EMBL Programme
2012–2016

EMBL

EMBL Programme

2012–2016

EMBL

Table of Contents

List of boxes

C. Research

D. Infrastructure and services to the member states

E. Training

F. EMBL's impact on the economy and society

A. Executive Summary

1. EMBL's place in Europe

Science is the driving force of progress. Basic research provides the understanding of the world that we need to responsibly interact with it and make sustainable use of its resources. It also lies at the heart of innovation. Throughout the world, countries have realised that their future growth and prosperity relies heavily on innovation arising from key areas of research and development (R&D). The three areas of R&D widely expected to drive progress in the first half of this century are information and communication technology, nanotechnology and the life sciences. In 2000 Europe, in the form of the EU governments, therefore adopted the goal of becoming the world's leading knowledge economy. In 2002, the EU set itself the Barcelona target of spending 3% of its GDP on R&D by 2010. In actual fact, EU expenditure on R&D only increased from 1.82% of combined GDP in 2002 to 1.9% in 2009 (Science and Engineering Indicators 2010, National Board of Science) and the EU2020 Vision therefore recently renewed the target of 3% of Europe's GDP, now proposed to be spent on R&D by 2020. In the past decade, the gap in R&D expenditure between the USA and Europe widened considerably and the EU countries' R&D budget, which had been greater than the combined efforts of the Asian countries in 2002, represented only 78% of Asian expenditure in 2007 (Figure A.1.). This trend means that during the first decade of the century there has been a significant reduction in the share of world R&D activity taking place in Europe.

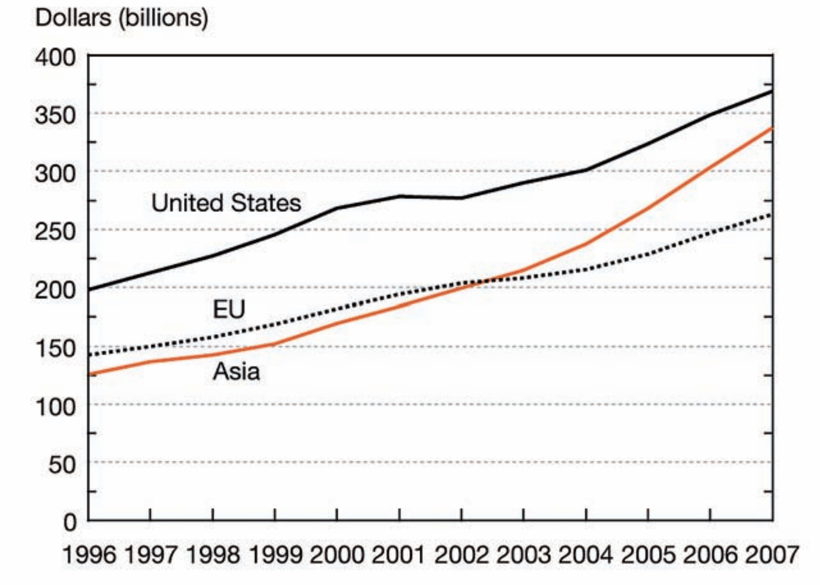


Figure A.1. R&D expenditures for United States, EU and Asia 1996-2007. *Source: Science and Engineering Indicators 2010, National Board of Science.*

If Europe is to remain competitive in this situation, it is critical that European countries invest their limited resources where they have the greatest impact and obtain the best possible return. As we will illustrate in this document EMBL – Europe's only intergovernmental life science research institution – is an obvious place to support world-class life science activities in Europe and provides an example of how limited R&D investment can be employed effectively and with broad impact.

EMBL is a world-leading basic research institute, as shown both by the results of regular, detailed, external reviews of its scientific activities and by analysis of the impact of EMBL's scientific output. Bibliometric studies have identified EMBL as one of the world's top molecular biology and genetics institutes during the past two decades, and it currently ranks number one in Europe and fourth worldwide.

However, the remarkable feature of EMBL is the way its unique combination of activities synergises with its research, thereby amplifying EMBL's positive impact on European life science R&D to the benefit of the EMBL member states.

	Institution	Papers	Citations	Citations per paper
1	Cold Spring Harbor Lab	669	63,570	95.02
2	MIT	1,995	163,596	82.00
3	Salk Institute for Biological Studies	707	49,996	70.72
4	European Molecular Biology Lab	1,435	94,736	66.02
5	Memorial Sloan-Kettering Cancer Centre	1,099	71,250	64.83
6	Wellcome Trust Sanger Institute	790	50,997	64.55
7	Rockefeller University	1,332	83,307	62.54
8	Dana Farber Cancer Institute	673	41,627	61.85
9	Massachusetts General Hospital	1,447	86,773	59.97
10	Cancer Research UK	752	44,343	58.97
11	Caltech	846	45,887	54.24
12	Harvard University	8,525	460,807	54.05
13	University of Texas Southwestern Medical Centre	1,768	95,517	54.03
14	Max Delbrück Centre for Molecular Medicine	674	36,335	53.91
15	Stanford University	3,029	159,610	52.69
16	Howard Hughes Medical Institute	660	34,534	52.32

Figure A.2. Thomson Essential Science Indicators citation ranking of science institutions in molecular biology and genetics 1999-2009.

1.1 What are EMBL's activities?

- **EMBL is a leader and pioneer in basic research and technology development:** With its outstanding track record and 35 years of experience in basic molecular biology research, complementary sets of tools and technologies and an extraordinarily interdisciplinary science community, EMBL is uniquely poised to take on the new challenges in modern life sciences and decipher the complexity inherent to living systems. Driven by the demands of its researchers, EMBL is heavily engaged in improving key techniques required for its ambitious research plans, including light microscopy and X-ray imaging technologies as well as developing computational biology software.
- **EMBL provides world class services and cutting-edge infrastructures** to the member state scientific communities, prime examples being the roughly 3000 structural biologists who visit EMBL Hamburg and Grenoble each year to make use of X-ray facilities and the more than 1 million yearly users of the EMBL-EBI data resources who span academia, medicine, environmental studies and the pharmaceutical, agriculture, food and biotechnology sectors.
- **EMBL is a hotbed of innovation:** EMBL engages in innovation both through pro-active technology-transfer activities and numerous formal and informal collaborations with industry.
- **EMBL helps to produce Europe's next generation of scientific stars:** EMBL offers outstanding training to all categories of postgraduate scientists. It has developed excellent programmes for both PhD and postdoctoral training and its fixed-term contract system ensures that EMBL's independent researchers, having established themselves and their scientific reputations, move on from EMBL to enrich the national communities in our member states.
- **EMBL is a platform for scientific exchange:** On behalf of the community, EMBL organises a greater number of scientific training courses, workshops and conferences than any other European life science research institute (over 130 per year from 2007 – 2009, with over 10 000 total participants). EMBL also plans and carries out outreach activities that are aimed at school science teachers, schoolchildren and the more general public.
- **EMBL is an instrument for European integration:** EMBL actively promotes the integration of European scientific activities in a number of informal ways via numerous scientific collaborations and networks and also formally by developing partnerships with institutions in the member states and helping to plan European research strategy, for example the next generation of European Research Infrastructures for the biomedical sciences.
- **EMBL serves as a role model of scientific organizations:** EMBL is built on a variety of highly successful organizational principles, including international recruitment of the most talented scientists, regular and stringent peer review its activities, the organization of shared equipment into central Core Facilities, and runs a set of "best in class" activities, e.g. the EMBL International PhD Programme and EMBL's technology transfer activities, that have served as models for multiple other scientific institutions.

All of these activities benefit our member countries, but they also involve and require working together with other European organisations including the European Commission, many other scientific organisations – in particular our sister organisation EMBO/EMBC – and the other European intergovernmental research infrastructure organisations that operate across many scientific disciplines (CERN, EFDA, JET, ESA, ESO, ESRF and ILL).

It is the cross-fertilisation and synergy between EMBL's leading research, services, training, innovation and integration endeavours and the demanding nature of our assessment procedures that allows us to stay at the forefront and achieve world-class standards. Together, they allow a uniquely integrated approach to supporting European molecular biology. Thus, the investments made in EMBL by its member states not only result in excellent science but have a multiplier effect that stretches far beyond EMBL and helps strengthen not only the European life sciences but also the entire European research landscape and helps pave the way for broader international cooperation.

2. Areas of EMBL focus in 2012-2016

The following is a brief summary of our aims and the new investments required to achieve them.

2.1 Research

- **Information Biology:** In the 5-year period encompassed by this document EMBL will pursue a multidisciplinary and multiscale approach towards unravelling the principles and logic of the complex functional organisation that defines all biological systems. We will purse an approach that combines EMBL's traditional strength in mechanistic studies and hypothesis-driven research, using mostly real-time live imaging, complementary structural biology technologies at different scales and biochemical studies with systems approaches employing 'omics' methods and computational biology.

  How the collection of genes in an organism – its genotype – gives rise to its developing and adult forms, and their defects, is a problem whose understanding far exceeds human intuition. The only way to disentangle and understand this complexity is to extract biological information from large and diverse experimental datasets as produced by the methods mentioned above. This approach turns biology into a very information-rich science and making sense and use of this information is the challenge that will define the future of life science research. EMBL is poised to take on this challenge, and for this reason we have titled our 5-year EMBL Programme "**Information Biology**". Aside from biologists this will require the integration of chemists, physicists, engineers and, above all, making optimal use of EMBL's unparalleled breadth and strength in bioinformatics and computational biology. Computational methods and aids are required, all the way from allowing adequate analysis of the large amounts of biological data being generated through modelling networks of gene and protein interaction through simulating complex biological systems and their defects, e.g. in disease states to pinpointing areas where detailed mechanistic understanding is essential to progress.

  EMBL's research plans should not be seen as an end in themselves. It will be both fascinating and satisfying to understand the principles by which model organisms and cell systems function, but an improved understanding of biological function will also be critical in two areas of applied interest. First, the next generation of biotechnology, medical and pharmaceutical discovery will be based on this understanding. Second, technology breakthroughs mean that the entire genome sequences of thousands, possibly even hundreds of thousands, of human beings will become available in the next decade. Unless we can use experimentally accessible biological systems to make considerable progress in understanding how genotype leads to phenotype, and how DNA sequence variation leads to phenotypic variation, e.g. in determining susceptibility to disease or to rare drug side effects, we will not be in a position to exploit this data to human benefit.

- **EMBL Centres:** EMBL Centres were introduced during the last Programme to support horizontal activities involving many EMBL Units, providing a focus for training in and dissemination of new technologies of broad application. They have carried out this function very successfully. As planned, the four initial Centres are now mature and ready to be incorporated into other structures. Two will be assimilated into either Core Facilities or existing Units. The staff, funding and activities of the Imaging Centre will be incorporated into the Advanced Light Microscopy Facility and those of the High Throughput Functional Genomics Centre into the newly created Genome Biology Unit, which grew out of and replaced the Gene Expression Unit in 2010. The activities of the Disease Mechanisms Centre will in future be carried out in the context of our partnerships for molecular medicine. The **Chemical Biology Centre** has commenced its activities in 2010 and will continue into the new Programme. Owing to increasing demand for specialised advice in the area of computational biology the last of the original Centres, in Computational Biology, will be replaced by three new Centres in the 2012-2016 Programme, each of which will be tightly focussed on areas of major need. They will be the **Centres for Statistical Data Analysis**, **Biomolecular Network Analysis** and **Mathematical Modelling**. In total, the funding requirement for Centres in the new Programme will therefore rise from five to six. As will be evident from later sections in which the new Programme is described in more detail, each of these Centres corresponds to a crucial need in our "Information Biology" plans.

- **Structural Biology in EMBL Hamburg:** For three reasons, we intend to increase the research activity in Hamburg in the next Programme. First, EMBL Hamburg is currently engaged in constructing what will be the most powerful X-ray beamlines in the world for structural biology as part of the PETRA-III project. Second, the host country intends to build a Centre for Structural Systems Biology (CSSB) on the Hamburg site – scheduled to begin construction in 2011 – in order to provide complementary activities to the X-ray expertise being built up in the EMBL outstation and thus make optimal use of the PETRA-III development. It would be a wasted opportunity if EMBL did not exploit these developments. Third, an X-ray Free Electron Laser, which will provide X-rays with orders of magnitude more energy than a synchrotron ring, is being constructed in Hamburg and will come online in 2014. Although pilot studies have been successfully carried out in Hamburg with the FLASH soft X-ray laser, it is not yet clear how beams of this energy will be used in biology, just as it was not clear that synchrotron-based X-rays would be useful for biology when they were first produced. It would, however, be negligent not to examine the possibilities and EMBL, as the only biological laboratory on the site, is the obvious place to do so. We therefore intend to increase the number of research groups at EMBL Hamburg by two.

2.2 Services and Infrastructure

- **Bioinformatics Services: EBI and ELIXIR.** EMBL-EBI is responsible for dealing with a significant part of what is possibly the biggest challenge in the contemporary life sciences: how to enable the scientific community to make optimal use of the avalanche of biological data being produced to support research and promote health and innovation. The 2012-2016 Programme will see the development of another aspect of this theme. ELIXIR is an ESFRI project whose aim is to plan and construct the next generation of bioinformatic data resources for Europe. Due to its existing resources and expertise, that are unique in Europe, EMBL-EBI is currently the only viable option for the ELIXIR hub. The hub will host the major biomolecular data resources and provide the connectivity needed for integration and interoperability of the distributed national data resources that will make up the ELIXIR nodes and be responsible for the delivery of bioinformatic services to the user community. While the creation of ELIXIR initially requires significant investment to create a hub at the EBI, at the same time it offers a sustainable solution for the problem of how to fund the collection and provision of biomolecular data in the long-term and is thus essential for the future health of the European life science community. Decentralisation and distribution of data resources will make the provision of bioinformatics services to European users feasible in spite of the ongoing rapid increase in data production and diversification of data types that will be required.
- **Core Facilities:** In order to pursue the Programme described here, it is clear that we will need to make considerable investments in imaging technologies. These include super-resolution light microscopy, the methods of which are just coming to maturity; robotics, for which additions to the facilities will be needed to adequately support the research in the new Genome Biology Unit in Heidelberg; new DNA sequencing technologies and the associated computational hard- and software; and cell sorters, a mass spectrometer and an iso-thermal calorimeter to keep EMBL's analytical facilities up-to-date. The greatest cumulative equipment investment will however be in computational infrastructure. This will be needed to support service, research and administrative activities both at EMBL-EBI and in the main Laboratory in Heidelberg.

2.3 Technology Transfer

- **Proof of Concept Fund:** EMBL set up its technology-transfer company, EMBLEM, in 1999 and stimulated the formation of EMBL Ventures, which from 2005-2009 managed ETFI, a venture capital fund that invested, in part, in EMBL technology. A second fund, ETFII, will begin its activities in 2010. However, the technology-transfer landscape has changed significantly in the past 5-7 years, with venture capital investing later and later in the translation process. To bridge the gap that this has created between basic research findings and application and between what is supported by EMBL and EMBLEM and what is attractive for outside investors, it is our intention to finance – EMBL Council support permitting – a 'Proof of Concept' fund that will enable us to develop our discoveries to the point at which outside investors are prepared to take them over and help them reach the market.

2.4 Training

- **Interdisciplinary training:** The complexity of current biological research questions frequently demands a combination of expertise, e.g from chemistry, physics, mathematics or engineering as well as from biology. During the last Programme, we piloted a new scheme, EMBL Interdisciplinary Postdoctoral Fellows (EIPODs), to help us achieve this goal. EIPODs bring two specific advantages. First, they must work on a project that involves at least two laboratories and two distinct sets of expertise. In this way cross-discipline and cross-Unit projects are greatly encouraged. Second, postdocs who are not biologists, and who would have been unlikely to choose to work at EMBL, are brought into the Laboratory. This enriches not only EMBL but also the fellows, who obtain training and expertise in a second discipline and are therefore better prepared for a future career in the life sciences. We believe that the need to support interdisciplinary projects means that the EIPOD scheme should start with 10 internally funded positions in 2012 and increase to 25 such positions during the course of the next Programme.
- **ATC activities:** The opening of the Advanced Training Centre (ATC) means that many more visitors will now come to EMBL Heidelberg. To make optimal use of this facility we must increase the conference support staff both in the Course and Conference office itself and in other areas such as the Photolab, which is responsible for audiovisual support and the Office of Information and Public Affairs, which generates all the publicity and information materials needed to run an efficient conference programme. A new activity that has come with the ATC is a symposium series through which EMBO and EMBL will jointly support five to six major scientific conferences per year, as discussed with EMBL Council and the EMBC. This series will enable us to underline the role of the ATC as a unique European platform for scientific exchange.

2.5 European Integration

EMBL plays a crucial role in integrating European research. The last Programme saw considerable growth of the workload of EMBL's small, and undermanned, scientific Administration. The four new remote Partnership laboratories, Australia's associate membership and in particular EMBL's intense efforts in developing the biomedical projects that are part of the ESFRI process all come with administrative responsibilities and require considerable time and effort. To be able to maintain these and other activities that contribute to EMBL's central role in the integration of life science research we must increase the size of the scientific administration. This will increase our capacity for interacting with the existing and potential new member states and for pursuing the ESFRI projects in which EMBL has a major role.

2.6 Administration

During the new Indicative Scheme the Administration will streamline its activities. In this framework a series of projects will be undertaken, most of which we will be able to manage internally. However, there are some areas that will require additional resources and some investment.

