer and transcript isoform usage. The human transcriptome contains in excess of 100 000 different transcripts, many times more than the number of genes; however, are all transcripts equally 'important'? By analysing transcript composition in 16 human tissues and five cell lines, we demonstrated that in a given condition most protein-coding genes have one major transcript expressed at significantly higher level than others, and that the same major transcripts are often expressed in many tissues. These observations show that the original dogma of molecular biology – one gene leading to one protein – may be closer to the truth than recently conjectured.

Future projects and goals

We will continue to focus on large-scale data integration and analysis with the goal of understanding gene expression patterns and their relation to phenotype. We will develop methods for RNA-seq data analysis and processing, and apply these to address important biological questions, such as the role of alternative splicing and splicing mechanisms. Together with our colleagues at the International Cancer Genome Consortium, we will investigate the impact of cancer genomes on functional changes in cancer development, for instance by exploring fusion genes and their role in cancer development. We will also work on integrating RNA expression data with proteomics data to understand all aspects of gene expression from the systems perspective.

Functional genomics and analysis of small RNA function

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.

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.

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.

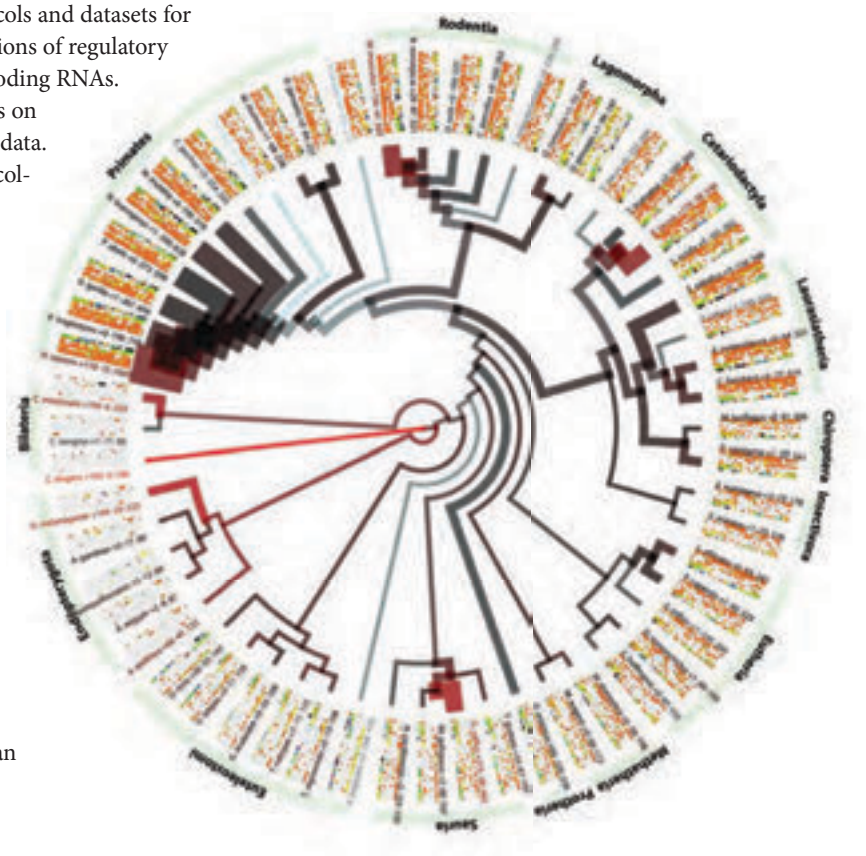
Our work involves the development of algorithms, protocols and datasets for functional genomics. We focus on determining the functions of regulatory RNAs, including microRNAs, piwiRNAs and long non-coding RNAs. We collaborate extensively with experimental laboratories on commissioning experiments and 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. 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, as 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.

Selected reference

Guerra-Assunção, J.A. & Enright A.J. (2012) Large-scale analysis of microRNA evolution. *BMC Genomics*, 13, 218



Evolutionary distribution of miRNA families. Phylogenetic tree representing miRNA family gains and losses. Branch width represents the number of miRNA families present among leaves of the branch, while the colour represents significant miRNA family loss (blue) or gain (red). For each of 408 miRNA families present at multiple loci in at least two species, we also build a graphical 'glyph'. This glyph can be used to quickly assess presence, absence or expansion of families between clades. Each square represents a specific miRNA family. Squares are coloured as follows: white indicates that this species does not contain a particular family, black indicates that this species contains at least 10 copies of miRNAs within that family. Copies between 1 and 10 are indicated as a rainbow gradient (red through violet). Groups of species are labelled according to the name of the evolutionary branch preceding them

Evolution of transcriptional regulation

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

Paul Flicek

DSc 2004, Washington University.

Honorary Faculty Member, Wellcome Trust Sanger Institute since 2008.

At EMBL-EBI since 2005.
Team leader since 2008.
Senior Scientist since 2011.

Selected references

Flicek, P., *et al*. (2012). Ensembl 2012. *Nucleic Acids Res*., 40, D84-90

Schmidt, D., *et al*. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell*, 148, 335-48

Mallon, A.M., *et al*. (2012). Accessing data from the International Mouse Phenotyping Consortium: state of the art and future plans. *Mamm. Genome*, 23, 641-52

Faure, A.J., *et al*. (2012). Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules. *Genome Res*., 22, 2163-75

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 Parkinson team (page 98), the International Mouse Phenotyping Consortium. We also collaborate on the development of 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 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.

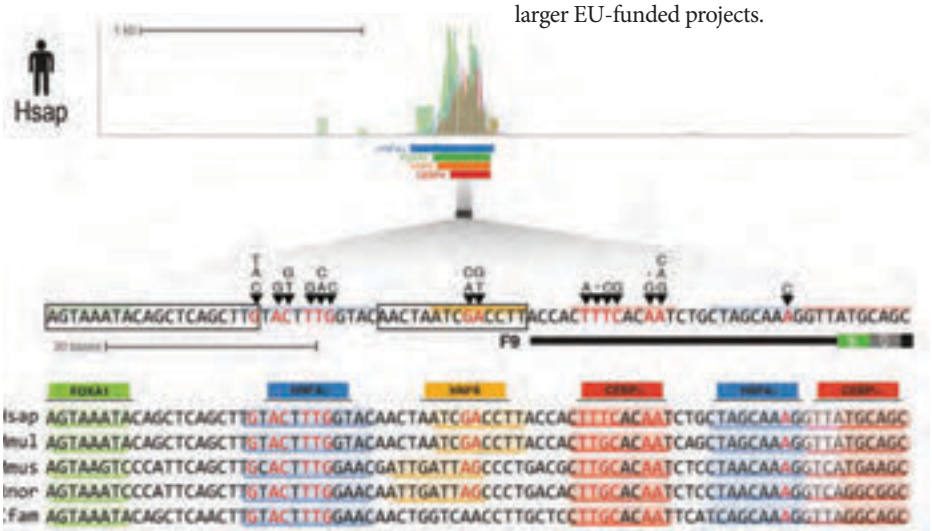
Our specific research projects focus on the evolution of transcriptional regulation. Recently, we have expanded our comparative regulatory genomics techniques, including mapping the same DNA–protein interactions in matched tissues in multiple species to understand how gene regulation has evolved while 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.

Future projects and goals

With the issue of major datasets from the EU-funded Blueprint project and the NIH-funded KOMP2 project, we 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. We will address these questions both in the context of our established collaborative projects with the Odom group at the University of Cambridge and as part of other collaborations, including larger EU-funded projects.

Evolutionarily conserved transcription factor binding around the Factor IX gene promoter in five species (human, macaque, mouse, rat and dog). Mutations known to cause haemophilia are indicated by triangles about the genome sequence and are consistently found at sites of conserved transcription factor binding



Evolutionary tools for genomic analysis

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.

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

Previous and current research

Evolution is the historical cause of the diversity of all life. The group's research focuses on the development of data analysis methods for the study of molecular sequence evolution and for the exploitation of evolutionary information to draw powerful and robust inferences about phylogenetic history and genomic function.

The evolutionary relationships between all organisms require that we analyse molecular sequences with consideration of the underlying structure connecting those sequences. We develop mathematical, statistical and computational techniques to reveal information present in genome data, in order to draw inferences about the processes that gave rise to that data and to make predictions about the biology of the systems whose components are encoded in those genomes.

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 these techniques to tackle biological questions of interest. We participate in comparative genomic studies, both independently and in collaboration with others, including 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.

Selected references

Goldman, N., *et al*. (2013). Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature*, 494, 77-80

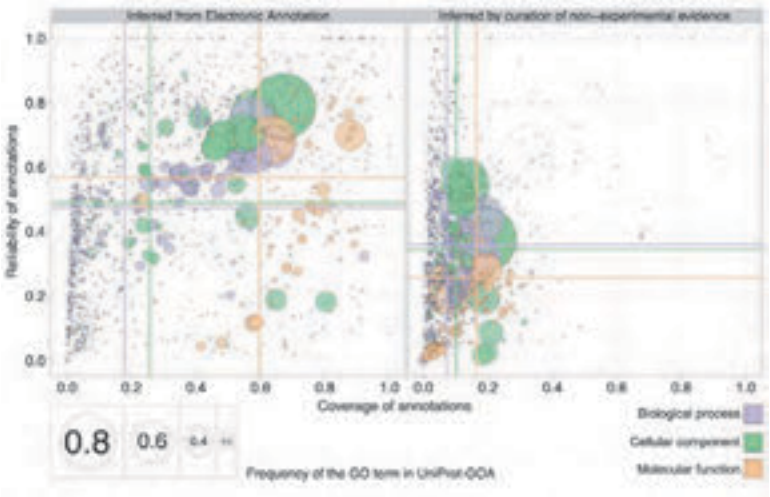
Altenhoff, A.M., *et al.* (2012). Resolving the ortholog conjecture: orthologs tend to be weakly, but significantly, more similar in function than paralogs. *PLoS Comp. Biol.*, 8, e1002514

Jordan, G. & Goldman, N. (2012). The effects of alignment error and alignment filtering on the sitewise detection of positive selection. *Mol. Biol. Evol*., 29, 1125–39

Löytynoja, A., Vilella, A.J., & Goldman, N. (2012). Accurate extension of multiple sequence alignments using a phylogeny-aware graph algorithm. *Bioinformatics*, 28, 1684–91

Future projects and goals

The group remains dedicated to using mathematical modelling, statistics and computing to enable biologists to draw as much scientific value as possible from modern molecular sequence data. We shall continue to concentrate on linked areas that draw on our expertise in phylogenetics, genomics and next-generation sequencing. Basic to all our work are the fundamentals of phylogenetic analysis, where we are investigating the use of non-reversible models of sequence substitutions and problems such as 'long branch attraction'. We remain committed to keeping abreast of evolving NGS technologies and exploiting them for new experiments. We will continue to develop novel data analysis methods, for example how to detect and represent the discordant evolutionary histories of different genes in an organism's genome. We will also continue our work to make the most advanced evolutionary multiple sequence alignment methods available to the broadest possible range of researchers. We are beginning to look to medical applications of NGS and phylogenetics as a source of inspiring collaborations, and hope to start to bring molecular evolution into a clinical setting where it may soon be applicable in 'real time' to help inform doctors' decisions and treatments.



Quality of electronic and curated annotations on a common set of GO terms Electronic annotations are more reliable than generally believed; their reliability (y-axis) rivals that of curated annotations. They also have greater coverage (x-axis). Terms associated with the three ontologies are drawn in different colours, with the area of the disc reflecting the frequency of the GO term in the UniProt-GOA release. The coloured lines correspond to the mean values for the respective ontologies and axes

Computational and evolutionary genomics

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

John Marioni

PhD 2008, University of Cambridge.

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

At EMBL-EBI since 2010.

Selected references

Kim J.K. & Marioni J.C. (2013). Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data. *Genome Biol.*, 14, R7

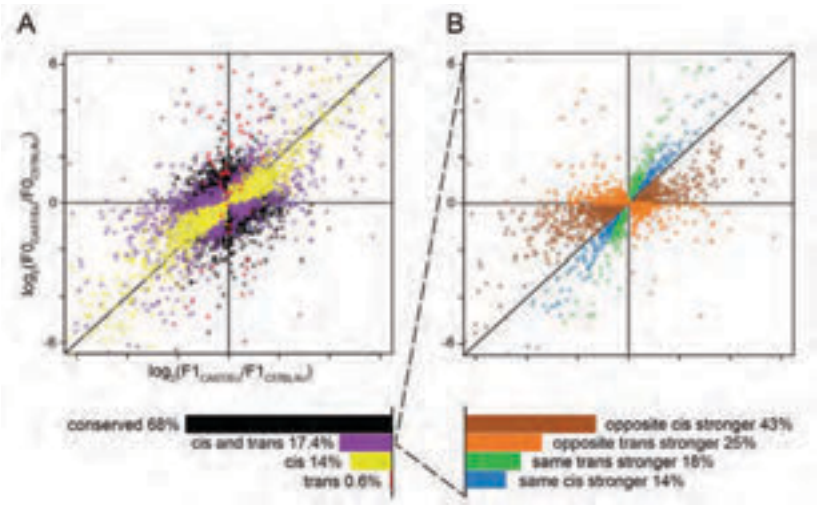
Goncalves, A., *et al*. (2012). Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. *Genome Res.*, 22, 2376-84

Fonseca, N.A. *et al*. (2012). Tools for mapping high-throughput sequencing data. *Bioinformatics*, 28, 3169-77

Previous and current research

Gene expression levels play a critical role in evolution, developmental processes and disease progression. Variability in the transcriptional landscape can help explain phenotypic differences both between and within species. For example, differential expression of the Tan gene between North American *Drosophila* species underlies divergence of pigmentation, while variation in expression levels of the MMP3 gene within the human population alters both vascular tissue remodelling and risk of developing atherosclerosis. As a result, identifying and characterising the regulatory mechanisms responsible for changes in gene expression is critically important. Recently, the advent of next-generation sequencing technology has revolutionised our ability to do this. By facilitating the generation of unbiased, high-resolution maps of genomes, transcriptomes and regulatory features such as transcription factor binding sites, these new experimental techniques have given rise to a detailed view of gene expression regulation in both model and, importantly, non-model organisms.

To make the most of these technological developments, it is essential to develop effective statistical and computational methods for analysing the vast amounts of data generated. Only by harnessing experimental and computational biology will we be able to truly understand complex biological processes such as gene regulation. With this in mind, my group focuses on the development of computational methods for interrogating high-throughput genomics data. Our work focuses primarily on modelling variation in gene expression levels in different contexts: between individual cells from the same tissue; across different samples taken from the same tumour; and at the population level where a single, large sample of cells is taken from the organism and tissue of interest. We apply these methods to a range of biological questions, from studying the regulation of gene expression levels in a mammalian system to the development of the brain in a marine annelid. In all of these projects we collaborate with outstanding experimental groups, both within and outside EMBL. Together, we frame biological questions of interest, design studies, and analyse and interpret the data generated.



Classifying genes by their regulatory function. We used RNA-seq data generated from F0 mice and their F1 hybrids to classify genes into sets depending upon their regulatory mechanism (Goncalves et al., 2012)

Future projects and goals

We will work with our experimental collaborators to apply our 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.

ChEMBL: a database of bioactive drug-like small molecules

John P. Overington

PhD 1991, Birkbeck College, London.

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

At EMBL-EBI since 2008.

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

Previous and current research

Although great progress has been made in developing medicines, synthetic small molecule and natural product-derived drugs still form the majority of novel life-saving drugs. The ChEMBL database stores curated chemical structures and abstracted quantitative bioactivity data alongside calculated molecular properties. The majority of ChEMBL data is derived by manual abstraction and curation from the primary scientific literature, and thereby covers a significant fraction of the structure–activity relationship (SAR) data for the discovery of modern drugs. Our research interests focus on data-mining ChEMBL for data that can be applied to drug-discovery challenges. We typically work with collaborators on integration with other data classes (e.g. toxicology data) or disease/patient datasets.

One current project involves the investigation of differential drug response as a function of age and gender: medicines are generally developed to be effective and safe in adult populations, and only a small subset of drugs has been formally studied for efficacy and safety in paediatric and geriatric populations. However, it is not easy to extrapolate doses, therapeutic indications, safety and efficacy. Lack of understanding of how factors such as weight, fat fraction, and metabolic state influence drug absorption, distribution, efficacy and elimination, especially at a molecular level, has been a hurdle in the transfer of medicines from one population to another. We try to identify age-related molecular differences in human populations, which could increase understanding of some of the differential responses to drugs reported in the literature. It could also help us to identify how these changes translate to species commonly used in drug development, in order to establish the best age-driven animal models.

Another project involves the characterisation of allosteric modulators within ChEMBL. These are compounds of significant interest to medicinal chemists because their mode of action keeps intact normal physiological regulation mechanisms, they allow modulation of proteins not easily accessible to orthosteric modulation (e.g. peptide binding proteins), and they are specific ligands that possess a self limiting on-target activity. Yet these qualities mean that they are very diverse, with an equally diverse array of modes of action. While several allosteric drugs are on the market and a large number of allosteric compounds are studied in preclinical research, there is no definitive course of action to discover a novel allosteric modulator of a pathologically relevant drug target. Using the ChEMBL database and targeted annotation, we try to capture what makes compounds allosteric modulators, and connect these observations to the targets that the compounds modulate, with the goal of developing predictive models that can quantitatively predict the likelihood of compounds modulating a certain target. This opens the door for elucidation of mode of action, focussed libraries, and screening prioritisation.

Future projects and goals

In 2013, we will focus on translational and safety biology and build our community around open data for neglected diseases. We aim to complete linkage to targets, deepen the data model to deal with post-translational modifications and macromolecular assemblies, and integrate population-variation data. We will enhance the delivery of ChEMBL via REST web services with an RDF triple-store representation.

Selected references

Chambers, J., *et al*. (2013). UniChem: a unified chemical structure cross-referencing and identifier tracking system. *J. Cheminform*., 5, 3

Gaulton, A., *et al*. (2012). ChEMBL: A large-scale bioactivity database for chemical biology and drug discovery. *Nucleic Acids Res*., 40, D1100-7

Hersey, A., Senger, S., & Overington, J.P. (2012). Open data for drug discovery – learning from the biology community. *Future Med. Chem.*, 5, 1865-7

Krüger, F.A., Rostom, R., & Overington, J.P. (2012). Mapping small molecule binding data to structural domains. *BMC Bioinform.*, 13, S11

ChEMBL-og drug monographs on the ChEMBL web interface

Systems biomedicine

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.

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.

Selected references

Pacini, C., *et al*. (2013). DvD: An R/Cytoscape pipeline for drug repurposing using public repositories of gene expression data. *Bioinformatics*, 29, 132-4

Iorio, F., *et al*. (2012). Transcriptional data: a new gateway to drug repositioning? *Drug Discov. Today*, 18, 350-7

Eduati, F., *et al*. (2012). Integrating literature-constrained and data-driven inference of signalling networks. *Bioinformatics*, 28, 2311

Terfve, C., *et al*. (2012). CellNOptR: a flexible toolkit to train protein signaling networks to data using multiple logic formalisms. *BMC Syst. Biol*., 6, 133

Previous and current research

The goal of our group is to acquire a functional understanding of signalling networks and their deregulation in disease, and to apply this knowledge to novel therapeutics. Our research is hypothesis-driven and tailored towards producing mathematical models that integrate diverse data sources. Because of this, we collaborate closely with experimental groups.

Our models integrate diverse types of data (from genomic to biochemical) with various sources of prior knowledge, with an emphasis on providing both predictive power of new experiments and insight on the functioning of the signalling network. We combine statistical methods with models describing the mechanisms of signal transduction either as logical or physico-chemical systems. Towards this end, 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.

Productive integration of data and computation requires an effective workflow that pulls together all the steps that link experiments to mathematical models and analysis. We therefore strive to develop tools that facilitate this process and incorporate public standards. We are also involved in a community effort to advance the inference of mathematical models of cellular networks: DREAM (Dialogue for Reverse Engineering Assessments and Methods).

Future projects and goals

We will continue to develop methods and tools to understand signal transduction in human cells, as well as its potential to yield insights of medical relevance. Our main focus will be on modelling signalling networks using phosphoproteomics 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 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?

Screenshot in Cytoscape of DvD, a tool for drug repurposing using public repositories of gene expression data.

Statistical genomics and systems genetics

Oliver Stegle

PhD 2009, University of Cambridge.
Postdoctoral fellow, Max Planck Institutes Tübingen.
Research group leader at EMBL-EBI since 2012.

The Stegle group develops computational and statistical methods to study the genotype-phenotype map on a genome-wide scale.

Previous and current research

Our interest lies in computational approaches to unravel the genotype–phenotype map on a genome-wide scale. How do genetic background and environment jointly shape phenotypic traits or causes diseases? How are genetic and external factors integrated at different molecular layers, such as transcription and translation?

To answer these pertinent biological questions, our work has focused on the development and use of statistical methodology to dissect the causes of molecular variation. We have shown how comprehensive modelling can greatly improve the statistical power to find genetic associations with gene expression levels, and provide for an enhanced interpretation of the interplay between genetic variation, transcriptional regulation and molecular traits. We are addressing these methodological questions in the context of close collaborations with experimental groups, where we apply our statistical tools to study molecular traits in model organisms, plant systems and biomedical applications.

Future projects and goals

We will continue to devise statistical methods to model and analyse data from high-throughput genetic and molecular profiling experiments. A major goal will be to derive techniques to make sense of molecular studies that combine profiling data at multiple molecular layers. To this end, we plan to utilise systems biology concepts to integrate genetic variation data, epigenetic marks, transcriptional profiling, and prior knowledge about regulatory relationships.

To deduce novel regulatory links from observational omics data, we will work on causal inference methodology. The uprise of specific molecular measurements, such as transcriptional information over a time course or single cell profiling, facilitates detailed modelling to testify causal relationships instead of mere correlation. In collaboration with experimental partners, we aim to utilise these methods to identify functional targets for molecular intervention that can be used for personalised treatment.

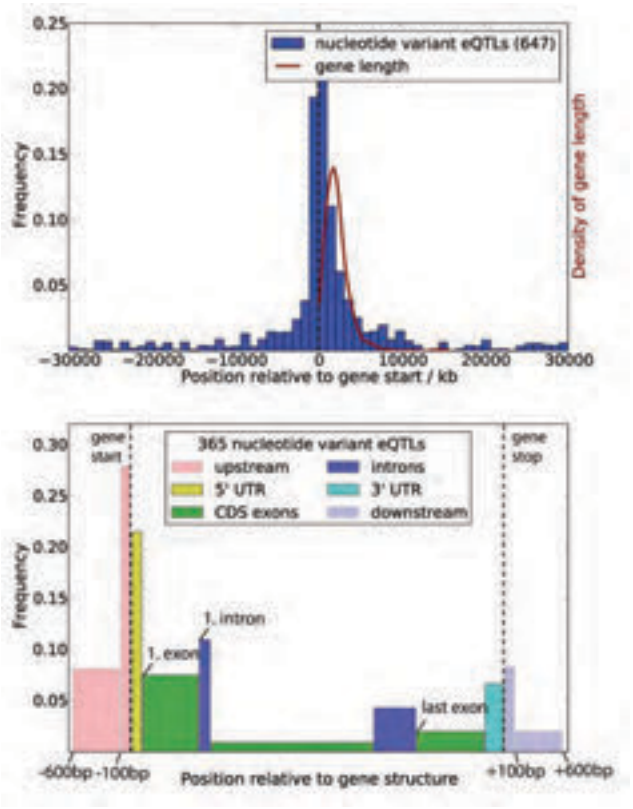
Selected references

Stegle, O., *et al*. (2012). Using Probabilistic Estimation of Expression Residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc*., 7, 500-7

Gan, X., *et al*. (2011). Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature,* 477, 319-423

Parts, L., *et al*. (2011). Joint genetic analysis of gene expression data with inferred cellular phenotypes. *PLoS Genet*. 7, e1001276

Stegle, O., *et al*. (2011). Efficient inference in matrix-variate Gaussian models with iid observation noise. *NIPS,* 17



Statistical inference pinpoints likely causal genetic variants with an effect on gene expression levels in A. thaliana *(top, adapted from Gan* et al.*, 2012). Genome reconstruction and gene re-annotation allow for mapping these associations to variant categories in the vicinity of coding genes (bottom)*

Small molecule metabolism in biological systems

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

Alcántara, R., *et al*. (2012) The EBI Enzyme Portal. *Nucleic Acids Res.*, 41, D773-80

Jayaseelan, K.V., *et al*. (2012) Natural product-likeness score revisited: an open-source, open-data implementation. *BMC Bioinformatics*, 13, 106

Pavelin, K., *et al*. (2012) Bioinformatics meets user-centred design: a perspective. *PLoS Comp. Biol.*, 8, e1002554

Steinbeck, C., *et al*. (2012) MetaboLights: towards a new COSMOS of metabolomics data management. *Metabolomics*, 8, 757-60

Christoph Steinbeck

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

Postdoctoral research at Tufts University, Boston, 1996-1997.

Group leader, MPI 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 research is dedicated to the elucidation of metabolomes, Computer-Assisted Structure Elucidation (CASE), the reconstruction of metabolic networks, biomedical and biochemical ontologies and algorithm development in chem- and bioinformatics.

Part of our research comprises the development and implementation of methods to analyse spectroscopic data in metabolomics. The chemical diversity of the metabolome and a lack of accepted reporting standards currently make analysis challenging and time-consuming. Typical mass spectrometry-based studies, for instance, generate complex data where the signals of interest are obscured by systematic and random noise. Proper data preprocessing and consequent peak detection and extraction is essential for compound identification.

Our group explores ways of producing chemical unification of existing metabolism databases; text mining of small molecules, proteins, tissues/cell types, and organisms; and chemical enumeration of generic reactions and lipids. These approaches allow us to provide species-specific molecule catalogues that make it easier for researchers in metabolomics to identify detected small molecules.

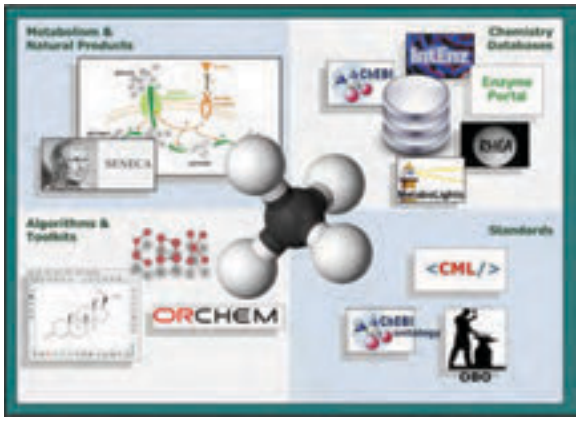
CASE methods developed by us provide a means to determine the structure of metabolites by stochastic screening of large candidate spaces based on spectroscopic methods. Our so-called SENECA system is based on a stochastic structure generator, which is guided by a spectroscopy-based scoring function. Simulated Annealing and Evolutionary Algorithms are at the core of the structure generation process, allowing us to explore the structural space of isomers.

Another of our activities involves employing machine learning methods to correlate graph-based molecular descriptors with database knowledge. We use the resulting prediction engines as judges in our SENECA scoring function. In order to narrow down our search space during the structure-determination process, we employ Natural Product (NP) likeness as a filter.