- The General Training and Development programme was introduced in the current EMBL Programme and has been very popular with staff and productive for the Laboratory. A small increase to its modest budget will enable us to provide courses complementing scientific skills with transferrable skills tailor-made to EMBL's staff.
- The provision of additional social benefits to fellows, as already decided by Council, has cost implications, as do the reorganisation of the Health Insurance Scheme and any update to the EMBL pension scheme, when decided by Council.
- EMBL has invested substantially in SAP as its operative system. In the new period we will need to make sure that maximum advantage is derived from that investment.
- After the completion of the ATC and the move of most of Administration to new offices, the Main Lab in Heidelberg will continue to be spatially reorganised in order to better accommodate changes in EMBL's research. Such changes will also affect space requirements at the EMBL-EBI.

B. Introduction

EMBL is a remarkable success story, and a tribute to the vision of both the scientists who conceived the Laboratory and the member states that have guided and supported its development over more than 35 years. EMBL was established in 1974 and is today recognised as one of the leading research institutions in the world. But EMBL is also much more than that. The Laboratory provides hundreds of thousands of external scientists with access to research infrastructures every year, and its training programmes have helped launch the careers of several thousand life scientists, many of whom now have leading positions in national research institutions. Indeed, the training programmes are regularly used as models of best practice by other research organisations. EMBL is broadly engaged in technology development and both the resulting inventions and EMBL's research discoveries form the basis for lively interaction with industry and a successful technology transfer programme that makes these findings available to society. Finally, EMBL engages intensively with the scientific communities in its member states, the European Commission, other research organisations and the general public to pursue its mission of promoting the molecular life sciences in Europe and beyond.

This unique mix of research-related activities is the key to EMBL's success. They function synergistically: the research, service provision, training, technology development and innovation activities cross-fertilise and stimulate each other such that the whole becomes much more than the sum of its parts, and ultimately produces EMBL's integrated approach to supporting European molecular biology.

Building on this foundation, the EMBL Programme 2012-2016 is geared towards achieving world-class standards across the entire range of EMBL's activities. We aim to stay at the forefront of research, service provision and training in the life sciences to the benefit of our member states. Goals for the EMBL Programme 2012-2016 are:

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

A unique combination of features underpins EMBL's excellence and equips it to achieve these goals:

- EMBL is the only intergovernmental organisation that performs research in the molecular life sciences, not only in Europe but worldwide. It is directly supported by 20 European member countries and one associate member, Australia.
- Rigorous quality control by external experts ensures excellence in all of EMBL's activities. The high standard of its research has allowed EMBL to further increase its scope by supplementing the funding it receives from the member states by successfully raising competitive external funding, a large part of which comes from the European Commission.
- We use stringent selection criteria to recruit Europe's most talented young scientists and continuous training and mentoring at EMBL allows them to realise their full potential.
- Across all sites, EMBL's staff is exceptionally international and – in 2010 – comprises 1529 people, including postdoctoral fellows and PhD students, from more than 70 different countries. The internationality and youth of our staff contributes to creating a distinct EMBL culture that stimulates creativity, interdisciplinarity and collaboration.
- A unique fixed-term contract system ensures continuous turnover of staff and creates a dynamic environment and constant renewal. This policy provides EMBL with the flexibility required to quickly react to new trends in the life sciences and to allow any areas that are underperforming to be rapidly identified and improved. This system promotes ongoing discontinuation of unsuccessful or outdated initiatives and, by permitting gradual thematic redirection over time, avoids the necessity for sudden, drastic strategic changes. After leaving EMBL, more than 80% of our alumni return to one of the member states and their experience, skills and the international network of contacts they acquire while at EMBL provide valuable contributions to national research and development efforts.
- EMBL's staff represents an unparalleled breadth of interdisciplinary expertise and complementary skill sets.

Here we present an integrated, cohesive strategy for the period 2012-2016 that details the plans for the entire institute and outlines how EMBL aims to achieve its ambitious goals over this five-year period. As part of this process we have defined a set of strategic objectives for each of the goals specified above to help steer progress, streamline EMBL's approach across its various activities and to serve as a basis for long-term strategic orientation of the institute. These objectives are briefly presented in the course of this Programme Introduction and the complete list can be found in Appendix B.1. Accompanying the goals and objectives we have developed ways to monitor the progress towards their fulfillment and to assess the performance of different programmes. We will collect data and generate performance indicators regularly once a year to systematically monitor the status of individual activities over time. Based on this system we are planning to further improve

EMBL's activities and the efficiency of our internal processes and increase accountability towards our member state in future.

The EMBL Programme 2012-2016 and the strategy it presents were developed in a process that involved input from faculty and administrative staff from all parts of the organisation. The document is not meant as a detailed directive, but rather as a broad guideline to how EMBL will develop in the upcoming period. It should be stressed that EMBL values and promotes the creativity and scientific independence of its young research staff and therefore allows them the freedom to pursue topics of interest to them. They make EMBL much more than the sum of its parts by drawing on and contributing to the pool of expertise, technology and collaborative possibilities available at EMBL, all of which are outlined in this document.

1. Forefront life science research: setting trends and pushing the limits of technology

Strategic Objective: *Promote EMBL's scientific excellence.*

Since its foundation, EMBL's foremost aim has been the creation and maintenance of a centre of excellence in molecular biology research that attracts highly talented young researchers and ensures that they pursue their career in Europe. Although there are difficulties in obtaining an objective measure of the value of the output of research institutes, EMBL can justifiably claim that this aim has been achieved. The most reliable indicator comes from the detailed reports of the committees of experts who regularly subject our research and service activities to external review on behalf of EMBL Council. Over the past decade, these have consistently judged all EMBL Units to be excellent on an international scale. EMBL produces on average 450 publications per year (making an average of five annual publications per research group), at least 10% of which are published in the top ten high-impact journals in molecular biology and genetics (Thomson Reuter's Essential Science Indicators). Considering the impact of these publications, as measured by citations in the scientific literature by the world's best known bibliometrics organisation, EMBL ranks as Europe's most successful institution, and the fourth highest ranking in the world, in molecular biology and genetics from 1999-2009 (see Executive Summary).

1.1 Information Biology

Strategic Objective: *EMBL's scientific focus 2012-2016 will be Information Biology.*

Our commitment to excellence also underpins the research plans for the next EMBL Programme. These are described in detail below, including many specific examples of our research plans (see Section C.2), and this Introduction is only intended to put these plans into context. EMBL's research is progressing in a framework first outlined in the 'Strategic Forward Look 2006-2015', which was prepared by the EMBL leadership with input and advice from EMBL Council and the Scientific Advisory Committee. Because research advances rapidly, it was not intended as a detailed blueprint but rather as a general indicator of the ways in which life science research would develop and how EMBL should help to drive these changes. The major theme of the Strategic Forward Look was systems biology, a term that is used to mean different things. We use it to describe an approach to understanding aspects of living systems and their inherent complexity by analysing the components of systems, their interactions and the structural and functional properties that emerge from these interactions.

Over the course of the current Programme we have made significant progress in understanding biological mechanisms and functions, and have begun to link different levels of biological organisation ranging from individual molecules to cells through to the entire organism (see Section C.1). We have now reached the stage at which data production is

no longer the bottleneck in systems biology. The challenge has shifted towards analysing and integrating large heterogeneous datasets and exploiting this data to build models of biological processes and generate new hypotheses. Making sense and use of complex information has now assumed a central role in biology. Our leading position and traditional strength in bioinformatics and computational biology, which provide the tools to extract meaningful information from data, the broad interdisciplinary expertise represented by our faculty and the close collaboration between experimentalists and theoreticians provide EMBL with an unusually strong platform from which to take on the new challenge of information biology.

EMBL has a long-standing tradition in detailed mechanistic studies of biological processes and of hypothesis-driven research. We will continue to rely on these approaches in future. Hypothesis-driven studies of specific biological processes and systems approaches are complementary strategies. On the one hand the new focus on systems and data integration approaches is now a natural component of the detailed study of specific processes, because it is a way to add meaning to specific data by seeing it in a broader context. On the other hand large-scale, global studies of systems can also help identify the best new targets for detailed mechanistic study.

Strategic Objective: *Combine global, high-throughput approaches with hypothesis-driven, mechanistic analyses.*

Two major research themes in the new Programme will be to obtain a better understanding of biological function and organisation, and of the complexity that is inherent to biological systems. We are rapidly moving towards a time when technical development will mean that genotyping, i.e. obtaining the entire DNA sequence of the genome of any organism, including human individuals, will be both technically trivial and affordable. In the majority of cases, however, we lack the ability to predict how the genes in the genome, and their variations between individuals, give rise to the structure, function and behaviour – the phenotype – of an organism or its component parts. In the new Programme we will concentrate on approaches that will help us bridge this gap. In particular, we will focus on:

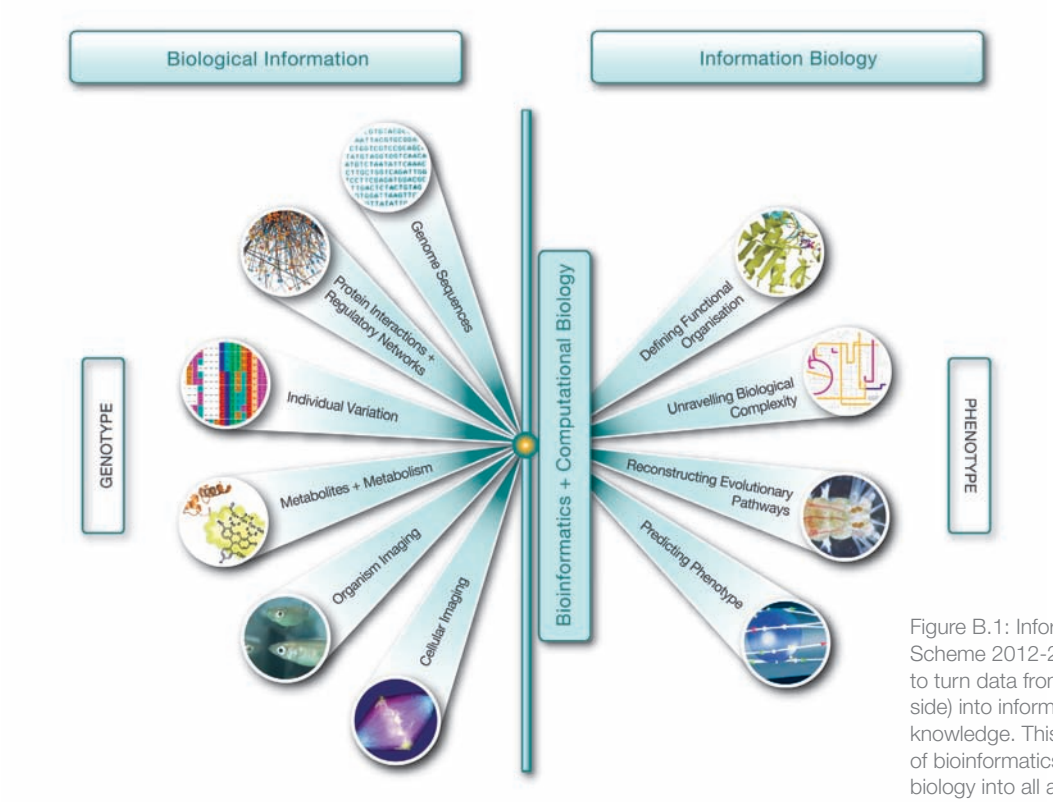


Figure B.1: Information Biology Scheme 2012-2016: The challenge is to turn data from various sources (left side) into information and then into knowledge. This requires the integration of bioinformatics and computational biology into all activities.

Strategic Objective: *Integrate structural and imaging technologies operating at different resolutions to bridge scales from molecules to cells to organisms.*

- **Bridging scales of biological organisation**: With remarkable developments in technology, some of them made at EMBL, and the breadth of complementary expertise that is available in-house, we are now better equipped than ever to bridge different scales of biological organisation. A new marriage between structural biology and cell biology – or cellular structural biology – combines *in vitro*-derived structural and biochemical information with functional studies of the same molecules and complexes in the context of the living cell. In future we will bring high-resolution structural biology methods such as X-ray crystallography, small-angle X-ray scattering (SAXS) and nuclear magnetic resonance (NMR) together with cryo-electron microscopy tomography, proteomics/mass spectroscopy, fluorescent light microscopy, various computational methods and chemical biology approaches to gain a structure-based, mechanistic understanding of cellular processes. Live imaging will be used to place insights at the cellular level into the context of the living organism and its physiology.

Strategic Objective: *Push the limits of research by technology and instrumentation development in imaging, structural biology and computational approaches and exploit new technical developments in these areas by applying them to challenging biological problems.*

- **Biology in four dimensions**: To date, molecular networks have been largely predicted in two dimensions, on the basis of systematic analyses using functional genomics approaches. However, the biological processes mediated by these networks do not occur statically in two dimensions but rather in four: three-dimensional space and time. The next generation of systems biology must therefore capture data in four dimensions, reflecting the physiological conditions and dynamics of the biological system under study. We will develop and apply new technologies for real-time live imaging at the molecular, cellular and organismal level to achieve an understanding of dynamic processes and how they unfold in space and time.
- **Predictive networks and models**: During the next Indicative Scheme a major emphasis will be on developing more predictive networks, models and simulations of biological processes. Predictive networks of complex systems at the level of gene regulation, protein–protein interactions, multimolecular complexes, cellular processes or collective cell behaviour can generate testable hypotheses and predict the quantitative outcome of perturbations, such as the effect of a gene deletion on global gene expression or the effect of a drug treatment on cell growth. Computer models and simulations guide future experiments and, when combined with experimental validation, are the ultimate way to test whether we understand biological processes.

Strategic Objective: *Pursue a combination of experimental and computational approaches; integrate computational biology and bioinformatics seamlessly with wet-lab research.*

- **Generating quantitative data**: Quantitative data is the prerequisite for building predictive networks and models. Systematic global measurements of diverse biological molecules (DNA, RNA, proteins etc.) that are provided by so-called 'omics' approaches (genomics, transcriptomics, proteomics) need to be integrated with kinetic biochemical measurements of real-time dynamics of biological processes. Many of the methods that can quantify molecular interactions and their dynamics use a host of non-invasive light microscopy-based imaging technologies that exploit molecularly defined fluorescence reporters. Developing chemical probes that track the activity of molecules and report molecular interactions will therefore be an important goal for the period of this Programme, as will the development of high-throughput approaches to biochemistry by incorporating tools and methods from nanotechnology.
- **Analysing, integrating and exploiting quantitative data:** As enabling techniques for systems biology, and underpinning all of the efforts described in this document, computational biology and bioinformatics will play a key role in EMBL's future research plans. During the next five years the scope of our bioinformatics research will be broad, to meet the challenges presented above, and it will also help to guide our experimental approaches by generating new knowledge and hypotheses. It will

centre on finding new ways to analyse, integrate and represent the vast quantities of heterogeneous data generated by experimental biologists and on building increasingly accurate models and simulations of a variety of biological processes.

A new dimension that we will incorporate in this Programme is the research theme of biological variation. The enormous advance in DNA-sequencing technology during the past few years has transformed the degree to which we can analyse the genetic variation between species and individuals and the genetic basis of healthy and diseased states. This will allow us to address how genetic variation influences basic biological function and organisation and its effects on complexity at various levels. In particular we will be exploring:

- **Inter-species variation:** The new capacity to perform detailed comparative studies of genetic variation between species has initiated a new era of evolutionary research at all levels of biological complexity and provides a window on the past and insight into the mechanisms of evolutionary change. EMBL researchers will contribute to the field by developing and applying new methods for comparative sequence analysis at multiple levels.
- **Intra-species variation:** DNA sequencing and high-density genotyping of single nucleotide polymorphisms (SNPs) and copy number variant (CNV) regions has provided an unprecedented advance in the capture of data on genetic variation between individuals of a species, primarily human beings. However, this increase in discovery has not yet been matched by an increase in the ability to understand the functional consequences of genetic variation. In the course of this Programme we aim to contribute to a change in our ability to understand these consequences. This will involve synergy between experimental and computational approaches and will make use of both experimentally induced and naturally occurring genetic variation. Our aim is to provide a general methodological framework with specific, cost-effective technologies to analyse the contribution of genetic variants to any phenotype, including disease.
- **Disease models and mechanisms:** Extensive expertise and infrastructure have been established at EMBL over the past decade to build and study increasingly refined animal, notably mouse, and cellular models of pathologies in a spectrum of body systems including brain, blood, muscle, heart, respiratory and reproductive systems. Over the course of the next Programme we plan to apply this expertise to help decipher the molecular basis of a range of genetic and infectious diseases and to share our knowledge by participation in several international activities established for the systematic generation of mouse models.
- **Towards human biology:** EMBL is and will remain a basic research institute with a focus on understanding the fundamental principles underlying the function of living systems. Basic biological research has largely focused on model organisms that allow ease of experimental manipulation and analysis. However, a combination of the application of the findings made in basic research and recent technology breakthroughs means that human biology has moved closer towards the centre of basic research and this is reflected in our future plans. We expect the human being to become biology's best-studied organism at the level of individual variation in DNA sequence. Understanding how this variation changes function to lead to specific aspects of human phenotype, such as propensity to disease, will rely on two skill sets in which EMBL is strong. First, bioinformatics will lead to predictions of the functional effects of variant gene and protein sequences; and second, studies in model organisms will help test, modify and improve these predictions to the point at which they can be tested directly in, or applied to, humans. Medicine and basic molecular biology have already started to converge with diagnosis increasingly moving from the phenotypic description of symptoms towards the molecular characterisation of

patients and disease states. More sophisticated genetic and biochemical methods will allow the engineering of increasingly better animal models of human disease and advances in computational biology and bioinformatics provide new ways to integrate and compare data from model organisms with human biology. It is in this light that we see many of the research plans presented in this document.