Future projects and goals

Our central theme of research is efficient methods and algorithms for the assembly, analysis and dissemination of information on small molecules of relevance for biological systems. This includes information about primary and secondary metabolites, and also on xenobiotics and other molecules of relevance, such as epitopes. To this end, we will continue our work in various related areas of ontology development, research on the computational representation of related data, inference of metabolomes from all types of available information, processing of metabolic and metabolomics information, and reconstruction of metabolic networks. We select these projects with an emphasis on applicability in our service foci. There, our focus is on the extension of the ChEBI database towards greater usability for metabolism and natural products research, improving the curation efficiency (which requires work on curation tools and narrow AI in the area of on-the-fly information retrieval) and the extension and establishment of our metabolomics database, MetaboLights.

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



Gene expression regulation and protein complex assembly

Sarah Teichmann

PhD 2000, University of Cambridge and MRC Laboratory of Molecular Biology.

MRC Career Track Programme Leader, MRC Laboratory of Molecular Biology, 2001-5 and MRC Programme Leader, 2006-12.

Fellow and Director of Studies, Trinity College, since 2005.

Group leader at EMBL-EBI and Sanger Institute since 2013.

Combining computational and wet lab approaches, the Teichmann group aims to understand general principles of gene expression and protein complex assembly.

Previous and current research

Our group seeks to elucidate general principles of gene expression and protein complex assembly. We study protein complexes in terms of their 3D structure, structural evolution and the principles underlying protein-complex formation and organisation. We also explore the regulation of gene expression during switches in cell state, and use mouse T-helper cells as a model of cell differentiation. We combine computational and wet-lab approaches at both EMBL-EBI and the Wellcome Trust Sanger Institute.

The wealth of genome-scale data now available for sequences, structures and interactions provides an unprecedented opportunity to investigate systematically principles of gene and protein interactions. We focus on the evolution and dynamics of regulatory and physical interaction networks, combining computational and mathematical approaches with genome-wide and gene/protein experiments. Our two main areas are transcription factors and the regulation of gene expression; and physical protein-protein interactions and protein complexes.

Differences in genes and their spatio-temporal expression patterns determine the physiology of an organism: its development, differentiation and behaviour. Transcription factors regulate this process by decoding DNA elements and binding to DNA in a sequence-specific manner. Our group has developed a prediction pipeline, **transcriptionfactor.org**, that identifies repertoires of transcription factors in genomes.

We are very interested in elucidating transcriptional regulatory networks that orchestrate T-helper-cell differentiation and plasticity. Using the T helper cell system, we explore the hierarchy and kinetics of molecular events that contribute to changes in gene expression, and whether the kinetics of these interactions are graded or switch-like.

Our group also investigates the principles that govern the folding and assembly of protein complexes. Using the informative power of genomic, proteomic and structural data, we capture the critical changes in sequence and structure that distinguish protein-complex formation from the sea of functionally neutral changes. The **3DComplex.org** database is a research tool for our work in this area. Our *in silico*, phylogeny-based methods predict critical ancestral mutations involved in changing protein complexes, and we test these using wet-lab biophysical and biochemical techniques.

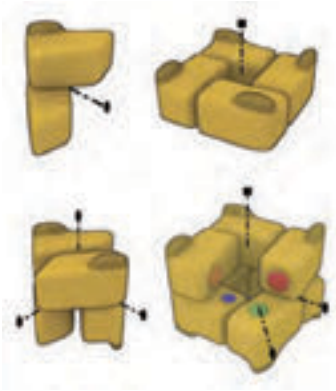
Selected references

Charoensawan, V., Janga, S.C., Bulyk, M.L., Babu, M.M., & Teichmann, S.A. (2012). DNA sequence preferences of transcriptional activators correlate more strongly than repressors with nucleosomes. *Mol. Cell*, 47, 183-92

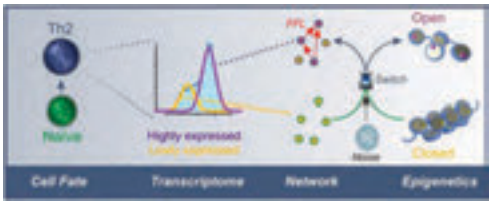
Perica, T., Chothia, C. & Teichmann, S.A. (2012). Evolution of oligomeric state through geometric coupling of interfaces. *Proc. Natl. Acad. Sci. USA*, 109, 8127-32

Hebenstreit, D., *et al*. (2011). RNA-sequencing reveals two major abundance classes of gene expression levels in metazoan cells. *Mol. Sys. Biol.*, 7, 497

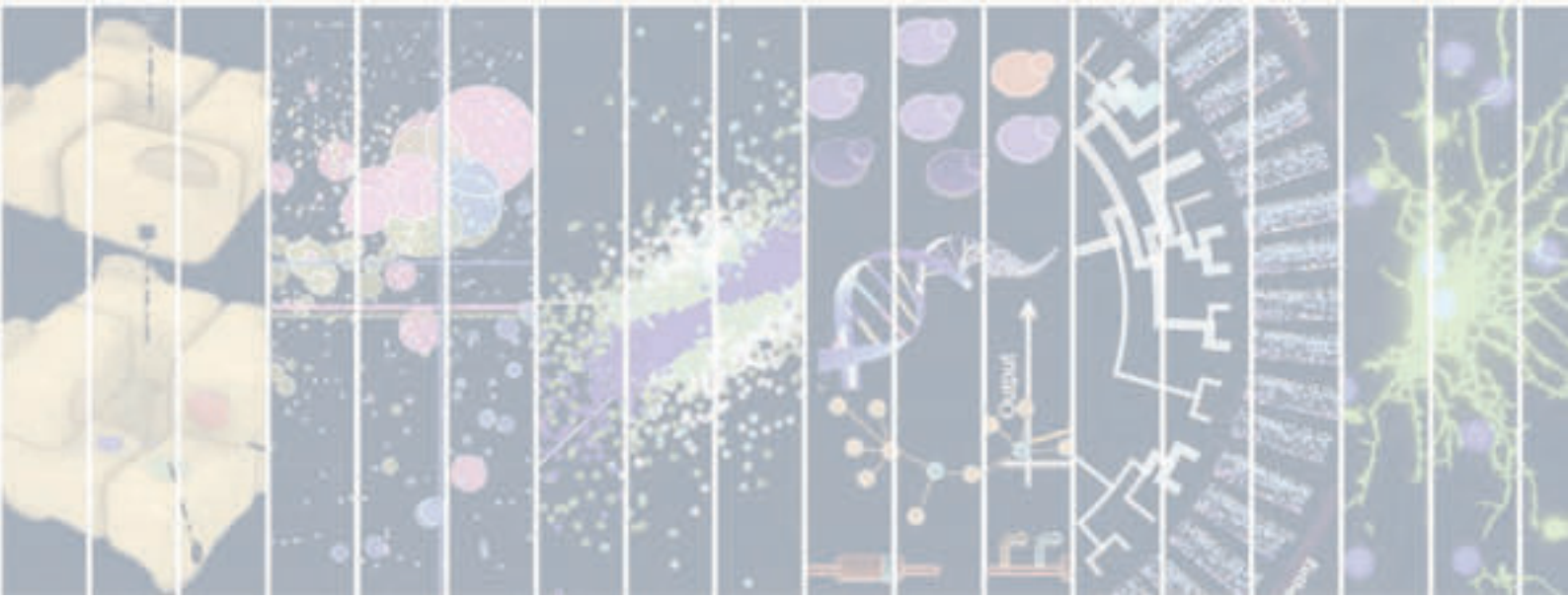
Teichmann, S.A. & Babu, M.M. (2004). Gene regulatory network growth by duplication. *Nature Genet.*, 36, 492-6



Above: Interactions between proteins are central to molecular biology, and determine protein structure and function. By analysing the protein complexes of known 3D structures, we discover patterns and principles of protein interactions, their assembly and evolution



Right: Transcriptional regulatory networks control cell fate, as in this example of CD4+ T cells, which play key roles in orchestrating adaptive immune responses. Stochastic gene expression and plasticity of cell types contrast the apparent need for precise control of mRNA levels.



Output

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 97) 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 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 96) 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-EBI 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

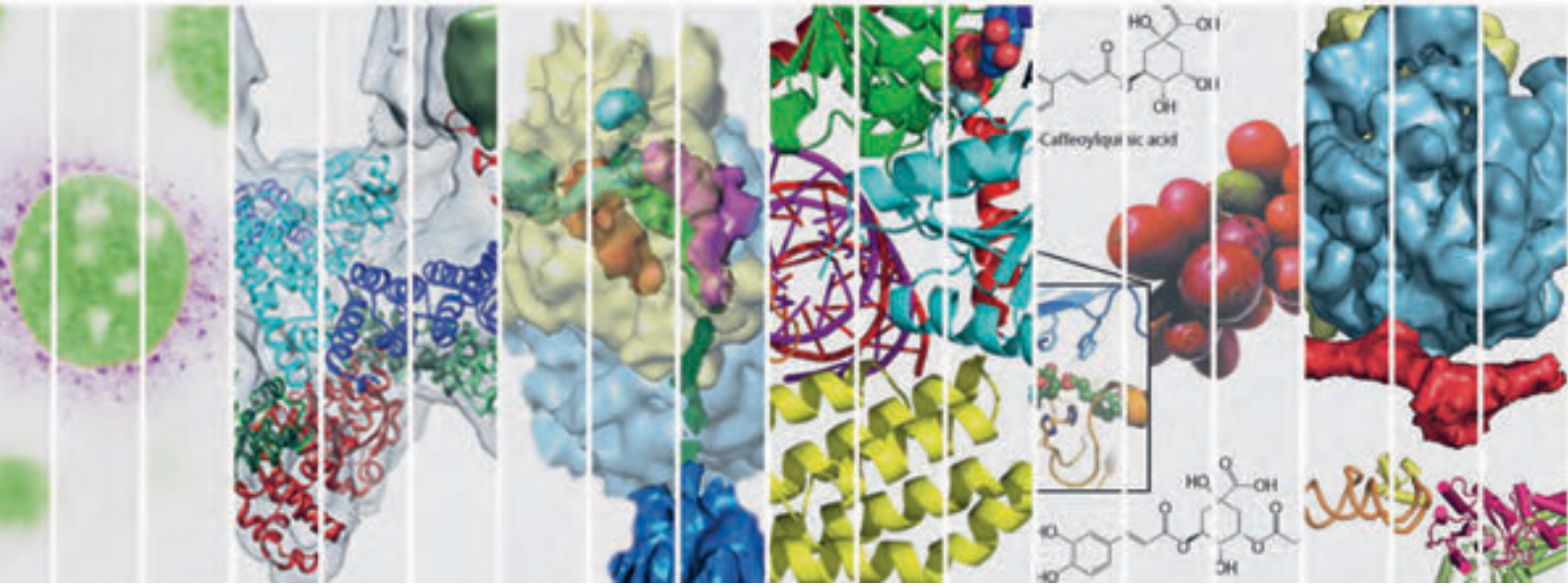
Sameer Valenkar

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.


HO
OH
OH
OH
HO
OH
OH

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 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 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 our involvement in several European-wide 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 further developing 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. One 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 focussed on the viral RNA pattern recognition receptor RIG-I, and another 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.
ERC Advanced Investigator.

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

Previous and current research

Aminoacyl-tRNA synthetases play an essential role in protein synthesis by charging specifically their cognate tRNA(s) with the correct amino acid. Structural information can help understand the catalytic mechanism of the enzymes and their substrate specificity for ATP, cognate amino acid and tRNA. In recent years we have focussed on leucyl-tRNA synthetase which contains a large inserted 'editing' domain able to recognise and hydrolyse mischarged amino acids – a proof-reading activity that 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 collaborated in the elucidation of the mechanisms of a new boron-containing anti-fungal compound (now in Phase III clinical trials) that targets the editing site of leucyl-tRNA synthetase and helped design related anti-bacterial compounds that are active against multi-drug resistant strains, including tuberculosis.

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

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 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: the cap-binding site in PB2 and the endonuclease in PA. These results give some insight into the polymerase mutations required to adapt an avian virus to be able infect humans and permit structure-based antiviral drug design. To pursue this we have co-founded a Vienna-based company called SAVIRA. This work is now being extended to the polymerases of other segmented RNA viruses such as bunyaviruses, which also perform cap-snatching. In collaboration with the Ellenberg (page 18) and Briggs (page 56) 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. Another major project is innate immune pattern recognition receptors such as NOD proteins and RIG-I like helicases. In 2011 we published the first structure-based mechanism of activation of RIG-I and are continuing to study this signalling pathway. Finally we have new projects on epigenetic complexes involving the histone acetylase MOF, such as the X-chromosome dosage compensation or male-specific lethal complex, which in *Drosophila* contains an essential non-coding RNA and the more recently discovered NLS complex (with the Akhtar group, MPI, Freiburg).

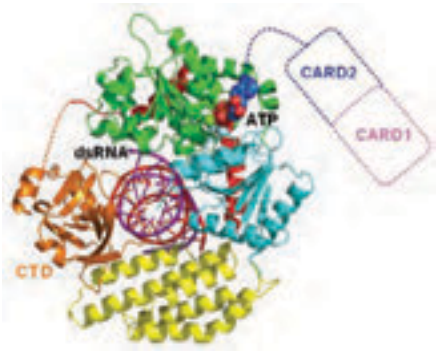
Selected references

Hallacli, E., *et al*. (2012). Msl1-mediated dimerization of the dosage compensation complex is essential for male x-chromosome regulation in *Drosophila*. *Mol. Cell*, 48, 587-600

Palencia, A., *et al*. (2012). Structural dynamics of the aminoacylation and proofreading functional cycle of bacterial leucyl-tRNA synthetase. *Nat. Struct. Mol. Biol.*, 19, 677-84

Avilov, S.V., Moisy, D., Munier, S., Schraidt, O., Naffakh, N. & Cusack, S. (2011). Replication competent influenza A virus encoding split-GFP tagged PB2 polymerase subunit allows live cell imaging of the viral life cycle. *J. Virol.*, 86, 1433-48

Kowalinski, E., *et al*. (2011). Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell*, 147, 423-35



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 since 2005.

Group leader at EMBL Grenoble since 2007.

Previous and current research

Human gene transcription requires the controlled step-wise assembly of the pre-initiation complex (PIC), comprising a large ensemble of proteins and protein complexes including RNA Pol II and the general transcription factors. TFIID is the first general transcription factor to bind to gene promoters, and is a cornerstone of PIC assembly. Understanding of TFIID and its crucial role in transcription regulation is hampered by a lack of detailed knowledge of its molecular architecture, assembly in the cell, and interactions with chromatin and other factors. The paucity and heterogeneity of TFIID in endogenous cells impede its extraction from cells for high-resolution structural and functional studies. Our lab addressed this challenge by creating new technologies for recombinant production of TFIID and other complex protein machines. Notably, our MultiBac system – a modular, baculovirus-based technology specifically designed for eukaryotic multi-protein expression – is now used in many labs worldwide, in areas including structural biology, vaccine development and gene therapy vectors. Recently, we determined the architecture of the 700 kDa heterodecameric human TFIID core complex by hybrid methods, combining MultiBac-based production, cryo-EM, X-ray crystal structures, homology models, and proteomics data. It is thought to represent a central scaffold that nucleates holo-TFIID formation and provided first impressions on how the functional holo-TFIID complex is assembled in the nucleus.

We collaborate with groups from academia and industry for technology development. We are striving to automate labour-intensive steps in the multiprotein complex structure determination process, and have harnessed homologous and site-specific recombination methods for assembling multigene expression plasmids. We have implemented a full robotics setup by developing ACEMBL, a proprietary automated suite for multigene recombineering on our TECAN EvoII platform. By using our technology, we produced numerous large multiprotein assemblies for structural studies, including multicomponent membrane protein complexes and the 1.6 MDa human TFIID holo-complex and expanded our multiprotein expression strategies to prokaryotic and mammalian hosts.

Future projects and goals

We continue to advance our expression technologies to entirely automate and standardise the process of production for eukaryotic gene regulatory multiprotein complexes including the entire human TFIID holocomplex, its various isoforms and other components of the preinitiation complex. In collaboration with the Schaffitzel Team (page 110) 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, aim to understand physiological function, and explore and challenge our 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.

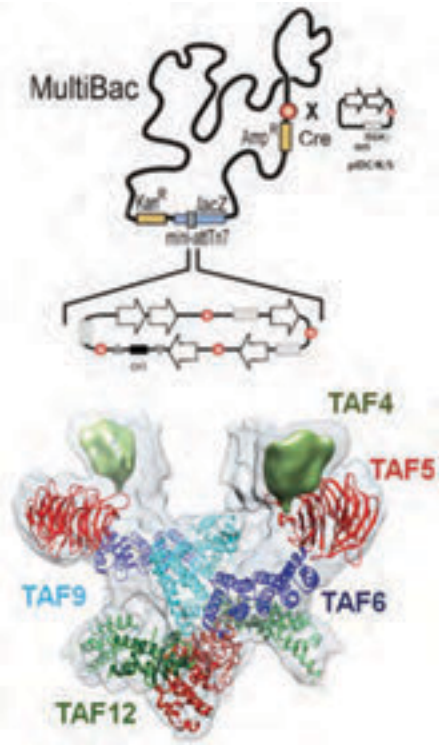
Selected references

Bieniossek, C., *et al*. (2013). The architecture of human general transcription factor TFIID core complex. *Nature*, 493, 699-702

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., *et al*. (2011). Architecture of the Mediator head module. *Nature*, 475, 240-3

Bieniossek, C., *et al*. (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.
Senior scientist since 2011.

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

The core activity of our team is to develop instruments and methods for X-ray scattering experiments in collaboration with the Márquez team (page 106) and with the ESRF Structural Biology group, as well as contributing to the development of the EMBL@PETRA3 beamlines. Benefiting from the support of EMBLEM, our mission is also to make our instruments available to the scientific community worldwide.

In line with the work done over the past decade in micro-crystallography, we have developed, in collaboration with EMBL Hamburg, a vertical Kappa diffractometer able to process micron-sized crystals. The MD3 prototype installed at the EMBL@PETRA3 MX2 beamline exhibits sub-micron stability at sample position. The quality of data obtained from shutterless 4D scans over thin needles has demonstrated the exceptional dynamic performance of the new goniometer. Concluding an EMBL/ESRF collaborative project started in 2008, we have equipped the EMBL@PETRA3 BioSAXS beamline with a fast automated sample changer. At the ESRF, a high-performance liquid chromatography system has been integrated to the BM29 BioSAXS beamline to characterise difficult proteins (part-supported through the EU-funded BioStruct-X programme).

Our current work mainly focuses on the development of an automated crystal harvesting system based on CrystalDirect™, a concept, conceived jointly with the Márquez team, where crystals grown on a thin film in a new vapour diffusion crystallisation plate are recovered by laser photo ablation (see figure). Chemical treatments like cryo-protection can be applied to crystals directly in the plates, before harvesting. CrystalDirect (CD) plates have been produced and a harvester prototype built. Model proteins and a number of new proteins have been crystallised, cryo-protected, harvested and successfully processed at MX beamlines. The CD plates are also ideal for *in-situ* X-ray crystal screening. A specific goniometer head was developed to process CD plates at the EMBL/ESRF BM14 beamline. Data with exceptionally low background was obtained, demonstrating the potential of the CD plates for screening micro-crystals *in-situ*.

Studies are ongoing to develop a compact and precise sample holder for frozen crystals. This new standard should reduce sample handling effort and further facilitate crystal alignment at beamlines, in particular when associated with automated harvesting systems like CD. Lead by the EMBL Grenoble, this collaborative project, called 'NewPin', is part-supported by BioStruct-X.

Future projects and goals

Our short-term plan is to open a crystal harvesting platform and a service for *in-situ* crystal screening. A new harvester with extended crystal treatment capabilities will be developed for projects involving ligands or derivatives. In the context of the ESRF–MASSIF upgrade program, we plan to equip BM14 with industrial robotics to evaluate the integration of CD and to test 'NewPin'. Our long-term ambition is to bridge, via CD, the Grenoble HTX lab and BM14 in a pilot automated MX facility that provides full remote service – from crystallisation to data collection.

Selected references

Cipriani, F., Rower, M., Landret, C., Zander, U., Felisaz, F. & Marquez, J.A. (2012). CrystalDirect: a new method for automated crystal harvesting based on laser-induced photoablation of thin films. *Acta Crystallogr. D Biol. Crystallogr.*, 68, 1393-9

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

Sanchez-Weatherby, J., *et al*. (2009). Improving diffraction by humidity control: a novel device compatible with X-ray beamlines. *Acta Crystallogr. D Biol. Crystallogr.*, 65, 1237-46

Cipriani, F., *et al.* (2006). Automation of sample mounting for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, 62, 1251-9

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.

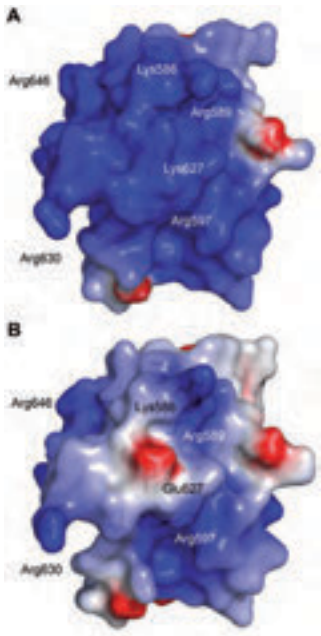
Selected references

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 multi-domain 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, however sometimes 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, Spain.

Postdoctoral research at EMBL.

Staff scientist at EMBL Grenoble since 2003.

Team leader since 2005.
Head of Crystallisation Facility since 2012

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 and one of the major resources of Grenoble's Partnership for Structural Biology, open to scientists working in European academic institutions through the EC-funded P-CUBE and BioStructX projects. The lab has offered services to hundreds of scientists, performing several million experiments. We are involved in the development of data management resources and new crystallisation techniques.

The Crystallisation Information Management System (CRIMS): CRIMS tracks experiments and makes results available to users via the web in real-time, along with all experimental parameters. It has been licensed to 10 other laboratories in Europe, three of them at synchrotron sites. Recently, data mining through CRIMS has allowed us to develop a new method to determine the crystallisation likelihood of a protein sample based on a simple assay measuring thermal stability (Dupeux *et al*., 2011).

Integration of crystallisation and synchrotron data collection facilities: While both highly automated platforms, recovering crystals and mounting them on supports remains a difficult and time-consuming manual process. In collaboration with the Cipriani Team (page 104), we developed Crystal Direct™, an approach that enables full automation of the crystal harvesting process (see figure). Crystals are grown on an ultrathin film in a vapour-diffusion crystallisation plate and recovered through laser-induced photo ablation. Advantages include: elimination of crystal fishing and handling; absence of mechanical stress during mounting; and compatibility with X-ray data collection. The first prototype is now in operation and could benefit many projects.

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 and shown how receptor dimerisation modulates ligand binding affinity leading to differential sensitivities towards the hormone (Dupeux *et al*., 2011). This provides a 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

In collaboration with the McCarthy (page 107) and Cipriani teams we will link the crystallisation screening service with the automated evaluation of crystals *in situ* with X-rays. New interfaces will be added to CRIMS to allow remote operation and we aim (with collaborators) to integrate CRIMS with synchrotron data management systems. We will develop vapour diffusion and microfluidic devices for crystal optimisation experiments. We aim to establish CrystalDirect for routine use in macromolecular crystallography and a series of pilot projects will be selected to establish standard protocols and develop new applications. Prototypes of the new plates will be distributed to other labs as part of the INSTRUCT project, and we will work towards the development of promising new approaches for crystal processing, including for cryo-protection, crystal soaking and crystal freezing. Towards this goal, an advanced version of the CrystalDirect harvester will be designed in collaboration with the Cipriani team.

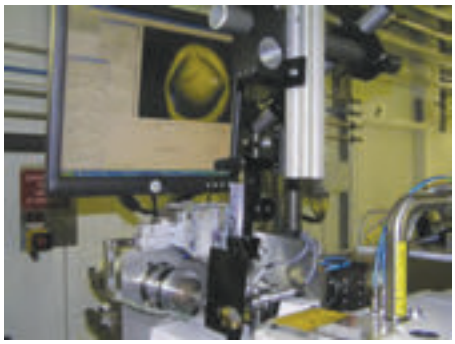
Selected references

Gueneau, E., *et al*. (2013). Structure of the MutLα C-terminal domain reveals how Mlh1 recruits the exonuclease I and contributes to Pms1 endonuclease site. *Nat. Struct. Mol. Biol.*, e-pub ahead of print

Cipriani, F., Rower, M., Landret, C., Zander, U., Felisaz, F. & Marquez, J.A. (2012). CrystalDirect: a new method for automated crystal harvesting based on laser-induced photoablation of thin films. *Acta Crystallogr. D Biol. Crystallogr.*, 68, 1393-9

Dupeux, F., Rower, M., Seroul, G., Blot, D. & Marquez, J.A. (2011). A thermal stability assay can help to estimate the crystallization likelihood of biological samples. *Acta Crystallogr. D Biol. Crystallogr.*, 67, 915-9

Santiago, J., *et al*. (2009). The abscisic acid receptor PYR1 in complex with abscisic acid. *Nature*, 462, 665-8



The CrystalDirect technique could benefit challenging structural biology projects, such as studies of membrane proteins or multi-protein complexes

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

Lallemand, L.A., *et al*. (2012). A structural basis for the biosynthesis of the major chlorogenic acids found in coffee. *Plant Physiol.*, 160, 249-60

de Sanctis, D. & Nanao, M.H. (2012). Segmenting data sets for RIP. *Acta Crystallogr. D Biol. Crystallogr.*, 68, 1152-62

Brockhauser, S., *et al*. (2012). The use of workflows in the design and implementation of complex experiments in macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, 68, 975-84

Griffin, J.L., *et al*. (2012). Near attack conformers dominate beta-phosphoglucomutase complexes where geometry and charge distribution reflect those of substrate. *Proc. Natl. Acad. Sci. USA*, 109, 6910-5

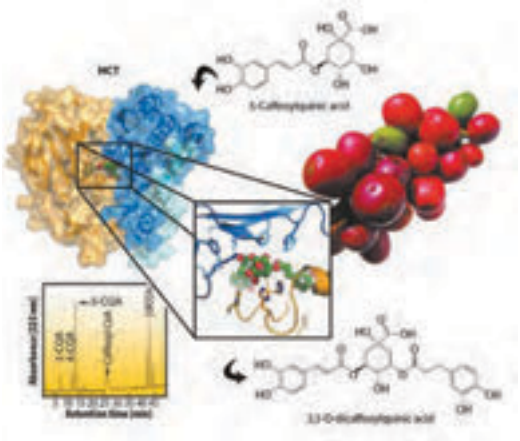


Figure 1: Coffee beans contain large quantities of beneficial compounds such as 5-CQA and 3,5-diCQA that are predominantly synthesised by HQT and HCT respectively. The lower left panel shows a chromatogram illustrating the synthesis of 3,5-diCQA catalysed by a HCT mutant. The central inset shows the best docking result of 3,5-diCQA in HCT

Figure 2: Diffraction Viewer (dViewer) software is currently under development for the visualisation of new generation pixel detector images

Previous and current research

Our team works in close collaboration with the Structural Biology Group of the European Synchrotron Radiation Facility (ESRF) in the design, construction and operation of macromolecular crystallography (MX) and biological X-ray scattering (BioSAXS) beamlines. We are currently responsible for two macromolecular beamlines, ID14-4 and ID23-2, the BioSAXS beamline at BM29, as well as the commissioning of MASSIF1 on ID30. The team also manages the operation of BM14, which is run as a partnership with the ESRF and the Indian government. The structural biology beamlines at the ESRF continue to perform, resulting in the deposition of more than 898 structures in the PDB last year. We work in close collaboration with the Cipriani team (page 104) to develop hardware and we have successfully commissioned a plate holder for the *in situ* screening of crystals on BM14, which will soon be available to external users. We are also developing novel methodologies for advanced sample screening and data collection possibilities. A recent example is the use of workflows in the design and implementation of complex experiments in MX.