1.2 Interdisciplinary collaboration

Strategic Objective: *Advance a network of formal and informal internal and external collaborations and joint activities that promote interdisciplinarity and excellence in European life science.*

As the above outline of future research plans show, systems biology and information biology require that the traditional methods of molecular biology are supplemented by techniques and expertise both from other biomedical fields and diverse scientific disciplines including physics, chemistry and engineering. However, the most pressing challenge is to bring together these 'wet' experimental disciplines with mathematicians, theoreticians and computational biologists. For EMBL this means becoming more interdisciplinary and diverse and engaging in a broad range of scientific collaborations and partnerships with research institutes and laboratories that have complementary expertise and interests (see Section G).

In the current Programme we have utilised EMBL's turnover system to increase the breadth of scientific expertise available in-house. As will be discussed later in more detail, we have hired new group leaders working in various aspects of instrumentation and technology development to support quantitative approaches, for example in developing new imaging methods and activity probes to measure functional processes in real time *in vivo* and *in vitro*. We have also appointed our first group leader in the area of nanotechnology to bring a new impetus to the development of high-throughput biochemistry methods and have increased both the number and the diversity of groups and teams taking computational approaches to biological problems. This has provided us with additional expertise in modelling and simulation, the mathematical expertise required for the accurate and efficient analysis of the enormous datasets produced by high-throughput methods, and further breadth in bioinformatics research and support.

As well as recruiting new cross-disciplinary expertise to EMBL it is equally important to ensure that groups with complementary techniques and skills at EMBL work together. As the structure of this document will show, the boundaries between the traditional fields of molecular biology research are blurring and the complexity of modern biology problems means that projects frequently span several EMBL Units. Collaboration has always been one of EMBL's strengths and, owing to the unique, communicative culture fostered in the Laboratory, usually happens naturally. However, to make sure that the breadth of expertise and skills represented at EMBL is put to optimal use we have implemented or will implement a variety of mechanisms, including the EMBL Centres and the new EMBL Interdisciplinary Postdoc (EIPOD) Programme. Efforts to maintain and further expand internal and external collaborations are an integral part of this Programme (see Sections C.3 and E).

2. Providing world-class research infrastructure and services to the member states

EMBL's second mission is to provide research infrastructure and services for the scientific communities in its member states. These service activities build on EMBL's excellent research base, which helps ensure that they remain at the cutting edge, and go hand-in-hand with EMBL's effort in technology development, which feeds into the services provided and makes new technologies available to the wider community. EMBL is very active in developing European and world-wide standards and protocols to be used in data collection and representation and often represents Europe in international consortia, for example for the integration of data resources across the world.

2.1 Structural biology

Strategic Objective: *Lead the provision of synchrotron radiation and integrated structural biology facilities in Europe.*

EMBL Hamburg and EMBL Grenoble both provide a broad palette of cutting-edge services to structural biology users. These are centred on synchrotron and neutron source-associated beamlines and have recently been extended to encompass up- and downstream aspects of sample preparation and data processing. Together, the two synchrotron sites have provided access for around 3000 users in every year of the current Indicative Scheme. The future of EMBL's structural biology services will be greatly enhanced by several ongoing activities. First, by the newly constructed PETRA-III storage ring at the German Synchrotron Radiation Facility (DESY) and the associated new integrated facilities for structural biology 'EMBL@PETRA3' that EMBL is constructing there; second, by EMBL's participation in the Centre for Structural Systems Biology in Hamburg (see Section G); and third, by the upgrade of the European Synchrotron Radiation Facility (ESRF) ring and beamlines as well as joint initiatives in the context of the Partnership for Structural Biology in Grenoble. With the new automated, integrated facilities at both sites it will be possible to host many more structural biology projects, and, more importantly, to host projects that are more difficult and have higher technical demands. This will maintain EMBL's leading role as a provider of access to synchrotron radiation for structural biologists in Europe.

2.2 Data resources

Strategic Objective: *Provide Europe with the biological data that serves basic research and innovation in biology, health and agriculture.*

The European Bioinformatics Institute (EMBL-EBI) offers Europe's most extensive and most widely used biomolecular database collection. The user community for EMBL-EBI data resources is both enormous and global, with roughly four million website visits per day at the beginning of 2010 and users from over 300 000 different locations per month, a significant number of whom work in medicine and industry. Ongoing major challenges for our bioinformatics services include finding efficient ways to keep up with the rapidly growing quantities of data produced by high-throughput methods and integrating heterogeneous data from different sources into seamless information repositories. The EBI combines its major service function with research programmes in software engineering and bioinformatics. Close interactions between research and service activities ensure that the service resources are consistently state-of-the-art and thus maximally useful. While EMBL as a whole obtains 25-30% of its annual funding through competitive external programmes, EMBL-EBI raises roughly 50% of its funding from external sources. There are many EMBL-EBI funders but the particularly generous support of the European Commission, the Wellcome Trust and the US National Institutes of Health, who provided 20.6%, 14.4% and 7.7% respectively of total cumulative funding of the EBI in 2007-2009, for which EMBL is extremely grateful, should be especially noted.

2.3 Core Facilities

Strategic Objective: *Provide excellent services with state-of-the-art equipment, user support and appropriate computational infrastructure through the Core Facilities.*

The scientific equipment at EMBL that is used by more than one research group is organised into shared facilities, which are operated and maintained by specialist service staff. These Core Facilities are designed to support internal and external research projects, to interact with industry – often as technology developers or B-testers –and therefore obtain optimum early access to new technologies, and to provide research scientists with state-of-the-art equipment. They help EMBL researchers to achieve ambitious goals at the forefront of today's life sciences, which often requires focusing diverse sets of expertise and multiple expensive technologies on a specific biological problem. The Core Facilities are an important instrument to recruit the most talented young scientists to EMBL and it is therefore crucial to maintain their instrumentation and know-how at the cutting-edge of science.

2.4 The future of European research infrastructures

As Europe's major provider of research infrastructures for the life sciences, EMBL is heavily involved in the intergovernmental European Strategy Forum on Research Infrastructures (ESFRI) process. EMBL scientists co-ordinate two of the ten ESFRI Biomedical Science projects and EMBL is a participant in no fewer than five of the remaining eight planned activities. Our input into these projects is in part scientific, but our unique experience in the organisation and management of European-scale life science research infrastructure means that we can also provide invaluable input on the governance and organisational models for new research infrastructures in the biomedical science field. Additionally, the EMBL leadership has been involved in many of the discussions of how to best select and organise world-leading research infrastructures for the European scientific community as well as in helping to generate the political support necessary to achieve this goal. The planning of some of the ESFRI projects will continue into the next Programme period whereas others will enter the construction phase. EMBL will have a central role in realising some of these infrastructures, particularly in the bioinformatics, biological imaging and structural biology areas (see Section H).

Strategic Objective: *As Europe's intergovernmental research organisation for life science research EMBL will be a major player in the construction and operation of the next generation of biomedical research infrastructures in Europe.*

ELIXIR, the European Life Sciences Infrastructure for Biological Information, is the ESFRI project in the field of bioinformatics and has the goal of developing the next generation of European biomedical data resources. As a crucial enabling and supporting resource, ELIXIR has a central position among all other biomedical research infrastructure projects. The intention is that EMBL-EBI will function as the central hub to link and integrate a number of ELIXIR nodes, which are distributed life science data resources located in the member states. The UK has already awarded a first tranche of funding to the EBI to enable it to begin to implement a new data centre, which will be at the heart of ELIXIR. A second funding application that will, among other things, allow the construction of a building to house the required personnel is currently being evaluated. It is now important that the EMBL member states provide the necessary support to enable EMBL-EBI to manage these new responsibilities, which are crucial for the future development of the biomedical sciences in Europe.

3. Training and inspiring the next generation of scientific stars

Providing advanced training to its staff and the European scientific community is EMBL's third mission. It implements this through a range of intramural and extramural training programmes that are tailored to the needs of both scientists and other staff categories at various stages of their careers. The major current challenge in this area stems from the increasing requirement for interdisciplinarity in the life sciences. EMBL is responding to this challenge in many ways. As described in other parts of the Introduction and below we have increased the diversity of our scientific staff – the most important providers of our training programmes – created Centres that provide training in widely used technologies, reorganised the courses taken by our PhD students and piloted a new scheme for interdisciplinary postdoctoral training.

3.1 Intramural Training

3.1.1 PhD students

Strategic Objective: *Maintain EMBL's International PhD Programme as a model for other PhD programmes.*

The EMBL International PhD Programme is Europe's 'best in class' PhD programme for the life sciences and has served as a model for similar programmes at science institutions throughout the world. It has an intake of roughly 50 PhD students per year. The students are carefully selected from a large multi-national pool of applicants that reached over 1200 applicants in 2008 and 2009. The successful candidates receive comprehensive training that promotes independence and results in greater than 95% of the intake successfully completing their PhD degree. The PhD training course, which takes place between October and December each year and involves the annual student intake and the entire EMBL faculty, networks the students as soon as they arrive at EMBL and allows them to become a driving force in the development of collaborations across the Laboratory. The content of the course has been completely reworked in the past two years with increased emphasis on the teaching of the concepts behind living systems and greater disciplinary breadth in the curriculum.

3.1.2 Postdoctoral fellows

Strategic Objective: *Implement and expand the EMBL Postdoc Programme fostering collaborative, interdisciplinary and outstanding science.*

During the last Programme we initiated a Postdoctoral Fellows Association. The postdoctoral fellows now organise a series of events including an annual retreat and training courses. In addition, we launched the EMBL Interdisciplinary Postdoc (EIPOD) Programme, which recruits postdoctoral fellows to work on interdisciplinary projects jointly co-ordinated by (at least) two supervisors with different scientific expertise and skill sets. The aim is to equip postdoctoral fellows with a broader range of skills to optimally prepare them for challenging future research careers. To date, the outcome of the initial phase of the EIPOD Programme has been excellent, allowing us to bring a new group of young innovators into EMBL each year who quickly develop cross-Unit contacts and exchanges. We intend to increase the numbers of EIPODs to roughly 25 per year during the next Programme.

3.1.3 Independent researchers

The typical EMBL Group or Team Leader comes to the Laboratory immediately after postdoctoral training and stays for a maximum of nine years. These young investigators are provided with training in transferable skills, receive scientific and professional mentoring from their senior colleagues and are also formally assessed once every four years by a panel of external experts as part of an independent review organised by the EMBL Scientific

Advisory Committee. In this way, they receive a wealth of useful input, advice and criticism from senior experts in their field. This, together with the input and support they obtain from within EMBL, ensures that they have a tremendous opportunity to launch their independent career in an optimal way and that they become a major asset when returning to the member states after completing their time at EMBL. The quality of the training provided by EMBL is reflected in the fact that our departing faculty moves into highly desirable senior positions in academia and industry on returning to the national systems.

3.1.4 Staff training

Strategic Objective: *Provide appropriate training to all staff to boost their performance and enhance their career progression at and after EMBL.*

To complement scientific training, EMBL started a limited formal programme of vocational training, the General Training and Development Programme, for all its scientific, administrative and support staff in 2008. We offer a variety of training courses from computer skills to language training to leadership and management expertise. This programme has been very popular with the staff – indeed it is enormously oversubscribed – and helps the Laboratory to fulfill its obligation to the member states by ensuring that all categories of staff members leave EMBL with a skill set that makes them attractive candidates for organisations and institutions in the national systems. This training programme is now proven and needs to be increased in size from its present modest beginnings.

3.2 Extramural training

Strategic Objective: *Deliver a cutting edge course and conference programme that enhances interactions between disciplines and more than doubles the current participants to make optimal use of EMBL's facilities.*

Our intramural training activities are complemented by a very active programme of EMBL-organised courses, workshops and conferences that take place at all of our sites and are mainly, but not exclusively, designed for scientific participants. Within Europe, EMBL is uniquely active in this area. In the period 2007-2009 more than 130 events were organised that attracted close to 10 000 participants. The huge demand for hands-on training in the use of the data resources at EMBL-EBI as well as the recent opening of the EMBL Advanced Training Centre (EMBL ATC), which is projected to increase the number of participants in Heidelberg events by two- to three-fold, will lead to a significant further increase in these activities in the next funding period. New initiatives that will be launched by 2011 include the EMBO/EMBL symposia and e-learning programmes at EMBL-EBI. There will be five to six EMBO/EMBL symposia per year at the EMBL ATC that will provide a platform to discuss and exchange ideas on challenging, forward-looking topics and new developments in the life sciences. The symposia will be selected, planned and funded together with EMBO.

3.3 Outreach

As a publicly funded research institute, EMBL strives to make the public aware of the value of life science research for society and human well-being. We see it as part of our mission to contribute to improving the public understanding of science and pursue a dialogue with a variety of target audiences, including the wider scientific community, journalists, students, teachers, decision-makers and the public.

3.3.1 Inputs into science education

Strategic Objective: *Promote the life sciences and their impact on society beyond the scientific community.*

The European Learning Laboratory for the Life Sciences has hosted more than 700 teachers from 22 countries on its training courses in molecular biology since its inception in 2003. With these efforts and 'Science in School', the interdisciplinary online journal for science teachers that EMBL publishes on behalf of our EIROforum partners (see below), we will continue to train teachers in multiple aspects of cutting-edge life science research and thereby help them to inspire the next generation of scientists in Europe's schools.

3.3.2 Science and Society

Since its inception in the 1990s, the EMBL Science and Society Programme has become an integral part of scientific life in the Laboratory. A variety of activities and events organised at EMBL bring together members of the life science community, scholars of other disciplines and members of the public, for discussion and communication extending beyond professional boundaries. Many of the EMBL Science and Society activities are open to the public, including a major conference organised each year together with EMBO and multiple smaller events that take place regularly at all EMBL sites. Other activities, including lectures given by invited speakers, are primarily targeted at the EMBL research community and aim to raise our awareness of the societal implications of research carried out at EMBL and by the life science research community.

4. Driving research, innovation and progress through technology development, interactions with industry and technology transfer

4.1 Technology development

Strategic Objective: *Develop new technologies driven by scientific needs to further improve EMBL's research performance.*

EMBL has a tradition of significant investment in technology development as a means to improving its research performance. In recent years this effort has included: the automated use of synchrotron beamlines for structural biology; new software approaches to the analysis of many sorts of data, including structural biology, imaging and bioinformatics methods development; novel methods of recombinant protein expression; the development of Digital Laser Scanning Light Microscopy; and light microscopy-based high-throughput screening of complex phenotypes (see Section C.1.3 for more detail and further examples).

In the next Programme there are several areas in which we will pursue technology development. The ongoing construction of new structural biology facilities at EMBL Hamburg as part of the PETRA-III project and in Grenoble in the course of the ESRF upgrade will provide opportunities for further developments of beamline instrumentation. In imaging, we will be active in developing new *in vivo* probes that report on biological functions in real time, methods for single-molecule imaging and in looking for new ways to bridge light microscopy and electron microscopy. Finally, we want to invest in nanotechnology-based methods that will allow *in vitro* biochemical experiments to be carried out at a higher throughput and enable single cell-level perturbation and analysis.

As these past and future examples of technology development make clear, our activities in this area are driven by the needs of our scientists, who often push the limits of technology to answer the questions in which they are interested. At EMBL, engineers and technology developers are embedded in the Research Units to facilitate direct exchange with biologists. New instrumentation is therefore customised to researchers' needs, so that research questions shape technology rather than the other way round. This leads to a mutual synergy through which research needs drive the technologists and technology developments enable novel approaches to research.

4.2 Technology transfer

Much of the technology and instrumentation that is developed, and many of the results of research at EMBL, is of potential use in innovation beyond the Laboratory. To allow the scientific community and society at large to benefit from technology development

Strategic Objective: *Actively engage in technology transfer activities through EMBLEM and make EMBL's discoveries and innovations available to society.*

and research discoveries made at EMBL, we actively engage in technology transfer. This is achieved through EMBL-Enterprise Management (EMBLEM), a company that is wholly owned by EMBL. EMBLEM sources inventions throughout EMBL and, where desirable, protects our intellectual property and actively engages in the development of EMBL inventions for transfer to the commercial sector. This can be achieved either by technology licensing or, in a small number of cases, by the creation of a spin-out company. Currently, EMBLEM manages an intellectual property rights portfolio of over 260 granted patents and patent applications and more than 90 copyrights and trademarks. During the current Programme, EMBLEM has been instrumental in the foundation of four start-up companies and has succeeded in becoming profitable at a very difficult time for the biotechnology market.

The environment for technology transfer is constantly evolving. EMBL produces a significant number of discoveries and inventions with the potential for commercial application. These applications are one way in which the member states can benefit from EMBL's research. However, research findings and inventions resulting from them need to be developed further and further toward proof of applicability in order to attract outside investors who will help by taking on the risk involved in their commercialisation. We therefore need to bridge this technology transfer gap with a 'Proof of Concept' fund (see Section F.2) if we wish to remain successful in this area.

4.3 Interactions with bio-industries

Strategic Objective: *Engage in formal and informal interactions with bio-industries to make available the expertise of EMBL scientists and service providers.*

Aside from its technology transfer activities, there are many different ways in which EMBL interacts with and supports European bio-industries. These include project-based collaborations or consultancy agreements between EMBL scientists and industry, the provision of research infrastructure, services and training to industry, joint technology development or β-testing of new technologies by EMBL scientists, and the participation in various European-scale public–private initiatives. It is our view that by working closely together with the life science industry, in partnership with other stakeholder groups, we can restore a competitive advantage to European researchers and enable better translation of fundamental research into new advances in medicine, health, and agriculture for the benefit of society.

5. Taking a leading role in the integration of life science research in Europe

As Europe's intergovernmental laboratory for molecular biology EMBL has a central and strategic leadership position in the European research landscape. It fulfills a multitude of different functions for its member states and the wider European community, and greatly contributes to the development and integration of research in Europe.

5.1 A centre of excellence and role model for life science research in Europe

EMBL is at the forefront of basic research and our scientists develop or test the cutting-edge technology and instrumentation required to pursue new trends, thereby paving the way for other researchers. Publishing a large number of high-quality papers, EMBL significantly contributes to the scientific knowledge base that drives and guides life science research worldwide. In addition, EMBL is a strong research partner for national organisations in the member states, Europe and beyond. In 2008 and 2009, EMBL Group and Team Leaders engaged in no fewer than 2000 external collaborations. The active participation of our staff in collaborative work is also reflected in the fact that more scientists from EMBL were involved in projects funded in the European Commission Framework Programme (FP) 6 and, to date, also in FP7 than from any other research organisation. In total, up to the end of 2009, EMBL researchers coordinated 29 and participated in over 130 FP projects. In addition EMBL interacts and collaborates with the European Commission in many areas of mutual interest to the benefit of the European life sciences.

Strategic Objective: *Remain a centre of excellence and serve as a role model for life science research in Europe.*

5.2 EMBL member states

EMBL maintains close relationships and interactions with its 20 member states and associate member state Australia**.** To further integrate European research efforts in the life sciences, the participation of as many European countries as possible in EMBL is desirable. We intend to continue discussions with the newer EU member states to encourage them to join EMBL and participate in and benefit from our unique network.

Strategic Objective: *Encourage all European countries to join EMBL.*

5.3 EMBL partnerships

Although collaborations mostly happen on the basis of individual projects and are driven by overlapping research interests, EMBL also engages in a number of formal partnerships with member state institutes with substantial vision and international orientation. The scope of these partnerships extends beyond research projects to include joint initiatives in technology development, service provision and training, and they ultimately promote the pursuit of scientific excellence throughout Europe.

The partnerships fall into two categories. There are local partnerships that grow up at all of our sites. Some of these have given rise to successful joint research and service activities, for example the Partnership for Structural Biology in Grenoble that engages three intergovernmental organisations (EMBL, ESRF, ILL), the Institut de Biologie Structurale (supported by the CNRS, CEA and Grenoble University Joseph Fourier (UJF)) and the EMBL-UJF-CNRS Unite Mixte Internationale for virus-host cell interactions. Another successful example is the Molecular Medicine Partnership Unit in Heidelberg that is located in the University Medical School and which houses several collaborative projects involving EMBL groups.

Strategic Objective: *Develop EMBL partnerships as platforms for scientific collaborations and exchange of know how.*

More ambitious, and with considerable potential for impact in our member state communities, are remote EMBL partnership units such as those created during the last Programme at the Centre for Genomic Regulation (CRG) in Barcelona, Spain, and the University of Umeå, Sweden, the University of Helsinki, Finland and the University of Oslo in Norway. These partnerships involve locally funded research units that are based on the EMBL model, i.e. with international recruitment of young independent investigators, fixed-term contracts, stringent external evaluation procedures and significant levels of both financial and infrastructure support for the young group leaders. EMBL provides help and advice at all stages of setting up these partnership units including participation in the crucial phase of recruitment of research group leaders. In this way, we provide a unique opportunity for national institutes to implement and take advantage of the successful EMBL model.

EMBL's function as a role model extends beyond its official partnerships. Many leading research institutes throughout Europe have requested help and advice in setting up PhD programmes, core facilities or technology-transfer activities. We are always happy to provide assistance to member state institutions in these or other areas. In addition, many of our senior scientists participate in advisory and review boards of member state institutions, research councils and at the European Commission and make their knowledge and experience available in this way.

5.4 EMBL alumni

Strategic Objective: *Grow the network of EMBL alumni, developing activities of mutual benefit.*

The EMBL turnover system means that there are is an extensive network of alumni, currently numbering more than 4600. Over 80% of EMBL alumni remain in the member states when they leave the Laboratory at the end of their fixed-term contract, acting as EMBL ambassadors and enriching the national systems with the experience and training gained at EMBL and benefiting from the international networks they have built during their time in the Laboratory. Our alumni are one of the major returns from EMBL to the member states and help us fulfill another important function: seeding Europe with top scientists. EMBL's alumni retain contact with EMBL and with other alumni through the Alumni Association and its national chapters, thereby building bridges between national scientific communities and linking them tightly to EMBL. In this way EMBL serves as a central meeting point and a platform for exchange for scientists in Europe.

5.5 EMBL visitors

Strategic Objective: *Make EMBL facilities and know-how available by hosting visitors at all sites.*

Aside from the many users of our research infrastructures, EMBL's Visitor Programme brings more than 400 visitors to EMBL every year to carry out collaborative projects, use and learn from our Core Facilities or acquire new techniques. In addition, as already described above, thousands of scientists attend EMBL-organised courses, conferences and workshops each year.

5.6 European initiatives and interest groups

Strategic Objective: *Play an active role in science policy through the interaction with decision-taking bodies and as a member of scientific interest groups such as EIROforum.*

EMBL participates in various European-level initiatives and is a member of several important interest groups through which it interacts with decision-taking bodies to help shape science policy. Apart from its active role in the ESFRI process, EMBL is a member of EIROforum, the joint organisation of seven intergovernmental research institutes that operate crucial research infrastructures (CERN, EFDA-JET, EMBL, ESA, ESO, ESRF and ILL). The exchange of information between the organisations is invaluable to all members, in particular in the area of research infrastructure strategy, management and planning. In addition to their important individual roles, common EIROforum activities and interests in science education and outreach benefit the entire European Research Area. EIROforum is an important conduit for discussion with the institutions of the European Union, as well as with the member states, in these areas of science and science policy. EMBL is also a founding member of the Initiative for Science in Europe, an organisation of European scientific societies and organisations, that contributes to important political debates on scientific topics.

C. Research

1. Backward look 2005-2009

1.1 Introduction

In total, EMBL scientists published 2246 publications in 2005-2009 (we chose this five-year period because this document was written in early 2010, in the middle of the 2007-2011 EMBL Programme). We will first illustrate the progress made in the Systems Biology framework set for the current EMBL Programme 2007-2011 by providing a selection of relevant research highlights from that time. Later in the document we present examples of EMBL's technology development and service activities from the same period.

1.2 Selected research highlights 2005-2009

In accordance with EMBL's long-term strategy, as first outlined in the 'Strategic Forward Look 2006-2015', EMBL's overall research theme for the period 2007-2011 has been Systems Biology, which we defined as 'the study of the connectivity between components of living matter and how this connectivity leads to emergent properties, the properties of a system not derivable by summing the properties of the system's components'.

Our approach to Systems Biology aimed to combine traditional reductionist biology with new large-scale, high-throughput technologies, integrate methods and expertise from different disciplines and complement experimental and computational approaches.

The first step to understanding the dynamic interactions within a system is the characterisation of its components through a quantitative description of their nature and properties. The main tools to generate these quantitative data were specified in the EMBL Programme 2007-2011 as 'omics' approach-

es, i.e. (functional) genomics and proteomics, imaging and bioinformatics and computational biology. With the help of specifically created EMBL Centres (for molecular and cellular imaging, for functional genomics and for computational biology) the use of these tools has been spread throughout EMBL and has become standard across its research Units.

Thanks to the implementation of these tools EMBL has been able to achieve significant progress towards the main goal of its Programme 2007-2011: the generation of quantitative data as the basis for an understanding of living beings. In the next Programme we will build on this foundation and focus on turning these data into information and ultimately into a better understanding of the complexity of living systems.

The following section provides a short overview of how far EMBL has come in its Systems Biology endeavour and presents a few selected research highlights of each of the four thematic areas that were outlined in the Programme 2007-2011. For a more comprehensive list of achievements, please refer to Appendix C.1.

1.2.1 Structural Biology

EMBL has a traditional strength in structural biology. Over the past five years many biomolecular structures have been solved that help to illuminate important biological problems. A combination of X-ray crystallography, small-angle X-ray scattering (SAXS), nuclear magnetic resonance (NMR), electron microscopy and tomography, proteomics and computational approaches has made structural analysis at different scales and functional levels possible. Comprehensive structural insight has been gained into a variety of processes. The first global analyses of entire proteomes and their organisation have been completed for two different organisms, the eukaryote *Saccharomyces cerevisiae* and the prokaryote *Mycoplasma pneumoniae*. The structures of close to 50 target proteins of the causative agent of tuberculosis, *Mycobacterium tuberculosis*, have been solved using X-ray crystallography as part of an EU-funded Structural Proteomics Project (XMTB). Some of them provide new starting points for future target-based drug discovery.

Forming a central hub for structural biology in Europe, EMBL groups have participated in a range of collaborations and international projects during the past five years, many of which were funded by the European Commission. Research groups from EMBL Heidelberg, Grenoble and Hamburg contributed their expertise in structural and computational methods to the European project 3D Repertoire, which aims in the longer term to obtain the 3D structures of the protein complexes found in yeast. EMBL Hamburg was a partner in the EU-funded XMTB project and EMBL Grenoble coordinated FLUPOL, a European project to determine the atomic structure of the influenza virus polymerase. Moreover, scientists at EMBL Hamburg and Grenoble shared their expertise in technology development in synchrotron-based structural biology with many European partners in the EC-funded projects BIOXHIT, a joint effort to develop and provide effective technology platforms for structural genomics, and SPINE (Structural Proteomics in Europe), a collaborative project to develop new methods and technologies for high-throughput structural biology.

A) Structural biology of the influenza virus polymerase

The influenza virus polymerase plays a central role in the virus life cycle and is a candidate molecular target for new anti-viral drugs. It is responsible for the replication and transcription of the viral RNA genome and its ability to adapt to interacting cellular host molecules is an important factor in inter-species transmission of the virus, notably between birds and mammals. The structure of the viral polymerase has proven intractable in the past, but thanks to new high-throughput methods for the expression of protein segments developed at EMBL Grenoble, EMBL scientists have recently made major advances in the structural understanding of the polymerase. They have identified and solved the high-resolution crystal structure of the two domains critical for the unique cap-snatching mode of transcription used by the virus. Both domains are potential targets for new anti-viral drugs and screens run by the EMBL Chemical Biology Core Facility have identified inhibitor leads that will be developed further in collaboration with a new start-up company established in Vienna to exploit these discoveries (see Section F.1 e)).

Dias A. et al. (2009) The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. Nature 458(7240):914-8.

Guilligay D. et al. (2008) The structural basis for mRNA cap-binding by influenza virus polymerase subunit PB2. Nat Struct Mol Biol 15(5):500-6.

Tarendeau F. et al. (2007) Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. Nat Struct Mol Biol 14(3):229-33.

Tarendeau F. et al. (2008) Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. PLoS Pathog 4(8).

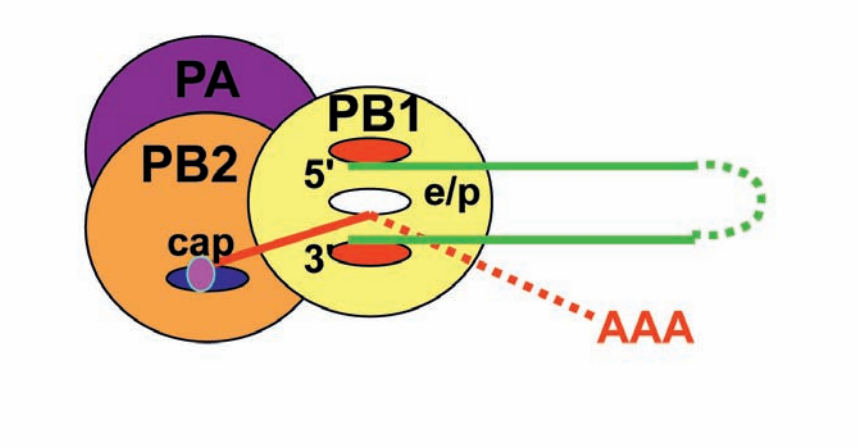


Figure C.1.1: The influenza virus polymerase is a complex consisting of many subunits.

B) Systems Biology of *Mycoplasma pneumoniae*

An integrated systems biology approach to the bacterium *M. pneumoniae*, the causative agent of atypical pneumonia, provided the first-ever blueprint of a minimal cell. Combining structural, biochemical and computational methods EMBL scientists in collaboration with the EMBL Systems Biology Partnership Unit at the Centre for Genome Regulation (CRG) in Barcelona, Spain, have characterised the proteome, transcriptome and metabolome of the bacterium. At all three levels the minimal bacterium, which has only 689 genes, was more complex and more similar to eukaryotes than expected. Many of the bacterium's molecules proved to be multifunctional, with metabolic enzymes catalysing multiple reactions, and many other proteins taking part in more than one protein complex. *M. pneumoniae* couples biological processes in space and time, with the cellular components required for consecutive steps in a biological process often being physically associated with one another.

Multiple groups in the Structural and Computational Biology Unit in Heidelberg investigated the *M. pneumoniae* proteome, applying a range of complementary techniques including mass spectroscopy, electron microscopy, electron tomography, NMR, X-ray crystallography and computational biology. They structurally defined many of the bacterial protein complexes and determined their location in the cell.

Yus E. et al. (2009) Impact of genome reduction on bacterial metabolism and its regulation. Science 326(5957):1263-8.

Guell M. et al. (2009) Transcriptome complexity in a genome-reduced bacterium. Science 326(5957):1268-71.

Kühner S. et al. (2009) Proteome organization in a genome-reduced bacterium. Science 326(5957):1235-1240.

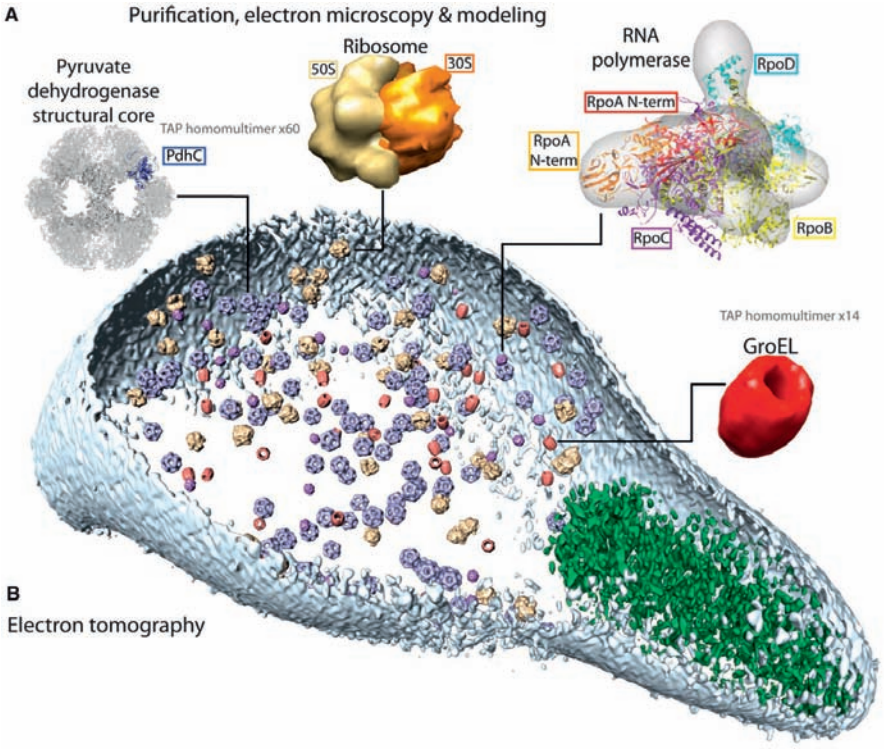


Figure C.1.2: Modelling the electron densities of individual molecules into an electron tomogram of an *M. pneumoniae* cell.

C) Structural investigation of HIV assembly in human cells

In 2009 researchers in the Structural and Computational Biology Unit in Heidelberg produced a 3D reconstruction of the human immunodeficiency virus (HIV), which shows the structure of the immature form of the virus in unprecedented detail. Immature HIV is a precursor of the infectious virus, the form that can cause AIDS. The study provides insights into the virus' maturation processes and describes how the protein coat that packages the viral genetic material assembles in human cells. Finding drugs that block this assembly process and prevent the virus from maturing into its infectious form is considered a promising therapeutic approach. The highest resolution 3D computer reconstruction images of the immature virus to date have been produced by a method called cryoelectron tomography, a technique with which a sample is instantly frozen in its natural state and then examined in an electron microscope from many angles. Images are obtained and assembled into an accurate 3D reconstruction by computer.

Briggs J.A. et al. (2009) Structure and assembly of immature HIV. PNAS 106(27):11090-5.