We study proteins involved in neuronal development, particularly the Slit-Robo signalling complex, and proteins involved in the synthesis of plant secondary metabolites (figure 1). Meanwhile, the BM14 group is actively involved in structural studies of proteins involved in *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 next phase of the UPBL10 project with our ESRF colleagues, the first user operation of the new MASSIF (Massively Automated Sample Selection Integrated Facility) suite of MX-beamlines on ID30A. 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. On BM14, we will optimise the *in situ* screening of crystals in plates and install a new sample changer design for commissioning. The 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 using Eclipse-RCP technology and as part of BioStruct-X and DAWN collaborations (figure 2). We will further develop the highly automated BioSAXS beamline in collaboration with the ESRF, the Cipriani team and colleagues at EMBL Hamburg, including the full integration of an online HPLC system with additional biophysical characterisation features. We hope that all our combined efforts will push the boundaries of structural biology to better understand the functions of more complex biological systems.

In our laboratory, we will continue research on the Slit-Robo signalling complex by trying to decipher exactly how Slit activates Robo on the cell surface. We will also extend work on phosphoryl transfer into human kinase signal cascades. 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 will be continued.

Structural biology of signal transduction and epigenetic gene regulation

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

Cellular control logic is ultimately embedded in the molecular architecture of the molecular machines that make up the living cell. Many molecular machines, especially complexes involved in cellular signalling, are transient, with a variety of states and a succession of structures. Transcription factors, for example, bind as complexes to enhancers ('enhanceosomes') in a combinatorial and dynamic fashion to regulate expression of genes. Kinases undergo major conformational rearrangements as part of their activation cycle. These signalling components interact with each other and with other molecules in highly structured but complex ways.

Understanding such transient and dynamic complexes of the cellular machinery is one of the most important challenges in biology today. One crucial first step towards characterising such dynamic processes is to determine the molecular architecture of essential components. We have been using the core approaches of structural biology to address the following questions: What is the architecture of signalling complexes that direct innate immune responses? How do these signalling pathways lead to assembly of higher-order transcription factor complexes? How does assembly of such transcription factor complexes ultimately lead to chromatin modification? How does chromatin modification direct nucleosome remodelling and gene regulation? (figure 1).

Future projects and goals

Cellular signalling ultimately results in assembly of transcriptional regulatory complexes that direct chromatin modification, remodelling and gene expression. The enhanceosome has served as a paradigm for understanding signal integration on higher eukaryotic enhancers (figure 2). Assembly of the enhanceosome results in recruitment of enzymes that modify chromatin. We ask how chromatin modification changes the structure of inhibitory nucleosomes and leads to a more permissive chromatin structure for gene expression. We also ask how chromatin modification enzymes read out histone modification patterns and how chromatin recognition and modification are coupled. Answers to some of these questions are likely to contribute to our understanding of epigenetic gene regulation and dysregulation in disease. Finally, we also are interested in understanding key regulators involved in the innate immune response. This is not only of fundamental importance for cellular signalling but also opens up opportunities for pharmacological targeting.

Selected references

Reynoird, N., *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., & 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

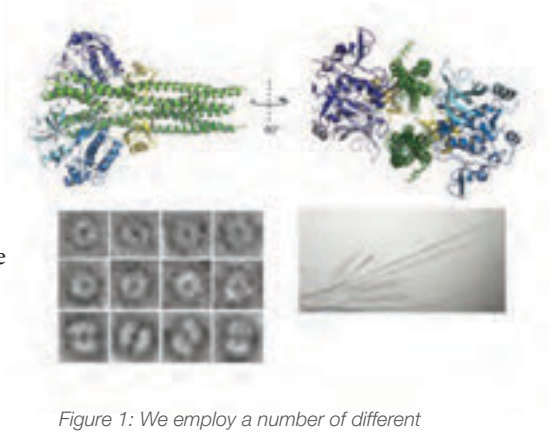


Figure 1: We employ a number of different resolution techniques to visualise the architecture of cellular components

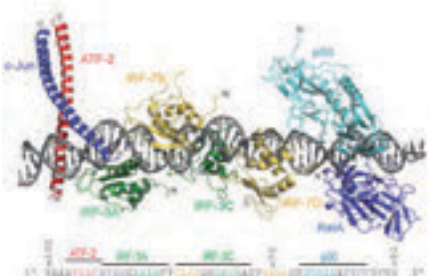


Figure 2: Atomic model of the INF-β enhanceosome

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.

Ramesh Pillai

PhD 2002, University of Bern.

Postdoctoral research at the Friedrich Miescher Institute, Basel.

Group leader at EMBL Grenoble since 2006.
ERC Investigator.

Selected references

Reuter, M., *et al*. (2011). Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature*, 480, 264-7

Simon, B., *et al*. (2011). Recognition of 2'-O-methylated 3'-end of piRNA by the PAZ domain of a Piwi protein. *Structure*, 19, 172-180

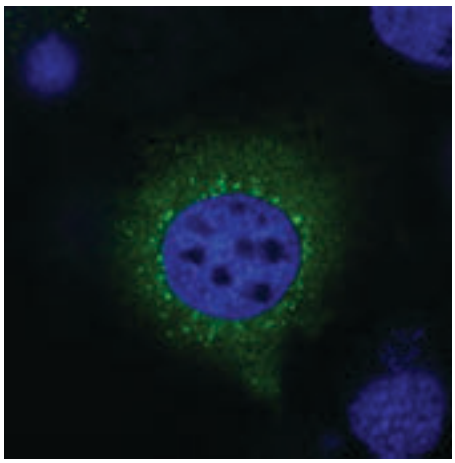
Zheng, K., *et al*. (2010). Mouse MOV10L1 associates with Piwi proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway. *Proc. Natl. Acad. Sci. USA*, 107, 11841-6

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

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 that 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.



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

Future projects and goals

We will continue to analyse additional factors identified in our complex purifications. Another goal is 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.
ERC Investigator.

The Schaffitzel team combines molecular biology, biochemistry and cryo-electron microscopy to study large macromolecular complexes in protein targeting, secretion and membrane protein integration.

Previous and current research

Research in our laboratory combines molecular biology, *in vitro* and *in vivo* biochemistry, and single-particle cryo-electron microscopy (cryo-EM) to study the structure and function of ribosomal complexes. Synthesis and folding of proteins require concerted interactions of the translating ribosome with translation factors, regulatory factors, molecular chaperones, and factors involved in the export of proteins. Structures of translating ribosomes in complex with these factors provide critical insight into the interaction networks, stoichiometry, and molecular mechanisms of these megadalton-size complexes. Using cryo-EM, we can study the multicomponent 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. We have established bacterial and eukaryotic cell-free translation systems for the *in vitro* generation of ribosomes that display homogenous nascent polypeptide chains or have 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 cross-linking/mass spectrometry (collaboration with Juri Rappsilber, TU Berlin) and single-particle cryo-electron microscopy. In collaboration with the Cusack (page 102), Hentze (page 14) and Kulozik (Molecular Medicine Partnership Unit) groups, we study mammalian ribosomal complexes involved in nonsense-mediated mRNA decay. We produce the eukaryotic factors involved using advanced recombinant eukaryotic technologies in collaboration with the Berger group (page 103). Finally, our team collaborates with a number of groups to solve the structures of large macromolecular complexes in transcription, epigenetics and cellular signalling.

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.

Schaffitzel, C., *et al*. (2006). Structure of the *E. coli* signal recognition particle bound to a translating ribosome. *Nature*, 444, 503-6

Mitra, K., *et al*. (2005). Structure of the *E. coli* protein-conducting channel bound to a translating ribosome. *Nature*, 438, 318-24

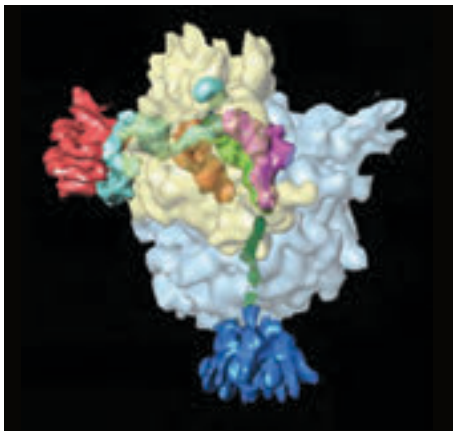


Figure 1: EM reconstruction of the ribosome nascent chain complex and the translocon

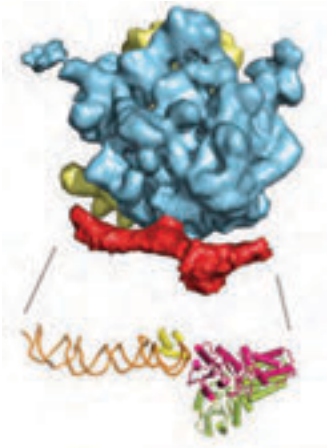
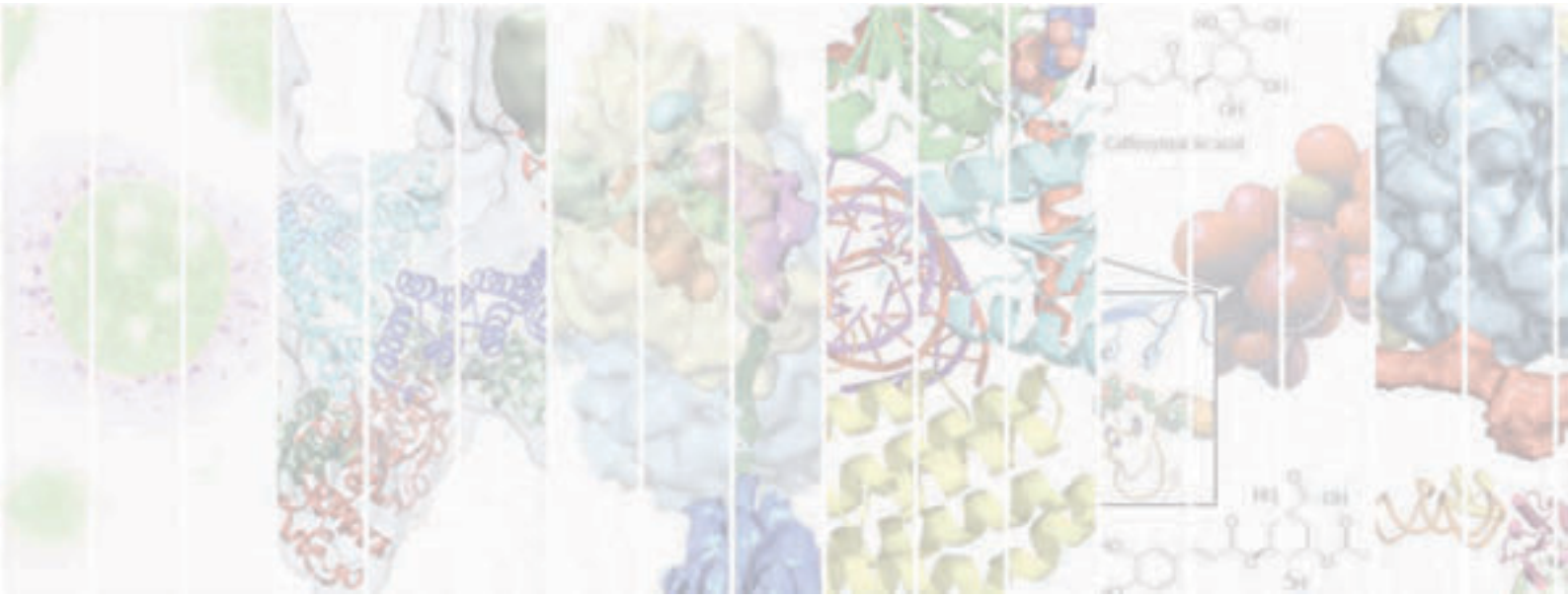
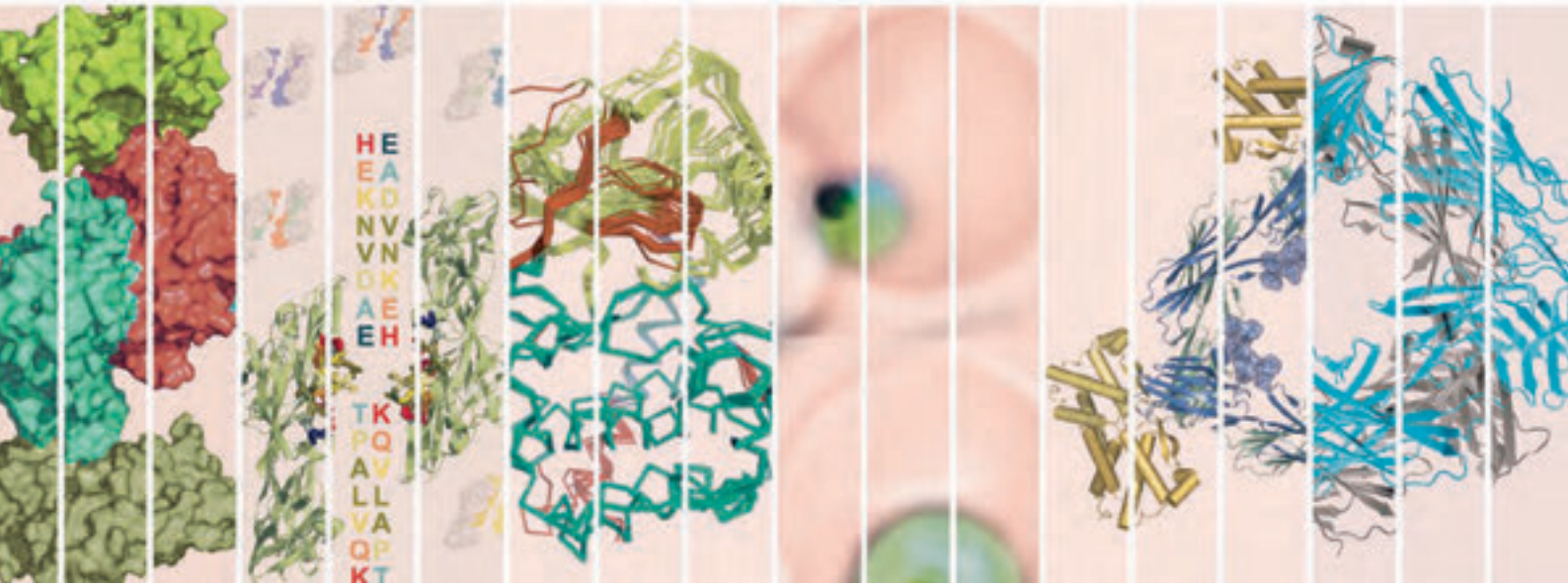


Figure 2: Cryo-EM structure of the ribosome (yellow/blue) bound to the signal recognition particle (SRP) and the SRP receptor (both in red). Below it is an atomic model of SRP (green-yellow/orange) and its receptor (pink)





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 (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 of 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 the goal 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 here 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. 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 the overall architecture of the elastic filament protein myomesin (Pinotsis *et al.*, 2008; Pinotsis *et al.*, 2012). 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 share a common C-terminal autoregulatory domain. 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). More recently, we have unravelled the structure of the apoptotic Death Associated Protein Kinase-1, in the presence of the regulatory scaffold calcium/calmodulin (CaM) (figure 1). This structure provides 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. For the first time, we have unravelled 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 2). Our goal, with support of two national and international research networks, is to provide insight into the overall architecture of the peroxisomal translocon, using a broad range of structural biology and imaging methods, complemented by genetic and cell biology-oriented approaches.

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 and with support of European research network systeMTb, we aim to use systems biology-orientated approaches (such as proteomics, metabolomics, lipidomics and transcriptomics) 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

Fodor, K., Wolf, J., Erdmann, R., Schliebs, W. & Wilmanns, M. (2012). Molecular requirements for peroxisomal targeting of alanine-glyoxylate aminotransferase as an essential determinant in primary hyperoxaluria type 1. *PLoS Biol.*, 10, e1001309

Pinotsis, N., *et al*. (2012). Superhelical architecture of the myosin filament-linking protein myomesin with unusual elastic properties. *PLoS Biol*., 10, e1001261

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

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

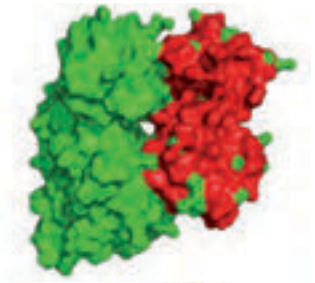


Figure 1: DAPK-CaM complex. Colour codes: DAPK, green; CaM, red. (de Diego et al., *2010)*

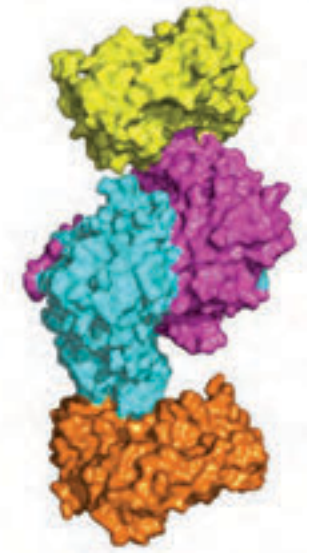


Figure 2: AGT-(Pex5p)2-AGT complex. Colour codes: AGT, yellow, orange; Pex5p, cyan, magenta. (Fodor et al., *2012)*

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

Siewert, F., *et al*. (2012). High angular resolution slope measuring deflectometry for the characterization of ultra-precise reflective x-ray optics, *Meas. Sci. Technol.* 23, 074015

Schultke, E., *et al*. (2011). Dual energy CT at the synchrotron: a piglet model for neurovascular research. *Eur. J. Radiol.*, 79, 323-7

Keyrilainen, J., *et al*. (2008). Toward high-contrast breast CT at low radiation dose. *Radiology*, 249, 321-7

Roessle, M.W., *et al*. (2007). Upgrade of the small-angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg. *Journal of Applied Crystallography*, 40, S190-4

Previous and current research

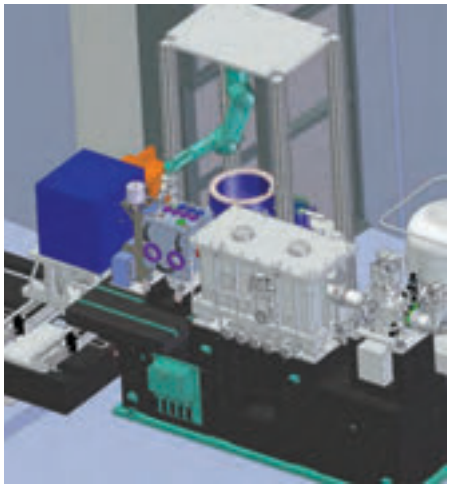
EMBL has designed, built and operates three beamlines for structural biology at the PETRA III synchrotron radiation source on the DESY campus in Hamburg. Beamline facilities are dedicated to the leading techniques for X-ray-based structural research of biological samples: small angle X-ray scattering and macromolecular crystallography. Our team provides expertise in X-ray optics, precision mechanical engineering, robotics, control software and electronics and is in charge of the X-ray optical elements, experimental endstations, vacuum system, cryogenic system, control system, data acquisition system, technical infrastructure and parts of the civil engineering.

Major projects in 2012: All three beamlines reached regular user operation. In preparation for this, it was necessary to install and commission adaptive focusing optics at all beamlines, develop a multi-degree-of-freedom heavy-duty detector stage for the large-area pixel detectors at the MX beamlines (P13 instrument installed) and to develop and install white beam X-ray monitor systems for all beamlines. A cryogenic supply system for the beamline endstations has been installed and is in operation and a two-floor control hutch area with computing infrastructure was set-up in order to allow beamline control and users to perform experiments in a comfortable manner.

Multilayer optics: An ongoing in-house development is the construction of a double multilayer monochromator that is intended to boost the flux density at the P12 SAXS beamline – this enables time-resolved solution scattering experiments in the microsecond range. This instrument has been further developed from a prototype installed at a beamline of the former DORIS storage ring that can preserve the highly increased coherence of the radiation delivered by the PETRA III synchrotron. We are also working on the development of downstream experimental instrumentation with the level of synchronisation necessary for dynamic measurements.

Automation: For all beamline facilities, it is important to develop capabilities such as automatic operation and remote access. In this context, a robotic sample mounting system named MARVIN is being developed for the MX applications (see figure). This is characterised by high sample storage capacity, high sample mounting speed and flexibility. It is integrated, like all beamline elements, into a software-based control system which allows for a heterogeneous control environment and provides distributed access. The prototype has been in user operation on the BW7b beamline at DORIS and adapted versions are undergoing installation on the MX beamlines at PETRA.

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



Future projects and goals

- Positional and intensity feedback and (active) vibrational stabilisation of the monochromatising optics.
- Automatic tuning of adaptive focusing mirror optics to vary beam size/shape.
- Integration of beamline elements into a global instrument protection system.
- Further automation of alignment and data acquisition and integration with sample preparation/crystallisation.
- Improved sample observation and positioning at or beyond the optical resolution limit for microcrystallographic investigations.
- Exploring preparative or bridging developments for the X-ray free electron laser with state-of-the-art synchrotron beamlines.

Integrative modelling for structural biology

Victor S. Lamzin

PhD 1987, Moscow State University.

Scientist, Inst. Biochemistry, Russian Academy of Sciences, Moscow, until 1991.

Postdoctoral researcher at EMBL Hamburg until 1995. Staff scientist until 1997.
Group leader and Deputy Head of outstation since 1997.

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

To fully understand the function of biological systems, accurate structures of their components – DNA, RNA, proteins, macromolecular complexes and assemblies – are required. We apply and develop cutting-edge computational methods and experimental approaches for sample quality control, experimentation and data interpretation in macromolecular crystallography that may also have potential use in electron microscopy and X-ray free-electron lasers.

Previous and current research

Methods for biological structure determination: 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 and Lamzin, 2012). We exploit inherent properties of macromolecular structures (Wiegels & Lamzin, 2012) and integrate additional information derived from biological databases. Pattern-recognition methods are the basis of the group's main focus, the ARP/wARP software project (Langer *et al.*, 2008). We have developed a molecular viewer – the ArpNavigator – which enables user control of the model-building process and provides easy access to a wide range of methods for quality assessment and model completion.

Structure-based drug design: We make use of various novel algorithms and, through their combination (Langer *et al.*, 2012), develop new tools for drug discovery. The ViCi software, available for use through an online web server (**www.embl-hamburg.de/ViCi**), allows for the *in silico* screening of known ligands to provide new leads for 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 focused on the probing of bacterial antibiotic resistance.

Biological imaging with Free-Electron Lasers (FEL): Breathtaking results from initial diffractive imaging experiments using coherent FEL radiation (Siebert *et al.*, 2011) show the potential for imaging cellular organelles and understanding dynamics of complex formation. In order to exploit the numerous 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 (Mancouso *et al.*, 2012).

Other targets of biomedical interest: We integrate X-ray crystallography, lower resolution imaging, biochemistry and biophysics to investigate targets of biomedical interest. These include the nuclear pore complex and hydrophobins (Kallio *et al.*, 2011), which may be used in drug delivery to solubilise hydrophobic pharmaceuticals, the pathway of amyloid fibril formation via class I hydrophobins, and fragments of human gelsolin.

Future projects and goals

The group will continue to focus on crystallographic software development driven by general academic interest, provision of state-of-the-art beamline facilities at PETRA III in Hamburg and by the potential use of such developments in projects of medical or biotechnological importance. Together with our international collaborators, we will undertake novel pilot projects aimed at interpretation of structural data obtained from various sources.

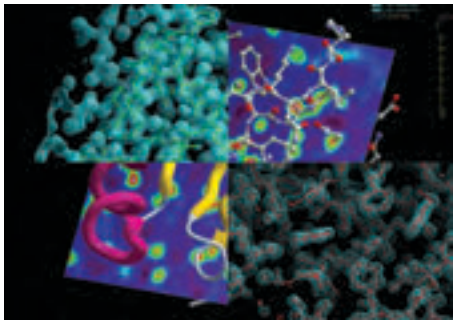
Selected references

Wiegels, T. & Lamzin V.S. (2012). Use of noncrystallographic symmetry for automated model building at medium to low resolution. *Acta Crystallogr. D Biol. Crystallogr.*, 68, 446-53

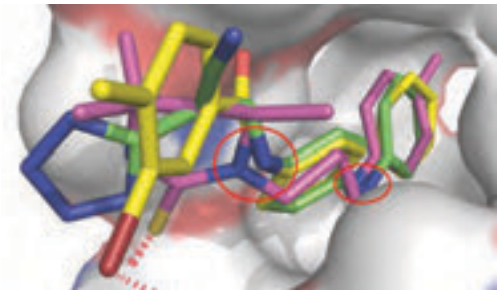
Langer, G.G., Evrard, G.X., Carolan, C.G., & Lamzin, V.S. (2012). Fragmentation-tree density representation for crystallographic modelling of bound ligands. *J. Mol. Biol.*, 419, 211-22

Mancuso, A.P., *et al*. (2012). Internal structure of an intact *Convallaria majalis* pollen grain observed with X-ray Fresnel coherent diffractive imaging. *Opt. Express*, 20, 26778-85

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



A selection of model-viewing options in ArpNavigator. Clockwise from the top left: a stick view in solid electron density; a ball-and-stick representation in planar density; a skeleton view of the electron density shown as a mesh; and the protein in cartoon view in planar density



A known inhibitor in green, aligned with two hits from the ViCi software, in the binding pocket of beta lactamase. Important interactions maintained are highlighted with red circles and those created for exploration with full red spots

Structural biology of cell surface receptors

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

Rob Meijers

PhD 2001, EMBL Hamburg/ University of Amsterdam.

Postdoctoral research at the Dana Farber Cancer Institute, Boston.

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

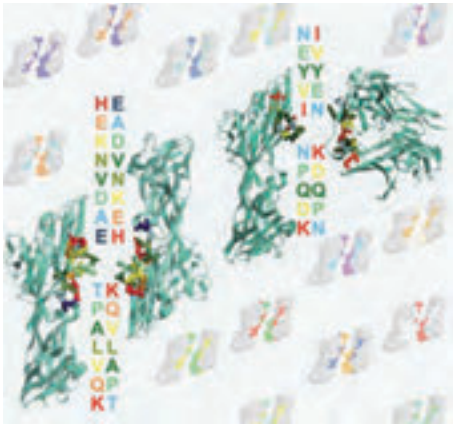
Group leader at EMBL Hamburg since 2009.

Selected references

Meijers, R., *et al*. (2007). Structural basis of Dscam isoform specificity. *Nature*, 449, 487-91

Meijers, R., *et al*. (2005). Crystal structures of murine MHC Class I H-2 D(b) and K(b) molecules in complex with CTL epitopes from influenza A virus: implications for TCR repertoire selection and immunodominance. *J. Mol. Biol.*, 345, 1099-110

Wang, J.H., *et al*. (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*

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 co-receptor 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.

Tools for structure determination and analysis

Thomas R. 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 MX@PETRA3

The Schneider group is operating two beamlines for macromolecular crystallography at the PETRA III synchrotron and develops new methods for structure determination.

Previous and current research

During the past six years, we built three beamlines at the new PETRA III synchrotron in close collaboration with the Cipriani (page 104) and Fiedler (page 115) teams. Since 2012, the beamlines for small angle X-ray scattering on solutions (SAXS) and X-ray crystallography on crystals of biological macromolecules (MX) are open for users. The beamlines are embedded in an Integrated Facility for Structural Biology that supports non-specialists in taking a project from producing a suitable sample to determination of the structure by SAXS and/or MX.