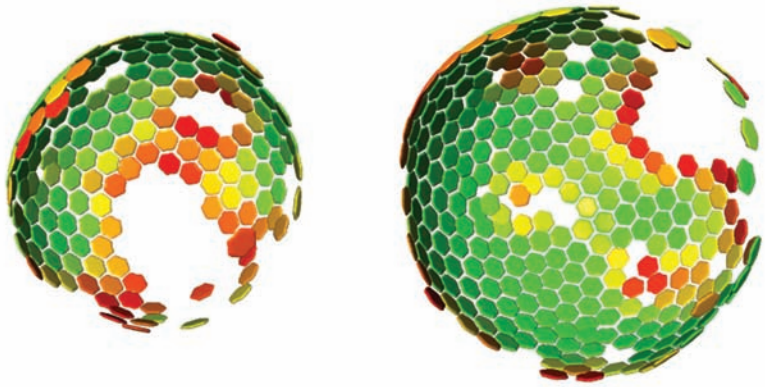


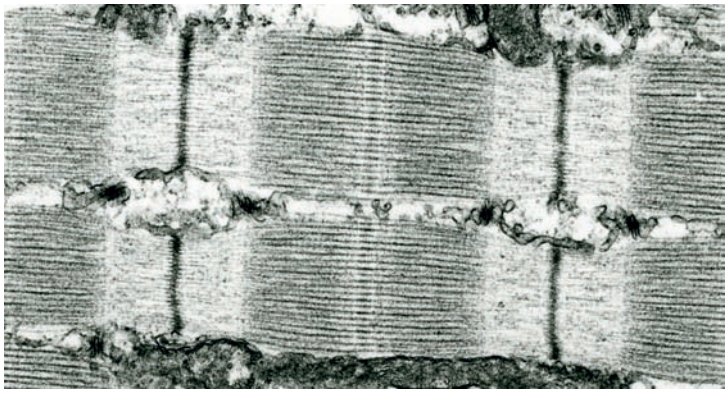
Figure C.1.3: Lattice maps for immature HIV particles. The 3D computer reconstruction shows the immature lattice of HIV that matures to form the protein shell of the infectious virus.

D) Insights into muscle assembly

For many years scientists have been investigating the assembly of structures called sarcomeres, the building blocks of muscle tissue that permit contraction and relaxation. Sarcomeres consist of many interacting proteins and are stacked end-to-end in long rows in muscle. Using X-ray crystallography researchers at EMBL Hamburg produced an atomic resolution image of the two sarcomere components titin and telethonin, which has shed light on how the basic units of muscle assemble.

Zou P. et al. (2006) Palindromic assembly of the giant muscle protein titin in the sarcomeric Z-disk. Nature 439(7073):229-33.

a)



b)

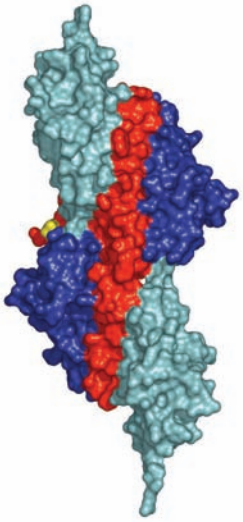


Figure C.1.4: a) Our muscles are composed of piston-like structures called sarcomeres, made up of many cells. Sarcomeres are linked by thick bands of proteins called Z-disks. Connections between specific molecules in these bands orchestrate how the entire structure is formed. b) The crystal structure of the two sarcomere components titin and telethonin.

1.2.2 Molecular, cellular and organismal biology

In order to obtain a deeper understanding of the properties of biological systems, throughout the Programme 2007-2011 scientists at EMBL have combined broad, genomic-level approaches with detailed analysis of individual functions at the molecular, cellular and organismal level.

1.2.2.1 Molecular analysis

At the molecular level major advances have been achieved in the areas of genome architecture and epigenetics, regulation of transcription, RNA metabolism and translational control through a variety of methods including functional genomics and proteomics, recombinant DNA-based approaches and biochemical analysis, and computer modelling.

E) Mapping a resistance factor that prevents the transmission of malaria parasites to mosquitoes

In a large-scale genomics approach in 2009 researchers in EMBL's Genome Biology Unit compared the genetic material of mosquitoes that are susceptible or resistant to infection by the malaria-causing parasite *Plasmodium berghei* and discovered that the major difference lies in a single gene. This gene, called TEP1, encodes a protein that binds to and promotes the killing of malaria parasites in the mosquito's midgut. Resistant mosquitoes have an allele of TEP1 that is different from the form found in non-resistant strains. Although this study focused on the parasite

that causes malaria in rodents, there is evidence that the same gene may also be involved in the mosquito's immune response to human malaria – a connection that may help to make malaria eradication programmes more effective.

Blandin S.A. et al. (2009) Dissecting the genetic basis of resistance to malaria parasites in *Anopheles gambiae*. Science 326(5949):147-50.

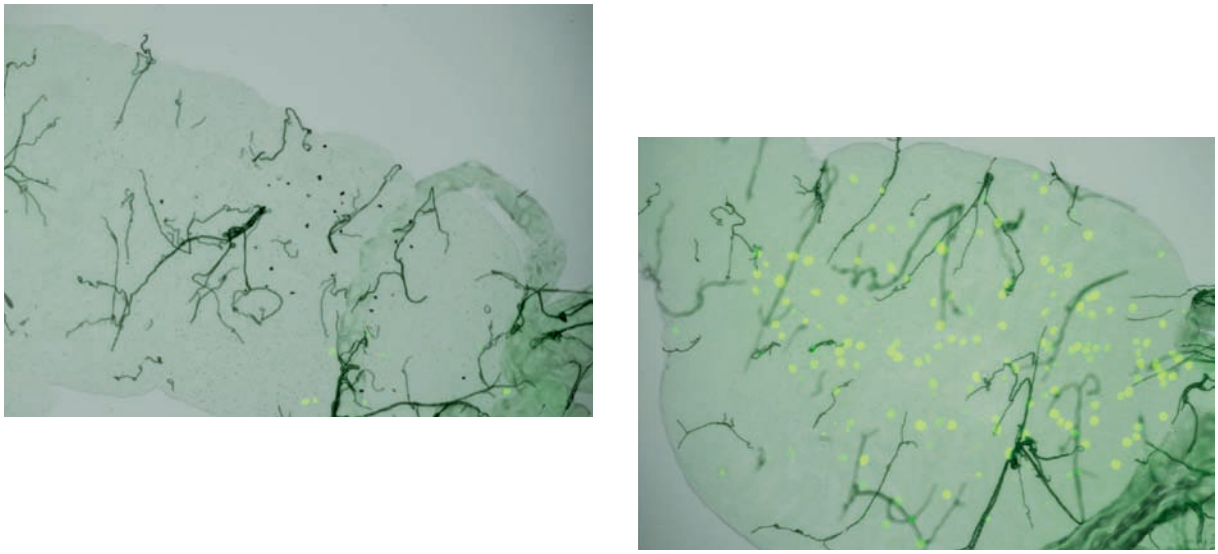


Figure C.1.5: These microscopy images show that, in *A. gambiae* mosquitoes, the different alleles of the TEP1 gene confer different degrees of resistance to malaria: the midgut of a mosquito whose only functional allele is the 'resistance' one (left) contains a number of dead malaria parasites (black dots), but very few live parasites (fluorescent green dots), whereas in another, genetically identical, mosquito with only the 'susceptibility' allele turned on (right), parasite survival was much higher.

F) Mapping gene regulatory networks and predicting their activity

To give each cell its unique identity during embryonic development, a special class of proteins called transcription factors makes sure the right genes are active at the right time. They act as molecular bookmarks by binding to regulatory regions on the DNA known as cis-regulatory modules (CRMs), thereby labelling the genes that should be transcribed. Applying a systems biology approach integrating genetics, biochemistry and computational biology, scientists in the Genome Biology Unit mapped the transcription factor network that controls the embryonic development of muscle tissue in *Drosophila*. Based on their high-resolution atlas of CRMs and the temporal and combinatorial patterns by which transcription factors bind to them, the researchers successfully predicted CRM spatio-temporal activity.

Sandmann T. et al. (2006) A temporal map of transcription factor activity: mef2 directly regulates target genes at all stages of muscle development. Dev Cell 10(6):797-807.

Sandmann T. et al. (2007) A core transcriptional network for early mesoderm development in *Drosophila melanogaster*. Genes Dev 21(4):436-49.

Wilczynski B. et al. (2010) Challenges for modeling global gene regulatory networks during development: Insights from Drosophila. Dev Biol 340(2):161-9.

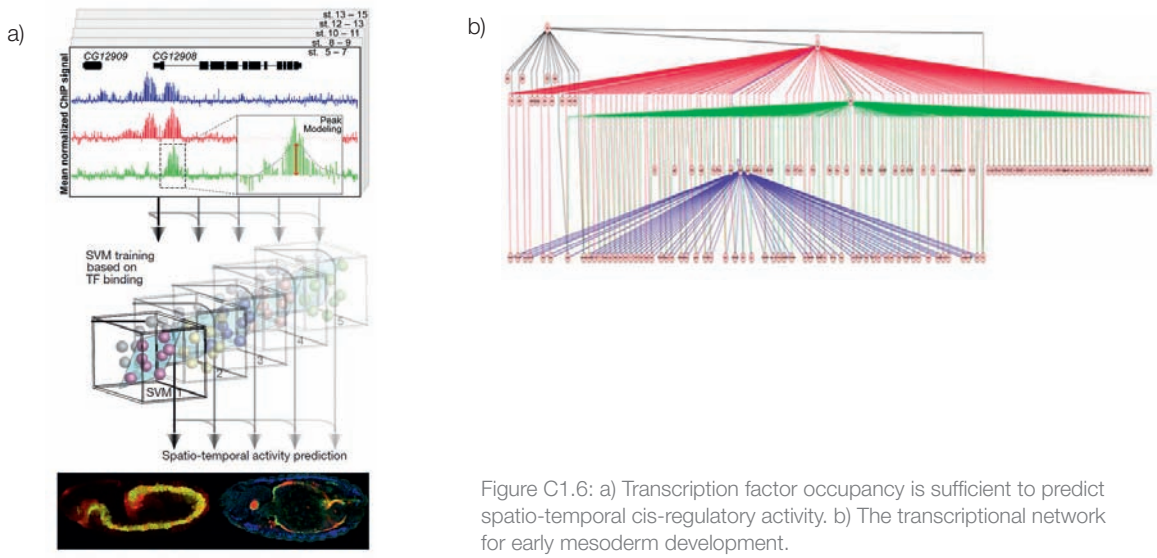


Figure C1.6: a) Transcription factor occupancy is sufficient to predict spatio-temporal cis-regulatory activity. b) The transcriptional network for early mesoderm development.

G) Deciphering how microRNAs control translation

More than 30% of human genes are under the control of small molecules called microRNAs. They prevent specific genes from being turned into protein and regulate many crucial processes such as cell division and development, but how they do so has remained unclear. In 2007 EMBL scientists developed an *in vitro* assay to investigate the mode of action of microRNAs and discovered that one way for them to achieve their effect is to block the initiation of translation, the earliest step in the process that turns genetic information stored on messenger RNAs into proteins. Specifically, they prevent the assembly of the molecular machinery required for translation. The new *in vitro* approach to studying microRNAs in action may pave the way to similar studies of human and other microRNAs. MicroRNAs play an important role in various diseases including cancer, diabetes and viral infections.

Thermann R. et al. (2007) Drosophila miR2 induces pseudo-polysomes and inhibits translation initiation. Nature 447:875-879.

1.2.2.2 Cell organisation

Major challenges at the cellular level have been the monitoring, dissection and measurement of the dynamics of cellular processes, including vesicle fusion and budding, intracellular transport, cell division, signalling events and the assembly and disassembly of subcellular structures such as the nuclear envelope, the cytoskeleton and the membrane system. Live imaging has been the method of choice to address many of these problems and imaging technology has been greatly advanced by EMBL scientists during the current Programme.

H) Phenotypic screening identifies genes involved in mitosis

As part of the international, EC-funded project MitoCheck, which is aimed at identifying all the proteins involved in cell division, or mitosis, EMBL scientists have developed a high-throughput phenotypic screening platform that combines gene silencing by RNA interference, time-lapse microscopy and computational image processing. The new technology was applied in a genome-wide phenotypic profiling of each of the ~21 000 human protein-coding genes. Each individual gene was silenced in a different group of cells, which were filmed for 48 hours under a microscope. Quantitative scoring by newly developed image-processing methods identified 600 human genes involved in diverse biological functions including cell division, migration and survival. The imaging data produced is freely available online for further analysis by scientists worldwide.

Neuman B. et al. (2010) Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. Nature 464(7289):721-7.

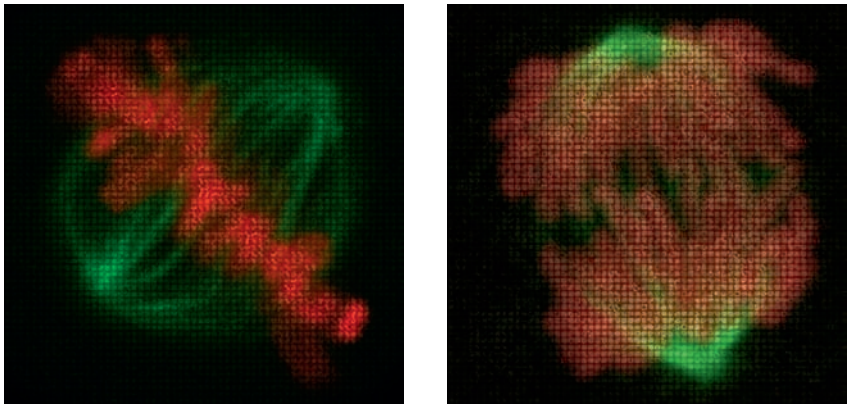


Figure C.1.7: Each of these large images of dividing cells is composed of several microscopy images of human cells in which different individual genes were silenced. The smaller images are placed according to genes' effects: images for genes that affect chromosomes make up the chromosomes (red/pink), while the mitotic spindle (green) is composed of images for genes that affect it.

I) Modelling microtubule behaviour

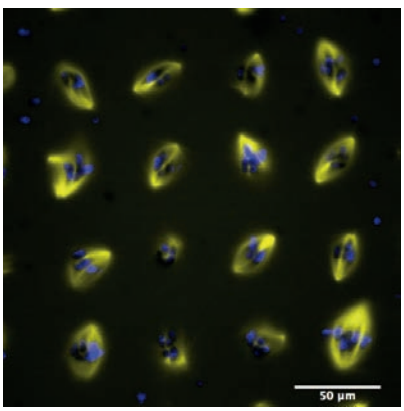
A prime example of systems biology at EMBL is work on the behaviour of microtubules that has been carried out in the Cell Biology and Biophysics Unit. Microtubules are long protein filaments that form part of the cytoskeleton and which, through their dynamic assembly and disassembly, influence many cellular processes. The best-studied cellular structure that consists of microtubules is the mitotic spindle, the apparatus that separates the chromosomes during cell division. Thanks to a new high-throughput spindle assembly assay recently developed by EMBL scientists, data on this important structure can be collected more efficiently and many spindles can be studied in parallel. Time-lapse images taken through a microscope allow microtubule behaviour over time and under varying conditions to be monitored. The quantitative data collected in this way are analysed with the help of a computer program called Cytosim, which EMBL researchers developed to simulate microtubule behaviour in a variety of scenarios. Cytosim models the collective behaviour of the fibres, their interaction with regulatory proteins and creates scenarios in which the molecular building blocks of the filaments are given hypothetical properties, and then explores whether this can lead to the construction of stable structures. The scenarios predicted by the simulation software can be tested experimentally and the new data generated are fed into the model to make it more realistic and improve its predictive power. With the help of Cytosim several key components, principles and dynamics of microtubule organisation have been deciphered in the past five years. Findings include the discovery that new microtubules are not generated randomly in the cell, but originate from already existing filaments and that their growth is influenced by physical forces acting on or within the cell.

Dinarina A. et al. (2009) Chromatin shapes the mitotic spindle. Cell 138(3):502-13.

Karsenti E. et al. (2006) Modelling microtubule patterns. Nat Cell Biol 11:1204-11.

Kozlowski C. et al. (2007) Cortical microtubule contacts position the spindle in *C. elegans* embryos. Cell 129(3):499-510.

a)



b)

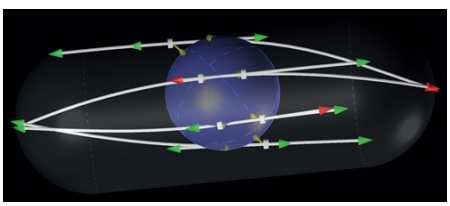


Figure C.1.8: Spindle arrays allow to study many spindles in parallel. b) The computer program Cytosim accurately models and simulates microtubule behaviour.

J) The first tomogram of a complete eukaryotic cell

In 2007 researchers at EMBL used electron tomography to generate the first 3D visualisation of a complete eukaryotic cell at a resolution high enough to resolve the cytoskeleton's precise architectural plan. The image of the unicellular organism fission yeast (*Schizosaccharomyces pombe*) revealed remarkable insights into the fine structure of the cytoskeleton as well as its interactions with other parts of the cell and serves as a reference map for all biologists interested in the architecture of the cell. Pictures were taken of sequential sections of a yeast cell from many different angles through an electron microscope and then combined into a 3D reconstruction with computational help. The principle is similar to that used to generate brain scans.

Höög J.L. et al. (2007) Organization of interphase microtubules in fission yeast analyzed by electron tomography. Dev Cell 12(3):349-61.

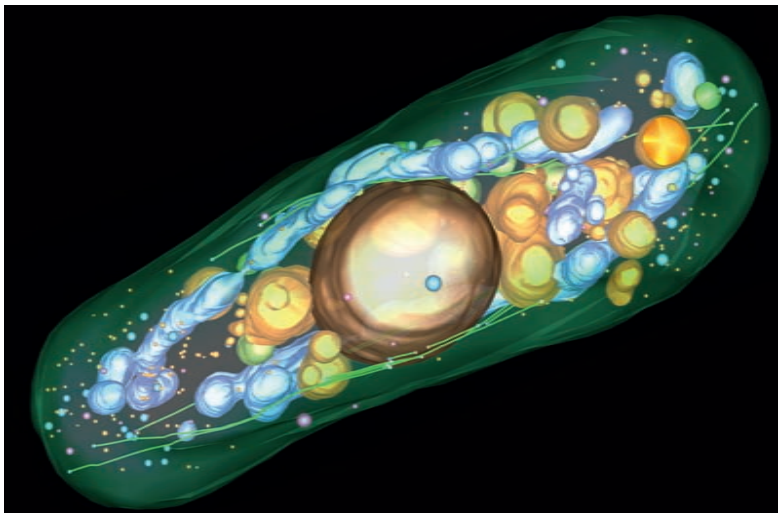


Figure C.1.9: The electron tomogram of a complete yeast cell reveals the cellular architecture. It shows plasma membrane, microtubules and light vacuoles (green), nucleus, dark vacuoles and dark vesicles (gold), mitochondria and large dark vesicles (blue) and light vesicles (pink).

1.2.2.3 Organismal Biology

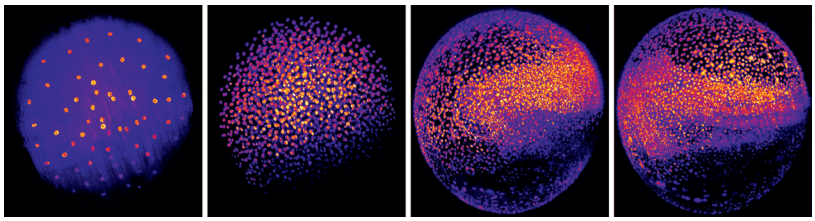
A combination of genetics, functional genomics and imaging shed light on a variety of biological phenomena at the organismal level, including the principles and dynamics of collective and single-cell migration, gene regulatory networks that govern embryonic development and the role of small regulatory RNAs in neurodegeneration, growth control, metabolism, germ cell development and regulatory feedback loops. Through the molecular comparison of cell types in different species and through studying the marine ragworm *Platynereis dumerilii*, a living fossil that has retained many ancestral traits over the course of evolution, scientists at EMBL gained insight into the evolution of the vertebrate eye, central nervous system, hypothalamus and microRNA expression in bilateria.

K) Digital zebrafish embryo

The first complete 3D developmental blueprint of a vertebrate was obtained by researchers in EMBL's Developmental Biology and Cell Biology and Biophysics Units in 2008. Technical advances in the single plane illumination microscope (SPIM), whose first prototype was developed at EMBL in 2002, allowed the movements of all the cells in a zebrafish embryo to be tracked for the first 24 hours of its life. The data were reconstructed into a 3D digital representation of the embryo. Movies of the digital embryo and the underlying database of millions of cell positions, divisions and tracks were made publicly available to provide a novel resource for research and scientific training.

Keller P.J. et al. (2008) Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. Science 322(5904):1065-9.

a)



b)

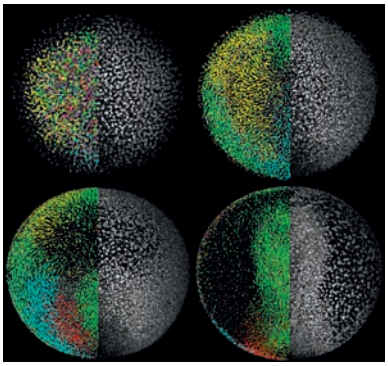


Figure C.1.10: a) The development of a zebrafish embryo observed with the SPIM. b) The montage shows the zebrafish digital embryo (left halves, colors encode movement directions of cells) and the microscopy data (right halves) at different time points in zebrafish development.

L) Uncovering the molecular basis of obesity

Why does the same diet make some of us gain more weight than others? Research by EMBL scientists suggests the answer could be a molecule called Bsx, which forms the molecular link between spontaneous physical activity and food intake. Mice lacking the molecule show less spontaneous physical activity, perceive hunger signals differently and have a lower concentration of feeding hormones in their brain than normal mice. As it is conserved across species, Bsx might be a promising target for controlling diet-induced obesity in humans.

Sakkou M. et al. (2007) A role for brain-specific homeobox factor Bsx in the control of hyperphagia and locomotory behavior. Cell Metab 5(6):450-63.

1.2.3 Genes and disease

Over the course of the current Programme EMBL's general focus on molecular mechanisms of disease has developed into a multidisciplinary, integrated approach that continues to support the talents of individual researchers whilst building on interactions between groups at EMBL and its partnerships. To further extend its molecular medicine network in 2007 EMBL established the Nordic Partnership for Molecular Medicine (Section G.2.2.7) in addition to the already existing Molecular Medicine Partnership Unit in Heidelberg (Section G.2.2.3).

Numerous approaches to investigate molecular mechanisms of disease have been applied during the current Programme. Functional genomics shed light on mitochondrial genes involved in disease and, in another study, fluorescent probes designed to monitor enzymatic activity in macrophages were used as an indicator of chronic inflammatory lung diseases.

Probably the most promising way to understand genetic disease is to generate and analyse mouse models. Over the past five years EMBL has developed new research activities that feature innovations in high-throughput technology for genomic, proteomic and imaging analysis of mouse mutants. Thematic strengths in mouse biology include developmental mechanisms, the neurobiology of behaviour and memory, regenerative biology and biomedical applications. In recent years mouse models have allowed insight into a wide range of disorders and conditions including stroke, leukaemia and other blood diseases, anxiety-related disorders, muscle degeneration, multiple sclerosis and defects in iron metabolism.

As an important hub for mouse biology in Europe, EMBL Monterotondo has played a central role in several EU-funded collaborative initiatives. These include EUMORPHIA, a project that developed standards for mouse phenotyping and a database to hold phenome data and EUMODIC, the European Mouse Disease Clinic that is undertaking a primary phenotypic assessment of up to 650 mouse mutant lines.

M) A mouse model for Sudden Infant Death Syndrome

Sudden infant death syndrome (SIDS) is a condition that unexpectedly and inexplicably takes the lives of seemingly healthy babies aged between one and 12 months. In 2008, researchers at EMBL Monterotondo developed a mouse model of SIDS that revealed that an imbalance of the neuronal signal serotonin in the brainstem is sufficient to cause sudden death in mice. This provides an important clue to the origin of SIDS that will be investigated further in future studies.

Audero E. et al. (2008) Sporadic autonomic dysregulation and death associated with excessive serotonin autoinhibition. Science 321(5885):130-3.

N) Identifying cholesterol-regulating genes

In the MMPU in Heidelberg, scientists at EMBL and at Heidelberg University have come a step closer to understanding how cholesterol levels are regulated. In a large-scale functional genomics study they identified 20 genes that are involved in this process, 12 of which were previously unknown. Besides giving scientists a better idea of where to look to uncover the mechanisms that ensure cholesterol balance is maintained, the discovery could lead to new treatments for cholesterol-related diseases. High levels of cholesterol in the bloodstream are a major risk factor for atherosclerosis and coronary heart disease, one of the current leading causes of death in developed countries.

Ludwiczek S. et al. (2007) Ca2+ channel blockers reverse iron overload by a novel mechanism via divalent metal transporter (DMT)-1. Nat Med 13(4):448-54.

O) Unravelling the molecular basis of memory

Combining genetic, electrophysiological and behavioural methods researchers at EMBL Monterotondo investigated the molecular basis of memory in living mice and identified both a molecule that is crucially involved in learning and the signalling pathway through which it affects memory. They engineered mouse strains with a defective version of a receptor molecule called TrkB. In the absence of TrkB, a signalling pathway in the brain that is important for memory formation could no longer be activated and the learning ability of the affected mice was impaired.

Gruart A. et al. (2007) Mutation at the TrkB PLCγ-docking site affects hippocampal LTP and associative learning in conscious mice. Learn Mem 14(1):54-62.

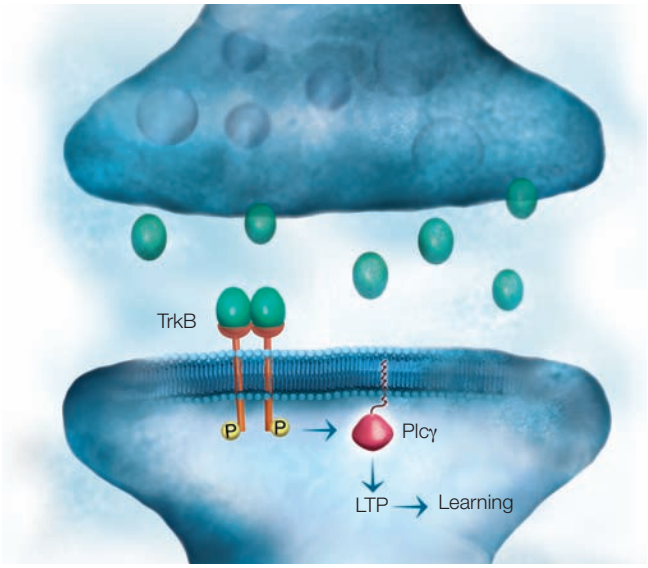


Figure C.1.12: Long-term potentiation (LTP), the strengthening of synapses between nerve cells in specific parts of the brain, is the molecular mechanism underlying learning. Some forms of LTP are triggered when a receptor called TrkB is activated by messenger molecules. Via a specific site the receptor activates a protein called PLC γ which then sets off a cascade that leads to the strengthening of the electrical signal that is passed from one cell to the other and allows mice to learn.

P) Comparing drug side effects

In 2008 researchers in the Structural and Computational Biology Unit developed a computational method to compare the side effects of different drugs and to predict how likely it is that the drugs act on the same target molecule. The computational approach revealed the molecular basis of many side effects, and it also has powerful therapeutic potential by hinting at new uses of marketed drugs. As an example it was found that the dementia drug Donepezil shares a target with the anti-depressant Venlafaxine, supporting the hypothesis that Donepezil could also be used to treat depression.

Campillos M. et al. (2008) Drug target identification using side-effect similarity. Science 321(5886):263-6.

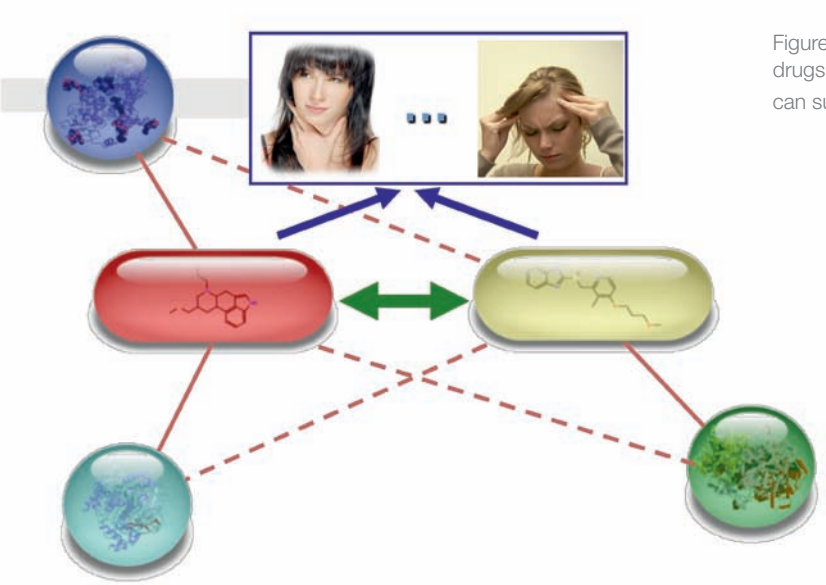


Figure C1.13: Similar side effects of different drugs hint at the same molecular target and can suggest new uses for old drugs.

1.2.4 Bioinformatics and Computational Biology

During the course of the current Programme, EMBL has maintained its position at the forefront of bioinformatics and computational biology worldwide, making major contributions in many research areas. In 'discovery biology', progress was made in understanding catalytic mechanisms, deducing evolutionary principles, elucidating the regulation of the bacterial transcriptome and metabolism, interpreting metagenomics data and modelling neuronal signaling, to cite a few examples. New algorithms were created for sequence alignment, structure classification, chemoinformatic methods for handling small molecules and text-mining tools were developed and used in numerous applications. New research data resources and tools to manage, integrate and visualise data were developed, for example to display the genome-wide transcriptional data produced by tiling arrays or for the analysis and visualisation of biomolecular networks, for gene expression atlases, for mathematical models of biological systems, for the classification of enzyme mechanisms and for exploiting evolutionary relationships between genes and genomes.

The bioinformatics groups within EMBL participated in many collaborative research studies worldwide, often as the lead partner. In Europe, EMBL coordinated three major bioinformatics research Networks of Excellence (BioSapiens, EMBRACE and ENFIN), promoting the distributed annotation of data, software access and harmonisation across Europe and the development of Systems Biology. Globally EMBL is leading the bioinformatics analysis of the ENCODE project on the functional annotation of the human genome, as well as the European human microbiome project. Increasingly EMBL's bioinformatics laboratories are involved in the design of experimental studies and actively 'commission' experimental data that are required to test their hypotheses, either through collaboration with other research groups or from commercial suppliers.

Q) Tree of life

In 1870 the German scientist Ernst Haeckel mapped the evolutionary relationships between plants and animals in the first 'tree of life'. Since then scientists have continuously redrawn and expanded the tree, adding microorganisms and using modern molecular data. Nevertheless, many parts of the tree have remained unclear. In 2006 researchers in the Structural and Computational Biology Unit at EMBL Heidelberg developed a computational method that resolved many of the open questions and produced what is probably the most accurate tree to date. The study identified 31 genes with clear relatives in 191 organisms, ranging from bacteria to humans, and allowed the reconstruction of their relationships. The high-resolution map of evolution also provided insights into the last common universal ancestor of all life on earth today.

Researchers at EMBL-EBI have developed another powerful tool that sheds light on evolutionary questions by performing highly accurate comparisons of genetic sequences between different species. The results challenge our understanding of how evolution occurs and suggest that sequence turnover is much more common than assumed.

Ciccarelli F.D. et al. (2006) Toward automatic reconstruction of a highly resolved tree of life. Science 312(5774):697.

Löytynoja A. et al. (2008) Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. Science 320:1632-1635.

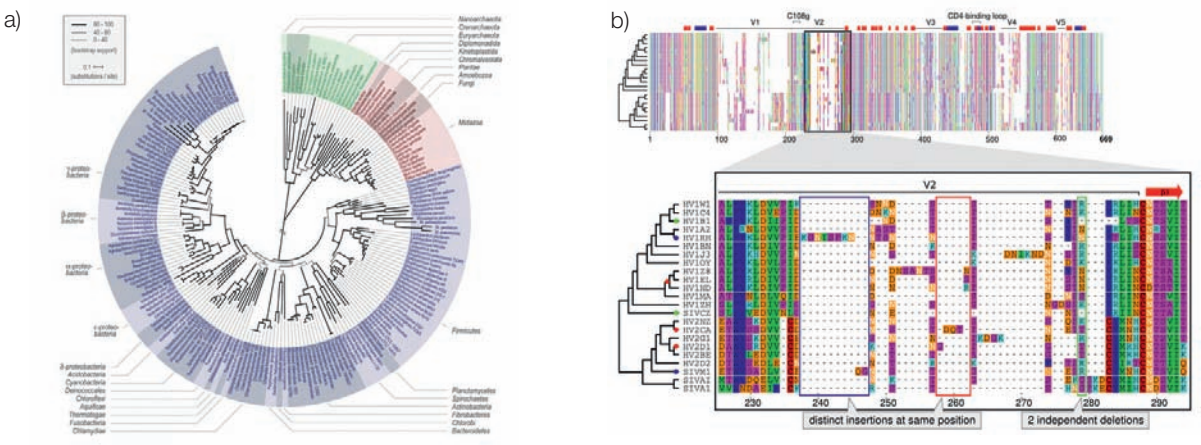


Figure C.1.14: a) The new tree of life. b) Sequence alignment according to the new method developed at EMBL-EBI.

R) ENCODE

Researchers at EMBL-EBI lead the analysis of the data produced in the context of the ENCyclopedia Of DNA Elements (ENCODE), which is an international research consortium organised by the NIH National Human Genome Research Institute. ENCODE aims to generate a comprehensive catalogue of all the components of the human genome that are crucial for biological function. In 2007 the EBI-led analysis and integration of 200 datasets generated by various high-throughput methods produced a 'parts list' of many categories of biologically functional elements in 1% of the human genome. The major findings include the discovery that most human DNA is transcribed into RNA and that these transcripts extensively overlap. This broad pattern of transcription challenges the long-standing view that the human genome consists of a small set of discrete genes, along with a vast amount of 'junk' DNA that is biologically inactive. The data indicate that the genome contains very few unused sequences; genes are just one of many types of DNA sequence that have a functional impact.