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

We will continue the commissioning work on the new beamlines and strengthen the user programme to full capacity. The two beamlines for macromolecular crystallography will offer excellent conditions for data collection on small (micron-sized) crystals, crystallographic phasing, and the development of new experimental protocols for challenging systems.

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/facilities/mx**

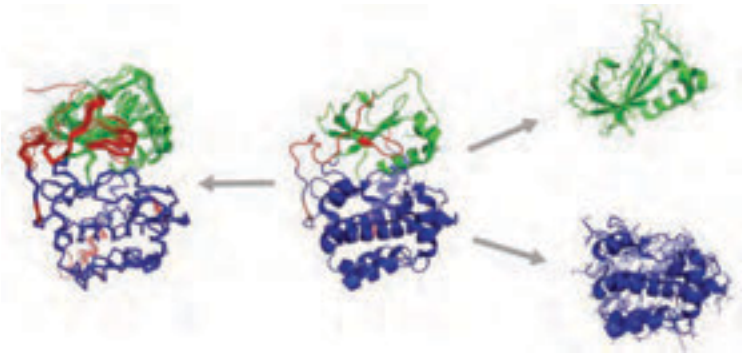
Selected references

Schneider, T.R. (2008). Synchrotron radiation: micrometer-sized X-ray beams as fine tools for macromolecular crystallography. *HFSP J.*, 2, 302-6

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, W42-6

Penengo, L., *et al*. (2006). Crystal structure of the ubiquitin binding domains of rabex-5 reveals two modes of interaction with ubiquitin. *Cell*, 124, 1183-95

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 resolution and cross-validate structural models.

Dmitri Svergun

PhD 1982, Dr of Science 1997, Institute of Crystallography, Moscow.

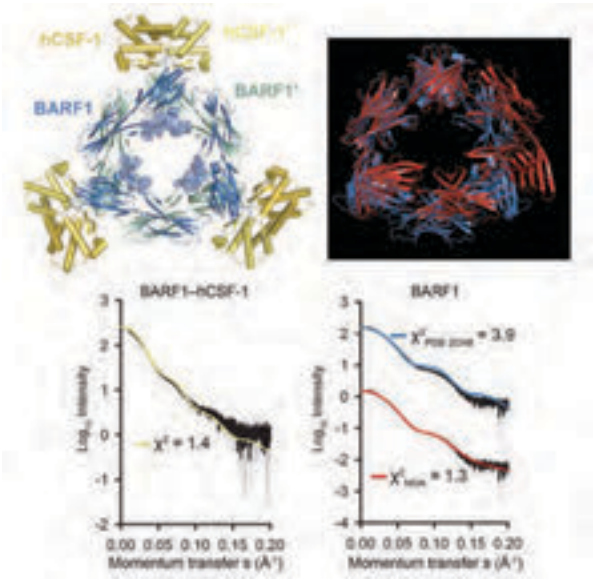
At EMBL since 1991.
Group leader since 2003.
Senior scientist since 2011.

Selected references

Mertens, H.D., Piljic, A., Schultz, C. & Svergun, D.I. (2012). Conformational Analysis of a Genetically Encoded FRET Biosensor by SAXS. *Biophys. J.*, 102, 2866-75

Petoukhov, M.V. *et al*. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr*., 45, 342-350

Elegheert, J. *et al*. (2012). Allosteric competitive inactivation of hematopoietic CSF-1 signaling by the viral decoy receptor BARF1, *Nat. Struct. Mol. Biol*. 19, 938-947.



Visualising a decoy protein of a herpes virus at work
The Epstein Barr Virus (EBV), a member of the herpes virus family, is a global human pathogen causing a diverse range of diseases. A multidisciplinary study using synchrotron radiation, electron microscopy, biophysical and cellular methods (Elegheert et al*., 2012) revealed at the molecular level how the virus can deactivate the alert system of the body's immune defence. The EBV secretes a protein BARF1 that neutralises a human colony-stimulating factor 1 (hCSF-1). It is shown that the flexible BARF1 protein locks the dimeric hCSF-1 into an inactive conformation, rendering it unable to perform its signalling function. The novel mechanisms have implications for the development of therapies and drug compounds. In the figure, the crystal structure of the BARF1-CSF1 complex (top left) is validated by SAXS data (bottom left), whereas the crystal structure of the free BARF1 (blue cartoon, top right) does not agree with SAXS (bottom right) and was refined using normal mode analysis (red cartoon, top right).*

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 in variable conditions. For many complicated biological systems which may be flexible or have a dynamic nature, SAXS is the only method capable of providing structural information. Recent experimental and methodical developments have significantly enhanced the resolution and reliability of the SAXS-based structural models. This versatility and universality – and the fact that it does not need crystals to characterise the structure – makes SAXS an ideal tool for systems structural biology, and the last decade saw a renaissance of biological SAXS worldwide.

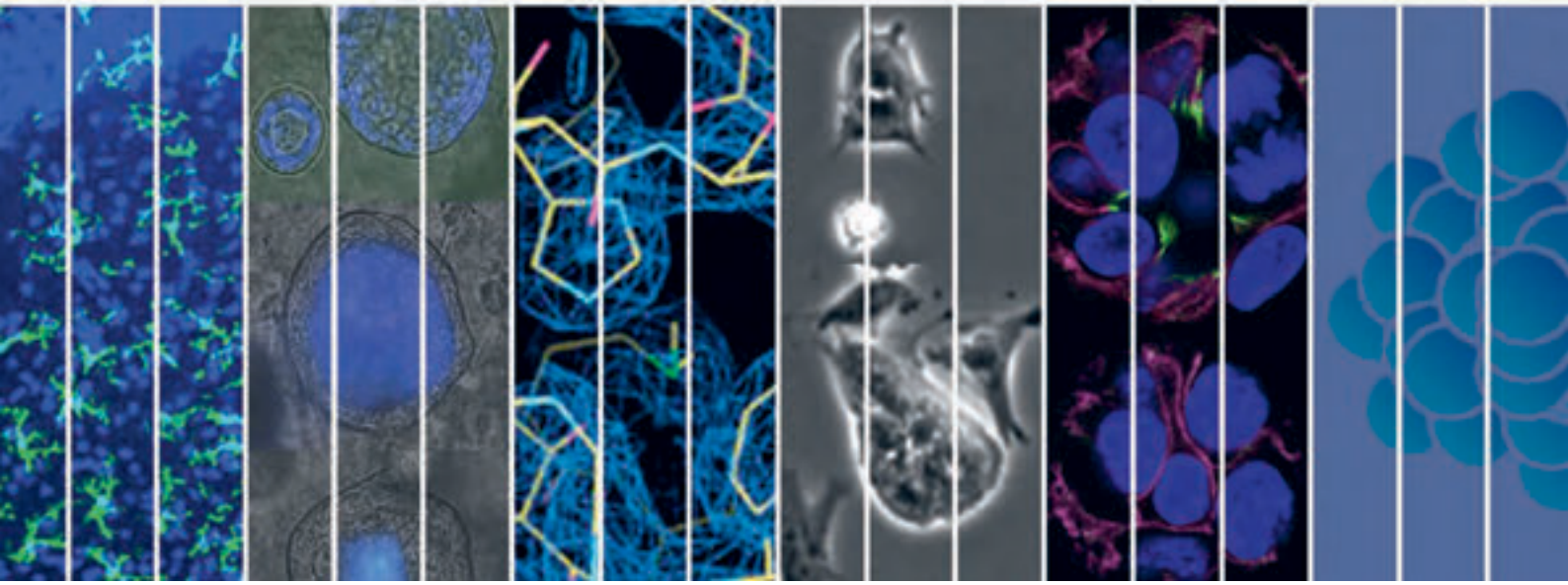
Our group leads the development of novel computational methods for constructing 3D structural models from the scattering data. Special attention is given to the joint use of SAXS with other structural, biophysical and biochemical techniques including: crystallography, NMR spectroscopy, electron microscopy, neutron scattering, and bioinformatics. We developed ATSAS, the world's most popular biological SAXS analysis program suite, which has been downloaded by more than 7000 users, and we continue to provide the scientific community with novel approaches.

We run a brand new high-brilliance synchrotron beamline, P12, at DESY's third-generation storage ring, PETRA III. The beamline, commissioned in 2012, is optimised and dedicated to biological solution SAXS. P12 is equipped with a robotic sample changer, utilising 96 well plates for rapid automated experiments. It possesses a data analysis pipeline for building structural models online, with FedEx-style and remote data access options. As of 2012 we also offer an in-line size exclusion chromatography setup, with biophysical sample characterisation using a triple detector Malvern box.

Most of the external users of P12 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 overall structural organisation and flexibility of individual macromolecules and complexes (see figure) and conformational transitions (for instance, upon ligand binding). It is also used to characterise oligomeric mixtures, flexible systems and intrinsically unfolded proteins, hierarchical systems, 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 structure of macromolecules and complexes from X-ray and neutron scattering data.
- Hybrid applications of SAXS with crystallography, NMR, electron microscopy and other methods, and the use of bioinformatics to construct and validate SAXS-based models.
- Participation in collaborative projects at the P12 beamline employing SAXS to study the structure of a wide range of biological systems in solution.
- Further improvements of the capabilities of P12, including complete automation of biological SAXS experiments and data analysis, online FPLC, and time-resolved scattering setups.



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 Director of Research.

Head of the Mouse Molecular Genetics Unit at the Institut Pasteur 1990-2011.

Head of the Developmental Biology Department at the Institut Pasteur 2006-2011.

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 emphasis 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

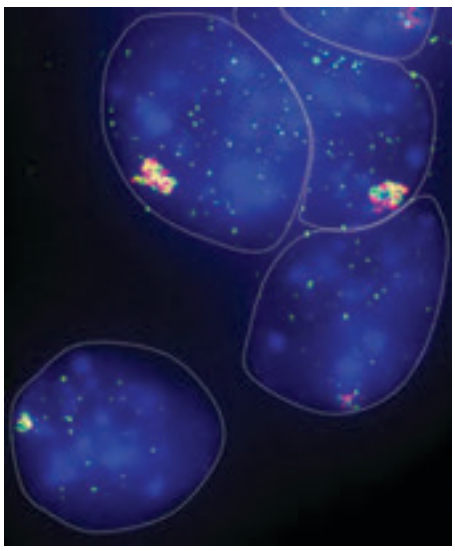
Chureau, C., *et al*. (2011). Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. *Human Molecular Genetics,* 20, 705-18

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

Navarro, P., *et al*. (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 behaviour

The Gross group uses pharmacological, histochemical, electrophysiological and behavioural genetic approaches to study the neural circuits underlying 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

Gross, C.T., & Canteras, N.S. (2012). The many paths to fear. *Nat. Rev. Neurosci*., 13, 651-8

Paolicelli, R.C., *et al*. (2011). Synaptic pruning by microglia is necessary for normal brain development. *Science*, 333, 1456-8

Gozzi, A., *et al.* (2010). A neural switch for active and passive fear. *Neuron*, 67, 656-66

Audero, E., *et al.* (2008). Sporadic autonomic dysregulation and death associated with excessive serotonin autoinhibition. *Science*, 321, 130-3

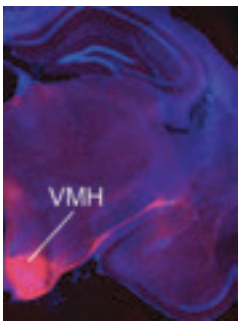


Figure 1. Expression of the hM4D/CNO pharmacogenetic inhibition tool in cell bodies and projections of the VMH in Nr5a1::hM4D-2A-tomatoF transgenic mice as revealed by farnesylated tomato (red) reporter protein (B. Silva)

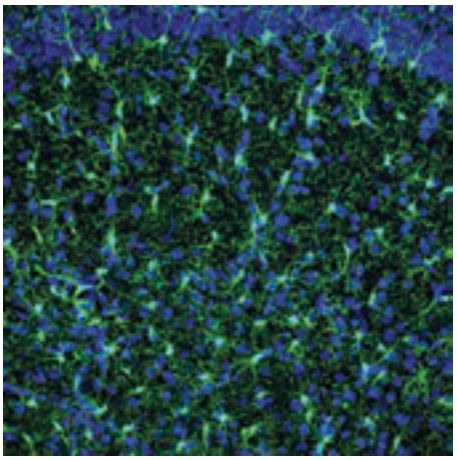


Figure 2: Microglia (green cells) visualised in the hippocampus of Cx3cr1 ^{GFP} transgenic mice (cell nuclei labelled in blue, R. Paolicelli)

Previous and current research

The laboratory is interested in understanding at a molecular and neural circuit level how early life events influence brain development in order to establish behavioural traits in adulthood, with a particular focus on fear and anxiety. We are currently pursuing two areas of research:

Neural circuits encoding fear and anxiety

Fear is a mental state that is elicited by exposure to threats or cues that signal those threats. Fear is part of an organism's natural defense mechanism and the accompanying behavioural and physiological responses are essential for it to cope with potential bodily harm. However, in its pathological form, fear can become excessive or inappropriate – features associated with anxiety disorders. It is accepted that the amygdala plays a central role in processing fear. However, it is less widely appreciated that distinct amygdala outputs and downstream circuits are recruited in response to different types of fear. Immediate early-gene mapping studies show that exposure to painful stimuli, predators, and aggressive members of the same species, activates distinct neural circuits that involve the amygdala, medial hypothalamus, and periaqueductal gray (Gross and Canteras, 2012). These data suggest that independent fear circuits may exist to respond to different classes of threat and imply that pathological fear may come in different flavours and be amenable to selective therapeutic treatment. Current work in the lab combines molecular genetic, electrophysiological, and genetically encoded neural manipulation tools (figure 1) with behavioural methods in mice to understand how amygdala, hypothalamic, and brainstem circuits support and adapt fear responses to diverse threats.