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

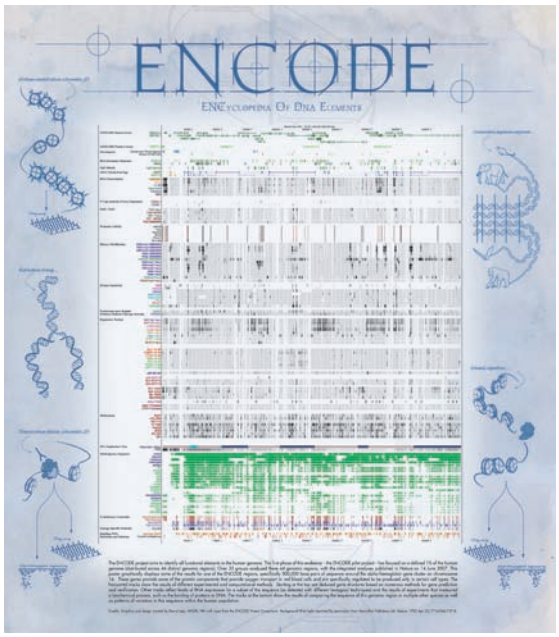


Figure C.1.15: Graphical display of some of the experimental results for 500,000 base pairs of sequence, in part of one of the 44 ENCODE regions spread around the genome.

S) Standards for systems biology

Over the past five years scientists at EMBL-EBI have made a major effort to develop standards and generalised ontologies for various areas of the life sciences including systems biology. Such standards facilitate the exchange and integration of different kinds of biological information. Successful examples are the Systems Biology Markup Language (SBML), a computer-readable format for representing models of biological processes, the Systems Biology Graphical Notation (SBGN), which standardises the graphical notation used in maps of biochemical and cellular processes, and the Systems Biology Ontology (SBO), a set of controlled vocabularies and ontologies for computational modelling. These standards are now widely applied in the systems biology community.

Le Novere N. et al. (2009) The Systems Biology Graphical Notation. Nat Biotechnol 27:735-741.

1.3 Selected technology development 2005-2009

To address pressing questions at the cutting edge of the life sciences, EMBL researchers often push the limits of technology. To advance the techniques required for their ambitious research plans scientists at EMBL engage in technology development across a broad range of biological disciplines. Many of the instruments and tools developed at EMBL are of benefit to the scientific community and are shared with other research institutes or made commercially available through technology transfer. This section features a selection of technologies developed at EMBL over the past five years that we are planning to build on over the course of the next EMBL Programme.

Imaging and microscopy

- **Single plane illumination microscopy/digital scanned laser light sheet microscopy (SPIM/DSLM):** The new microscope allows the study of large, living specimens from many different angles, under physiological conditions and with minimal harm to the specimen. In 2008, SPIM underwent a major upgrade with the introduction of illumination of the specimen by a thin laser sheet (DSLM) allowing living organisms to be studied over extended periods of time.

- **High-throughput screening systems for automated microscopy.**
- **Prototype of an automated high-content microscope** (Olympus, ScanR).
- **Automatic liquid dispenser for microscopy** (PRODesign, PROcellcare).
- **A set of FRET reporters** to characterise patients with lung inflammation.
- **Correlative LM/EM microscopy**: using a) laser etching on glass coverslips with cultured cells, and conventional transmission electron microscopy (TEM) after chemical fixation and b) carbon landmarks on sapphire disks, followed by cell monolayer cryofixation by high-pressure freezing.
- **Whole mount reflection confocal laser scanning microscopy**: Taking advantage of the reflective properties of NBT-BCIP crystals, this protocol visualises highly sensitive, non-fluorescent whole mount *in situ* staining by confocal microscopy.
- **Tools to visualise neuronal apoptosis in real time** (mitochondrial permeability and effector caspases) in a living organism.
- **Genetically encoded live-cell imaging tools** for DNA-damage mediated activation of PARP enzymes.

Chemical Biology

- **New *in vivo* cross-linking techniques.**
- **ChemPALM**: a new method to determine at the single molecule level where a chemical reaction takes place in living cells.
- **Implementation of transition metal catalysis in the living cell**: This methodology will be used to convert macrophages into drug-producing cells that deliver active compound at the site of inflammation.
- **Photoactivatable lipid derivatives and tools for *in vivo* lipid labelling.**
- **Large-scale protein-lipid interaction screens** enabling new types of network research.
- New method for solid-phase synthesis for phosphoinositides.
- **New biochemical purification method** to identify and characterise ADP-ribosylated proteomes.
- **Development and validation of glutathione alphascreen beads** for use as a general method in high-throughput chemical screening.

Genetics, genomics and proteomics

- **New *in vivo* transposition-recombineering approach** that facilitates and speeds up the development of mouse models of human chromosomal aneuploidies. This comprises a resource of more than a hundred mouse lines with mapped insertions, providing entry points to generate duplications/deletions that cover about 10% of the mouse genome.
- **Novel *cre* lines based on BAC-transgenesis**: Owing to a high demand for these specific lines in the field of neurobiology they have been shared with a number of other laboratories worldwide.
- **TRAWLER**: A *de novo* regulatory motif discovery pipeline for chromatin immunoprecipiation.
- **Tiling array for the yeast genome**: This has become the main commercially available tiling array for yeast (developed together with Affymetrix).
- **Whole mount *in silico* expression profiling**: A newly developed protocol uses image registration to align gene expression patterns to an averaged axonal scaffold. For the first time, this allows expression profiling of thousands of single neurons simultaneously in the entire brain of simple organisms.
- **A new method for *in vivo* cell-specific proteomics**: This is based on new genetic tools that allow the purification of cell-type specific complexes from the mouse brain.

Bioinformatics tools, software and standards

- **Cytosim**: A high-speed simulation tool applicable to many aspects of cell biology that has been applied to various microtubule self-organisation problems.
- **System for Information Management in BioMedical Studies (SIMBioMS):** SIMBioMS is a web-based open source software system for managing data and information in biomedical studies. It provides a solution for the collection, storage, management and retrieval of information about research subjects and biomedical samples, as well as experimental data obtained using a range of high-throughput technologies.
- **Bioconductor, open source software for analysis of genomic data:** EMBL researchers have been leading organisers and contributors to this user-extensible and programmable software system offering state-of-the-art analysis and interpretation of genomic data, including statistical algorithms, machine learning, visualisation, metadata, data integration, and modelling.
- **MIRIAM resources:** The Minimum Information Required in the Annotation of Models (MIRIAM) is a standard for the curation and annotation of computational models in Systems Biology.
- **Onion analysis pipeline:** To facilitate analysis of all protein sequences being deposited in the public domain (21 million at the end of 2009) using protein signatures (~100 K), it was necessary to create a new pipeline system to keep calculation time to a minimum.
- **Protein Identifier Cross-Reference Service (PICR):** Each major protein database uses its own conventions when assigning protein identifiers. The PICR service is a web application that provides interactive and programmatic access to a mapping algorithm that uses the UniProt Archive (UniParc) as a data warehouse to offer protein cross-references based on 100% sequence identity to proteins from over 84 distinct source databases loaded into UniParc. 3.2 M web hits/month in 2009.
- **Systems Biology Graphical Notation (SBGN):** This project is an effort to standardise the graphical notation used in maps of biochemical and cellular processes studied in Systems Biology.
- **Systems Biology Ontology (SBO):** This is a set of controlled vocabularies and ontologies tailored specifically for the kinds of problems being faced in Systems Biology, especially in the context of computational modelling.
- **EB-eye:** EB-eye is a high-performance, full-text search engine that can search across multiple biological data resources. It provides a single, easy and uniform way of accessing more than 64 distinct data resources hosted primarily at the EMBL-EBI.
- **New UniProt website**: The UniProt consortium is the main provider of protein sequence and annotation data for much of the world's life science community. The new http://www.uniprot.org website is the primary access point to this data and to documentation and basic tools for the data. These tools include full text and field-based text search, similarity search, multiple sequence alignment, batch retrieval and database identifier mapping.
- **Taverna:** The Taverna Workbench allows users to create, execute and share workflows built from component bioinformatics (and other domain) web and grid services.
- **Phenostat**: a visualisation and statistical tool for analysis of phenotyping data.

Bioinformatics databases

- **Atlas of Gene Expression**: An added value database providing information about gene expression in different cell types, organism parts, developmental stages, disease states, sample treatments, and other biological conditions.
- **BioModels Database:** A data resource that allows biologists to store, search, retrieve and analyse published mathematical models of biological interest. The models stored are annotated and linked to relevant data resources, such as publications, databases of compounds and pathways, controlled vocabularies, etc.

- **ChEMBL:** A new medicinal chemistry database of ca. 500 000 bioactive compounds, their quantitative properties and bioactivities (binding constants, pharmacology and ADMET, etc). The data are abstracted and curated from the primary scientific literature and then made available. ChEMBL was made possible by funding from the Wellcome Trust.
- **Ensembl Genomes:** This exploits and extends the Ensembl software framework (developed by EBI and the Wellcome Trust Sanger Institute for the analysis and display of vertebrate genomes) to five new domains: bacteria, protists, fungi, plants and invertebrate metazoa.
- **Ensembl Functional Genomics**: This database provides microarray annotation, a collection of genome-wide epigenetic data sets and the results of integrative analysis of genome function.
- **Proteomics Identifications Database (PRIDE):** The PRIDE database provides a standards-compliant repository for mass spectrometry-based identifications of proteins, peptides and post-translational modifications, together with the mass spectra that provide evidence for these identifications. Started in 2006, PRIDE now contains more than 2.5 million protein identifications based on 50 million mass spectra from direct data submissions. PRIDE as a central repository is a key element in overcoming the fragmentation of proteomics data and allowing validation of experimental results.
- **Scientific Literature Database (CiteXplore):** A database containing biomedical abstracts primarily from PubMed, Agricola, CiteSeer, and European patents.
- **Sequence Read Archive**: A data repository that provides access to public domain next-generation sequence data and supporting technology.
- **Trace Archive:** A scalable system that allows EBI to store large quantities of sequencing data. All archived data are stored in two independent archives that are located in separate data centres.
- **wwPDB Consortium:** Four institutes – two in the USA (RCSB, BMRB), one in Japan (PDBj) and one in Europe (PDB at EMBL-EBI) – joined forces to maintain the single worldwide archive of structural data on biomacromolecules. They are also developing a common deposition and annotation tool and have remediated the entire archive together. This provides a range of advantages: improved integrity of the structural archive (through remediation); consistent deposition and annotation procedures (through new common software); improved procedures at each of the sites (mutual validation); and better value for money (less duplication of effort and collaboration instead of competition).

Structural Biology

- **INPHARMA**: Ongoing development and validation of a method allowing the detection of the binding mode of small molecules on macromolecular receptors with NMR spectroscopy.
- **ESPRIT**: The ESPRIT construct screening technology is a unique process to identify soluble constructs of 'difficult-to-express' protein targets that resist the classical approach of bioinformatics and PCR cloning.
- **Software packages for automated data interpretation and structure determination from crystallographic data:**

  A) ARP/wARP is a software suite to build macromolecular models in X-ray crystallography electron density maps automatically, with a reproducible computational procedure. During 2005-2009, the package has been downloaded almost 8000 times. In addition, over 80 pharmaceutical research groups have obtained ARP/wARP software licenses.

  B) The program package ATSAS covers virtually all data processing and interpretation steps, from primary data reduction to 3D modelling methods, in SAXS. To date, ATSAS has been downloaded by over 2800 users from more than 1000 laboratories. In addition, in 2009 the package was used online for 6000 tasks. More than 50% of all biological solution SAXS publications cite ATSAS programs.

  C) Auto-Rickshaw provides a pipeline of software packages to allow automatic X-ray structure determination and data validation. Is has been used by 230 research organisations for determination of over 1000 novel structures (presently published in 60 peer-reviewed papers).

Technological developments for X-ray crystallography and small-angle scattering beamlines

- **Kappa diffractometer family for micrometre-sized crystals** (MD2x, MK3): In many challenging projects like the study of viruses or membrane proteins, only crystals of sizes in the micron range can be obtained. The MD2x diffractometer family allows crystals down to 10 mm size to be processed routinely. Among several advantages, reorienting crystals with the MK3 miniKappa goniometer head can reduce radiation damage (producing more data with a given dose of X-rays), allows efficient data collection over several micro-crystals, and to obtain precise data from Bijvoet pairs.
- **Automation of MX beamlines** (SPINE sample holder standard, SC3, C3D): A key contribution was made to the automation of MX beamlines. Firstly in driving the development of the European SPINE sample holder standard, secondly in developing an automatic sample changer (SC3), and finally in developing C3D, a software for centreing and aligning frozen crystals with an X-ray beam. The SC3 sample changer and C3D are implemented on all the ESRF MX beamlines and are commercially available.
- **Piezoelectric shutter:** On third-generation synchrotron beamlines, only a few minutes are necessary to collect a full dataset. Several thousand multi-pass images are taken in a day, thus reducing dramatically the lifetime of electromechanical shutters. A piezoelectric shutter was developed to satisfy the demand of high-throughput MX beamlines.
- **Crystal dehydration device** (HC1b): A dehydration device that adjusts the amount of water in crystals through an air stream of precise relative humidity has been developed. It is used to optimise the diffraction quality of poorly diffracting crystals by altering the amount of water in crystals and thus the packing of molecules in the crystals.
- **Automated sample environment for small angle X-ray scattering beamlines (BioSAXS**): We have developed an automated sample environment for the BioSAXS beamlines of ESRF and EMBL@PETRA3. Microvolumes of solutions stored in 96 wells plates can be processed automatically, taking full benefit of the X-ray beams of third-generation synchrotrons.
- **Beamline refurbishment and services at DORIS-III:** A new tuneable beamline (X12) for applications in macromolecular crystallography was constructed and opened in 2005. A novel fluorescence detection system (*X-Flash*) and software developments allow simultaneous analysis of the composition of the sample. The SAXS beamline X33 was entirely refurbished. The new beamline setup resulted in increased flux and higher beam stability as a precondition for unattended operation. Automated data acquisition at the SAXS X33 beamline has become possible owing to the installation of a liquid handling robot (in-house development) and a pipeline for automated SAXS data collection and analysis. Remote experiments in biological SAXS have been performed since summer 2009.
- **High-throughput crystallisation:** An automated high-throughput crystallisation facility with a storage capacity for 10 000 plates has been installed at EMBL Hamburg. A series of additional services, using microfluidics and offering advanced solutions for experiment optimisation have been added.
- **EMBL@PETRA3 – Planning, preparation and construction:** EMBL Hamburg has planned, prepared and started construction of a new integrated facility 'EMBL@PETRA3'. It will include three state-of-the art beamlines with applications in macromolecular crystallography and SAXS, a platform for sample preparation and characterisation, and another platform for automated data evaluation. It is scheduled to be operational in 2011.

2. Research Themes

2.1. Bridging dimensions: from molecules to cells to organisms

2.1.1 Bridging molecular and cellular resolution: from protein-protein interactions to networks in cells

In all living systems, structural organisation is intimately linked to function. The reason for this is that living systems must be able to organise themselves into functional states. Individual cellular components, such as proteins, RNAs, lipids and other molecules, often have intrinsic functions. However, these functions are only useful if they are carried out in the correct place at the correct time and in conjunction with other components of the same system. Major drivers of spatio-temporal organisation in living systems are stereo-specific interactions between components of the system and the dynamics and regulation of these interactions over time. Through such interactions molecules themselves, but also entire molecular ensembles, can acquire new functions and properties. Understanding these so-called emergent properties is a central goal of systems biology and dissecting molecular interactions is a prerequisite to achieving this goal.

Structural biology provides a powerful set of tools that allows us to investigate the interactions between biological (macro)molecules at different ranges of resolution. In addition, to gain insight into interaction dynamics requires both biochemical tools – for instance, to measure the affinities between molecules and to determine the mechanisms by which interactions are regulated, e.g. by protein modification – and lower resolution imaging techniques that can reveal the networks of interactions and how they change over time in a living cell. EMBL's plans for this area in the next five years therefore involve making use of our multi-disciplinary expertise and culture to bridge the lev-

Box C.2.1 Cellular systems biology of a minimal bacterium

EMBL scientists and collaborators from the EMBL-CRG Systems Biology Partnership Unit have generated the first blueprint of a minimal cell by conducting a comprehensive and quantitative analysis of the proteome, metabolic network, and transcriptome of the human pathogen *Mycoplasma pneumoniae* (Section C.1.2). This dataset creates unprecedented opportunities for systems biology. *M. pneumoniae's* simple genome makes it amenable to global studies, but in terms of molecular organisation and interactions the bacterium shares many features with more complex cells, making it a very useful model organism. The comprehensive characterisation of the bacterium's protein, RNA and metabolite household will serve as the basis for further interdisciplinary efforts at EMBL involving several groups from different research Units. Future research will focus on two aspects: i) increasing the resolution of the global overview with additional quantitative data on several aspects of cell organisation such as molecular interactions and ii) integration of the data ensemble to construct computational models of aspects of the organism that will allow prediction of various features of *M. pneumoniae* biology.

A variety of experimental approaches will be employed to complement the emerging scaffold of molecular architecture in *M. pneumoniae* with quantitative data on individual molecular processes. Combining genetics with quantitative mass spectrometry analysis of Mycoplasma grown under various conditions, networks of protein phosphorylation will be generated from which we hope to deduce general principles of post-translational regulation. Large-scale expression studies will address the regulatory function of metabolites and second messengers in transcription and identify as yet unknown or little-studied transcription regulators. To obtain a more general understanding of how they affect proteome function, chemical biology approaches, including the use of affinity- or activity probes, will provide insights into the interactions between metabolites and proteins. This data will be integrated with data from structural analyses and computational modelling of molecular interactions to lead to a better understanding of the biomolecular recognition code, which in turn will aid the design of small molecule modulators to systematically perturb molecular interactions in vivo.