Developmental programming of brain wiring by microglia

Microglia are non-neuronal cells of the hematopoetic lineage that infiltrate the brain during development and are thought to play a role in brain surveillance. Recent studies from our group and others have shown that microglia are particularly abundant during the period of postnatal brain development when synapses are formed (figure 2) and that they play a key role in the elimination of synapses during this period, a phenomenon called 'synaptic pruning' (Paolicelli *et al*., 2011). Mice with deficient synaptic pruning show deficits in neural circuit maturation and we are currently investigating the long-term consequences of deficits in synaptic pruning on circuit wiring and function.

Future projects and goals

Together these approaches are aimed at discovering the neural circuits and molecular mechanisms that support individual differences in behavioural traits in health and disease. A better understanding of the molecular signals that influence the formation and remodelling of these circuits will allow us to form specific hypotheses about how human behaviour is determined and lead to improved diagnostic and therapeutic tools for mental illness.

Molecular physiology of sensation

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.

We use a combination of molecular, imaging and electrophysiological techniques to examine functional properties of sensory neurons at their peripheral and central terminals. 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 the cytoskeleton in regulating the composition and function of this complex. Using cellular, electrophysiological and molecular imaging techniques we are characterising the contribution of the cytoskeleton, in particular microtubules to mechanotransduction.

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 defined subpopulations of sensory neurons.
- 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, will 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

Wang, H., Schupp, M., Zurborg, S. & Heppenstall, P.A. (2013). Residues in the pore region of *Drosophila* transient receptor potential A1 dictate sensitivity to thermal stimuli. *J. Physiol*., 591, 185-201

Kalebic, N., *et al*. (2012). The Tubulin Acetyltransferase alphaTAT1 Destabilizes Microtubules Independently of its Acetylation Activity. *Mol. Cell Biol.*, epub ahead of print

Lapatsina, L., *et al*. (2012). Regulation of ASIC channels by a stomatin/STOML3 complex located in a mobile vesicle pool in sensory neurons. *Open Biol*., 2, 120096

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

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, 1841-4

Jechlinger, M., *et al*. (2006). Autocrine PDGFR signaling promotes mammary cancer metastasis. *J. Clin. Invest*, 116, 1561-70

Jechlinger, M., *et al.* (2003). Expression profiling of epithelial plasticity in tumor progression. *Oncogene*, 16, 7155-69

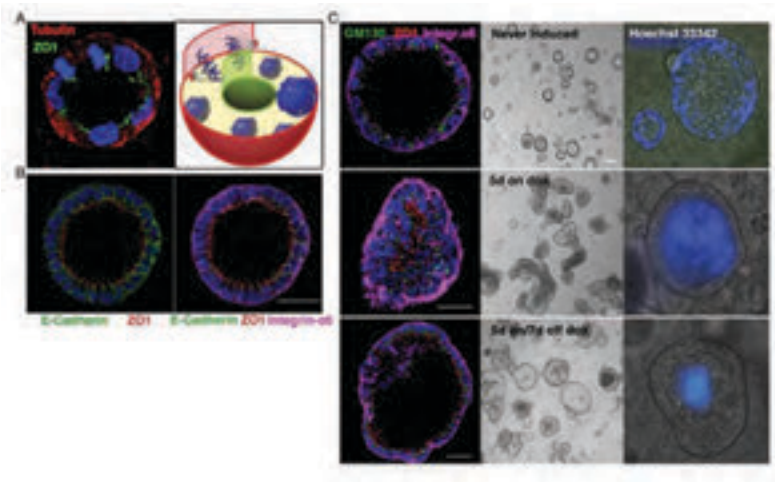
Previous and current research

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 tumourigenesis 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 localised 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 so 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 types, 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 types 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 cell generation from ESC or iPSC for regenerative medicine.

Selected references

Lancrin, C., *et al*. (2012). GFI1 and GFI1B control the loss of endothelial identity of hemogenic endothelium during hematopoietic commitment. *Blood*, 120, 314-22

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., *et al*. (2010). Blood cell generation from the hemangioblast. *J. Mol. Med*., 88, 167-72

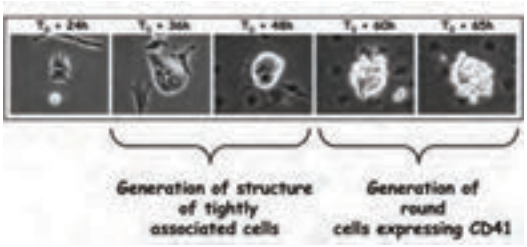


Figure 1: Time-lapse microscopy analysis of haemangioblast differentiation. The formation of a blast colony from the haemangioblast can be retrospectively divided in two consecutive phases: a generation of a structure of tightly associated endothelial cells and the production of round non adherent cells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

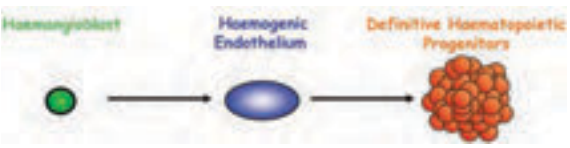


Figure 2: New model of blood cell origin: The haemangioblast and the haemogenic endothelium are part of the same developmental pathway to generate blood cell progenitors during embryonic life

Non-coding RNA function and RNA modification in germ/stem cell biology

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.

Selected references

De Fazio, S., *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., *et al.* (2010). Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. *Proc. Natl. Acad. Sci. USA*, 107, 15163-8

Rasmussen, K.D., *et al* (2010). The miR-144/451 locus is required for erythroid homeostasis. *J. Exp. Med.*, 207, 1351-8

Dónal O'Carroll

PhD 1999, Research Institute of Molecular Pathology, Vienna.

Postdoctoral research at The Rockefeller University, New York.
Adjunct member of faculty, The Rockefeller University, since 2007.

Group leader at EMBL Monterotondo since 2007.
ERC Investigator since 2012.

Previous and current research

The goal of my laboratory is to explore the contribution of non-coding RNA as well as RNA modification pathways to tissue development and homeostasis. We have garnered much experimental expertise in both haematopoiesis and the germline. Within the laboratory, an emphasis (non-exclusive) is now placed on studying the immortal lineage. The integrity of the genome transmitted to the next generation intrinsically relies on cells of the germline. Processes that ensure germ cell development, genomic stability, and reproductive lifespan are essential for the long-term success of a species. We tackle fundamental questions regarding the mammalian male germline and heredity from an RNA perspective. Specifically, our research explores the contribution of non-coding RNA (miRNA, piRNA and lncRNA) as well as RNA modification pathways within germ cell development as well as testicular homeostasis and regeneration.

The precise identity of the spermatogonial stem cell (SSC) *in vivo* that supports spermatogenesis throughout life remains unknown. Capitalising on the fact that SSC maintenance is dependent upon several RNA-binding proteins, we hope that investigation into these pathways may reveal the identity of this stem cell *in vivo*. The maturation of RNA sequencing techniques, in combination with refined genetic approaches, now renders the identification and functional evaluation of non-coding RNAs and RNA modifications *in vivo* within the realm of experimental feasibility. Our research objectives focus on the contribution of these emerging pathways on the underlying circuitry of self-renewal that underpins the SSC, as well as the coordination of the various cellular/differentiation processes of spermatogenesis.

The acquisition of both pluripotency and totipotency is associated with the deregulation of transposable elements – our goal is understand the mechanisms by which the germ cells manage this formidable threat to gametes, and thus transgenerational genome stability. Specifically, transposon silencing in the germline by the Piwi-interacting RNA (piRNA) pathway as well as epigenetic mechanisms will be extended upon from our previous findings.

Future projects and goals

- The identification and characterisation of the spermatogonial stem cell populations *in vivo*.
- Dissection of the pathways required for spermatogonial stem cell self-renewal and testicular regeneration.
- Post-transcriptional RNA modification in germ cell and hematopoietic development.
- Long non-coding RNA function in spermatogenesis.
- Establishment and maintenance of epigenetic transposon silencing in the male germ line.

Mitotic chromosomal instability and oncogene dependence

Rocio Sotillo

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

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

At EMBL since September 2010.

ERC investigator since 2011.

HHMI International Early Career Scientist.

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 focuses on understanding the molecular mechanisms that lead to CIN and the consequences it may have in tumour initiation, suppression and relapse. We hope 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).

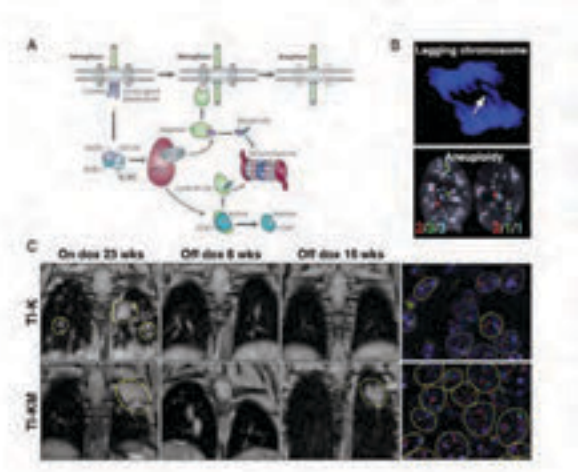
Selected references

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, 701-14

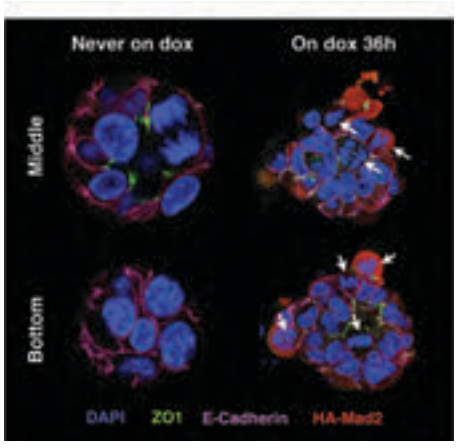
Sotillo, R., Schvartzman, J.M., Socci, N.D. & Benezra, R. (2010). Mad2-induced chromosome instability leads to lung tumour relapse after oncogene withdrawal. *Nature*, 464, 436-40

Schvartzman, J.M., Sotillo, R. & Benezra, R. (2010). Mitotic chromosomal instability and cancer: mouse modelling of the human disease. *Nat. Rev. Cancer*, 10, 102-15

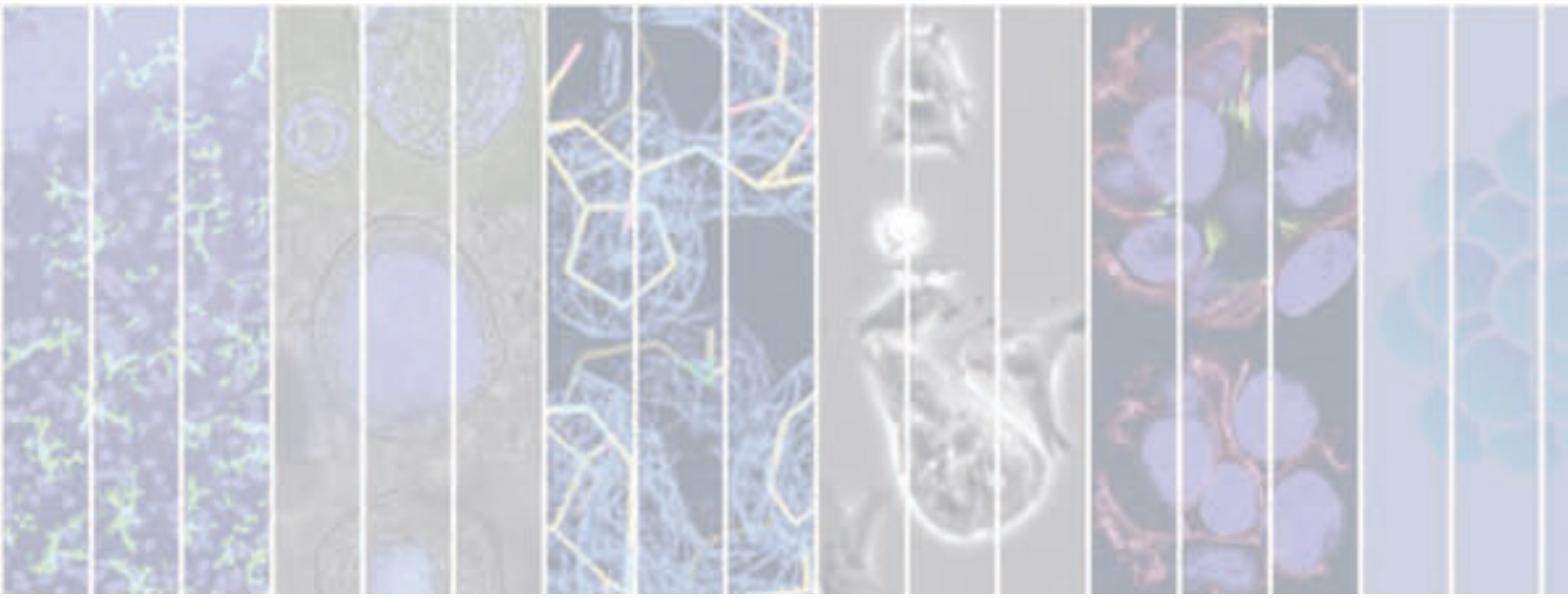
Sotillo, R., *et al*. (2007). Mad2 overexpression promotes aneuploidy and tumorigenesis in mice. *Cancer Cell*, 11, 9-23



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



Primary mammary cells from TetO-Mad2/TetO-Myc/MMTV-rtTA mice grown in 3D culture. Left panel shows never induced cells that are beginning 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



Index

Published by the Office of Information and Public Affairs

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

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

EMBL Grenoble
6, rue Jules Horowitz, BP181
38042 Grenoble, Cedex 9
France
Tel. +33 (0)4 76 20 72 69, Fax +33 (0)4 76 20 71 99

EMBL Hamburg
c/o DESY
Notkestraße 85
22603 Hamburg
Germany
Tel. +49 (0)40 89 90 20, Fax +49 (0)40 89 90 21 04

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

European Molecular Biology Laboratory

Heidelberg • Hinxton • Grenoble • Hamburg • Monterotondo

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

www.embl.org