The knowledge gained through the global analysis of the *M. pneumoniae* proteome will allow EMBL scientists to extend current analyses to the structure of protein networks that have remained elusive in the past, such as the membrane proteome. Membrane proteins, especially multi-component complexes, are difficult to study with traditional structural biology techniques. Taking advantage of the relative simplicity of the model bacterium and of the powerful combination of mass spectrometry and cryo-electron microscopy EMBL scientists will for the first time comprehensively study the topology and dynamics of a membrane proteome.

The wealth of quantitative data produced in these ways will be integrated with the help of computational technologies and will form the basis for systemic models that simulate and hopefully accurately predict various aspects of the cellular biology of *M. pneumoniae*. Ultimately this will be complemented by synthetic biology approaches aimed at introducing novel genes or regulatory networks into the organism's genome to generate new phenotypes, such as a reduction in the pathogenicity of the bacterium.

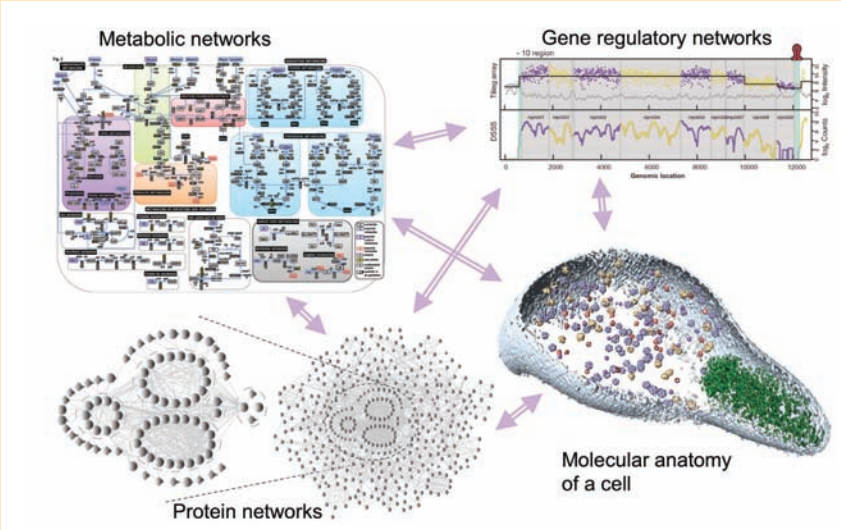


Figure C.2.1: Molecular anatomy of an *M. pneumoniae* cell.

els of atomic, molecular, multi-component, cellular and organismal resolution. We have made good progress towards this goal over the current five-year period (Section C.1.2) and in this document we describe ambitious future aims to build on this experience.

2.1.1.1 Structural Biology and Biochemistry

Structural biology has experienced spectacular advances in technology in the past two decades that encompass developments in synchrotron crystallography techniques, interpretation of small-angle X-ray scattering (SAXS) data, high-field nuclear magnetic resonance (NMR) and single particle and tomographic electron microscopy (EM) reconstruction methods. Together with the increasingly efficient production of recombinant proteins and complexes and the exponentially increasing availability of complete genome sequences, these advances give structural biologists freedom to work on virtually any system of choice.

One of the unique characteristics of EMBL is the breadth of and its long-term commitment to technology development and it is therefore no surprise that EMBL has been a major driving force in these advances (Section C1.3). Recent examples have been achieved in part through coordination or active participation in large EU-sponsored integrated projects (e.g. SPINE, BIOXHIT, 3D Repertoire, SAXIER, SPINE2-complexes). The technology development activities across the three structural biology research Units in Grenoble, Hamburg and Heidelberg are complementary. Due to their vicinity to major synchrotron facilities EMBL Grenoble and Hamburg have a strong focus on X-ray based methods and have made cutting-edge contributions to state-of-the-art, automated synchrotron crystallography, software development for X-ray based structural studies and novel protein expression technologies. The Structural and Computational Biology Unit in Heidelberg, supported by EMBL-EBI groups with expertise in computational structural biology, pursues an integrated approach using a broad range of techniques for biophysical, biochemical and computational characterisation of proteins and complexes. The integration and coordination of technology development and research efforts across sites is ensured through regular meetings of joint interest groups, such as the bilateral meetings of the beamline instrumentation groups of EMBL Hamburg and Grenoble. In addition, complementary external expertise is leveraged in three structural biology partnerships with member state institutions, the Partnership for Structural Biology and the Unit for Virus and Host Cell Interactions in Grenoble and the Partnership for Synchrotron Radiation Applications in Hamburg (Sections G.2.2.2 and G.2.2.5).

Despite the technological progress, *in vitro* structural studies of large multi-subunit complexes remain very challenging and EMBL will continue to pursue developments in this area as illustrated by the following examples:

In Grenoble, two unique platforms for protein expression will continue to be developed. These will also be made available to external member state users via the EU-funded PCUBE I3 grant. The ESPRIT platform is a high-throughput screening process that identifies soluble domains in large proteins that are poorly expressed. This technology will now be extended to handle binary complexes and further advances are expected. The Eukaryotic Expression Platform is a new technology that focuses on automated expression of multi-subunit complexes in insect cells. This will be coupled to state-of-the-art mass spectrometry to identify stable, soluble sub-complexes obtained by partial proteolysis and their subsequent automated re-cloning.

In synchrotron crystallography, a number of new developments will be associated with the opening of the new PETRA-III beamlines in Hamburg and the ESRF upgrade programme (described in Section D.2). An important development in Hamburg is the plan to build a Centre for Structural Systems Biology (CSSB) – scheduled to begin construction in 2011 – in order to provide complementary activities to the X-ray expertise being built up in the EMBL outstation and thus make optimal use of the PETRA-III development. Research carried out in the context of the CSSB will connect structural biology and systems biology approaches with key projects of biomedical relevance. Furthermore, an X-ray Free Electron Laser, which will provide X-rays with orders of magnitude more energy than a synchrotron ring, is being constructed in Hamburg and will come online in 2014. Pilot studies have been successfully carried out in Hamburg with the FLASH soft X-ray laser, but it is not yet clear how beams of this energy will be used in biology, just as it was not clear that synchrotron-based X-rays would be useful for biology when they were first produced. It would, however, be negligent not to examine the possibilities and EMBL, as the only biological laboratory on the site, is the obvious place to do so. EMBL will exploit both of these developments by extending its research activity in Hamburg.

In SAXS, a method that determines the overall shape of macromolecules and is especially suited for large complexes, the planned developments will enable higher quality data to be collected from smaller volumes and more dilute

samples. This will make more complex systems accessible to structural study. High photon fluxes in combination with high-throughput and micro-fluidic sample handling will pave the way for time-resolved structural studies in solution using SAXS.

In NMR, we will focus on studying interactions between components of complexes. This will involve hybrid approaches that use a combination of selective NMR-derived distance constraints (potentially also from solid-state NMR) together with SAXS data, biochemically derived distance constraints and molecular modelling to characterise ligand binding or to derive models of complexes assembled from structures of sub-complexes and individual components. More extensive ligand-binding studies are a crucial step towards gaining a better understanding of small molecules and metabolites, their interactions with proteins and their regulatory functions in molecular networks. In combination with mass spectrometry approaches NMR provides a powerful tool to map metabolic processes in the context of a cell's "metabolome". Tracking metabolite fluxes *in vivo* will be an important future goal.

In EM, we will continue to expand throughput and data quality through automated data acquisition and improved instrumentation. This will not only yield EM reconstructions at higher resolution, where the number of particles of a given sample analysed is limiting, but will also allow EM analysis to be extended to more heterogeneous samples, expanding the scope of the method to complexes that are more difficult to prepare. The interface between high-resolution EM and tomography will also be a particular focus for further development.

In computational structural biology, new tools for automatically fitting ligands to electron densities will be developed, as will improved methods to validate small molecule ligands bound to large biomolecules. Methods development for the prediction of protein function from structure will continue, based on evolutionary relationships as well as the study of physico-chemical principles and the growing number of known protein structures (currently 60 000). Areas in which we expect progress include enzyme catalytic mechanisms and membrane protein channel specificity as well as the prediction of the structural and functional consequences of human genetic variation on proteins. Tools to exploit structure–activity relationship data from the new ChEMBL database at EMBL-EBI will link structures to drug modes of action, supporting the *in silico* design of improved small molecule probes and drugs.

2.1.1.2 Cellular Structural Biology

The purpose of further technology development is to support a profound, biology-driven change in the focus of structural biology. Biologists now strive to combine *in vitro*-derived structural and biochemical information with functional studies of the same proteins and complexes in the context of the living cell. The goal of this marriage of structural biology and cell biology, or *cellular structural biology*, is the structure-based, mechanistic understanding of cellular processes. This move is aided by the addition of cryo-EM tomography, proteomics/mass spectroscopy, fluorescence light microscopy, various computational methods and chemical biology approaches to the palette of techniques available to structural biology laboratories.

EMBL's culture of interdisciplinarity, collaboration and shared interests puts us in an extremely favourable position to pursue cellular structural biology. We have reinforced our strength in this area through focused recruitment in 2010 and 2011 in high-resolution EM and EM tomography, single molecule fluorescence-based analysis methods, super-resolution light microscopy and quantitative mass spectrometry, as well as by making use of the EMBL Interdisciplinary Postdoctoral (EIPOD) Programme (described in Section E.3) to promote interdisciplinary projects. The integration of *in vitro* and cellular structural biology approaches that will result from this forward planning has tremendous potential for providing insight into the detailed molecular mechanisms of certain cellular processes. A range of topics are under study and we confidently expect major new insights in diverse fields such as transcription in the context of chromatin, cellular and viral RNA metabolism, translocation across membranes via the nuclear pore complex (Box C.2.4) or the peroxisomal translocon, and cellular- and viral-induced vesicle budding (Box C.2.2).

Box C.2.2 Correlative microscopy of viral and cellular budding systems

EMBL scientists will continue to develop and apply novel correlative microscopy techniques that allow the study of cellular events at high resolution over time. Electron microscopy and tomography can be used to map the spatial organisation of all biomolecular complexes in a cell in three dimensions and allow detailed measurements of organelle dimensions, inter-object distances, membrane curvature and other crucial cellular parameters. Complementing these static snapshots with correlative light microscopy adds the dimension of time and makes dynamic cellular processes amenable to microscopy at nanometer resolution.

The new methods, which rely heavily on state-of-the-art light and electron microscopy equipment available in the Core Facilities, will be used to produce high-resolution three-dimensional movies of cellular and viral budding systems. Vesicle budding is involved in various cellular processes, such as the trafficking of proteins and lipids between different cell compartments or the release of signalling molecules such as neurotransmitters. Budding also plays an important role in the spread of many enveloped animal viruses, including human immunodeficiency virus (HIV) and influenza virus. Budding events are complex processes that require the coordination of a dynamic, multi-component protein machinery that reshapes the lipid bilayer to generate vesicles. Three-dimensional movies of budding events will reveal the structure and composition of this protein machinery, as well as the topology of the lipid bilayer, at each time point during the process. The stage of budding and the quantitative protein composition of a large number of individual buds will be monitored by fluorescence microscopy, and the three-dimensional structure of the same buds will be imaged using electron tomography. Sub-tomogram averaging techniques will resolve repetitive units, such as the coat proteins of the bud, to a resolution at which the atomic structures of protein domains can be modelled into the density, allowing the arrangement of the proteins over the surface of the bud to be precisely defined.

The high-resolution, four-dimensional data obtained from correlative microscopy will be used to build computational models of the budding process. These models will generate new hypotheses to be tested experimentally. This systems biology approach will gradually build up a comprehensive understanding of the function of the vesicle budding machinery and at the same time reveal general features of the transient assembly of multi-component complexes to catalyse cellular events. In future, similar correlative microscopy approaches can be applied to describe various other cellular processes structurally and quantitatively.

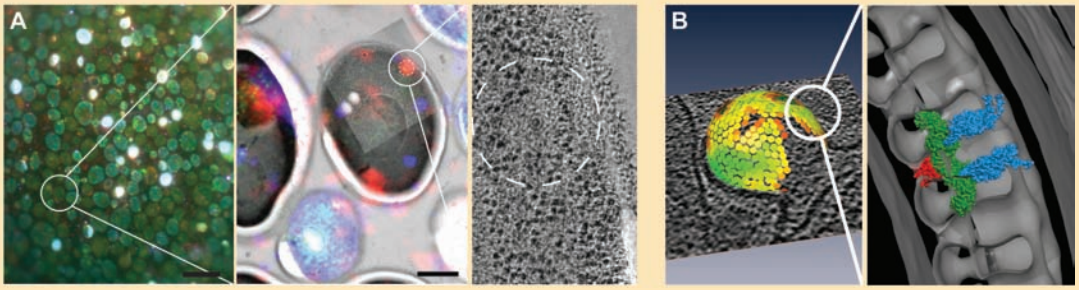


Figure C.2.2: A. Correlative microscopy allows precise identification of a fluorescent endocytic event imaged in a light microscope in an electron tomogram. The combination of budding components present (represented by different colours) can then be used to assign the stage of budding. B. From a 3D reconstruction of a budding virus particle, the positions of the viral structural proteins can be mapped, and a 3D reconstruction of the virus protein lattice can be calculated.

2.1.2 From cells to organisms: dynamic organisation and imaging

2.1.2.1 Cell Biology and Biophysics is next-generation Systems Biology

The basic unit of life is the cell. To understand biological systems at the molecular level, it is therefore essential to investigate the sub-cellular and cellular organisation and dynamics of the biological macromolecules that form the cell and determine their shape and function. To date, protein and signalling networks have been largely predicted in two dimensions based on systematic analyses using functional genomics approaches. However, the biological processes mediated by these networks do not occur statically in two dimensions but rather in four: three-dimensional

space and time. The challenge is therefore to move from mechanistic information on individual proteins and low-resolution analysis of the dynamics of cellular structures and organelles to understand the function of protein complexes, supramolecular machines and their self-organised networks to obtain a mechanistic view of the dynamic inner workings of the cell. The next generation of systems biology must therefore capture data and build models in four dimensions, reflecting the physiological conditions and dynamics of the biological system under study. The methods required to tackle these problems include new imaging technologies and probes that come from physics and chemistry. EMBL has recruited experts working in these fields and is thus prepared for the next steps.

2.1.2.2 New imaging methods to quantify the dynamic action of molecules inside the cell

The methods that can deliver the data that meet these requirements with sufficient sensitivity, as well as sufficient spatial and temporal resolution, in living cells are largely based on a host of non-invasive light microscopy imaging technologies that exploit molecularly defined fluorescence reporters. The ongoing revolution in imaging methods and probe development provides access to biophysical and biochemical parameters that describe biological macromolecules as well as small molecules in their natural habitat. These parameters include localisation (down to a resolution of tens of nanometers and single molecules), abundance, half-life, diffusion coefficient and flux, interactions and enzymatic activity. Exploiting these unprecedentedly powerful imaging tools requires an interdisciplinary combination of expertise in cell biology, biophysics and chemistry that combine to meet the needs for new instrumentation, data analysis methods and probe development. EMBL has been at the forefront of technology developments in these areas in the current Scientific Programme, through the development of fundamentally new imaging technologies including selective plane illumination microscopy (SPIM, DLSM) as well as the systematic development of many imaging modalities into high-throughput microscopy (HTM) for systems biology (Section C.1.3). It will be key for us to maintain strength in technology development and we have consequently appointed new groups trained in physics, who combine modelling and instrumentation expertise, as well as building a new focus in chemistry and probe development (Box C.2.3)

2.1.2.3 From gene networks to protein-interaction networks in living cells

An entry point into cellular systems biology is the systematic identification of proteins required for a cellular function, i.e. the definition of the components of a functional network. Thanks to technology developed at EMBL during the last Programme, such experiments can be carried out in human cells as shown, for example, by the genome-wide RNAi screens for genes required for cell division, protein secretion from the endoplasmic reticulum to the cell surface and for DNA repair. Microscopy-based RNAi screening is now available to EMBL groups through the Advanced Light Microscopy and Genomics Core Facilities (Sections D.3.1.6 and D.1.1.1). These facilities are also open to scientists from EMBL member states, who can take advantage of the facilities to either "clone" the technology to their own home institute with our help or, to the extent made possible by our limited capacity, by use of the EMBL set-ups. Mined by computational image analysis methods also developed at EMBL, these rich datasets reveal networks of genes related by the phenotypic signature produced by their absence, and thus generate a first hypothesis for the identities of proteins that form complexes or networks to carry out a specific function.

Although microscopy phenotypes are very information-rich, their prediction of molecular function needs to be validated by direct analysis. The power of imaging methods now allows many aspects of functional analysis to be performed in the context of intact living cells. The first level of information that can be gained is dynamic protein localisation. The ability to express functional, fluorescently tagged transgenes, including their native *cis*-regulatory sequences, in human cells with concurrent knockdown of the untagged endogenous gene by RNAi enables the extension of this experimental system of *in vivo* functional studies in near-native conditions. The development of the quantitation of molecules by high-resolution three-dimensional microscopy calibrated with single molecule precision by fluorescence correlation spectroscopy (FCS), recently put in place at EMBL, will allow the concentration of proteins to be determined with high sub-cellular spatial and real-time temporal precision in living cells. Knowing the location of a protein in the cell, e.g. on the kinetochore, and when it is there, e.g. during mitosis, provides a solid first hypothesis on the biological process in which the protein is involved. An ongoing effort during the coming Programme will therefore be to compile a four-dimensional atlas of absolute protein concentrations inside the dividing human cell. Part of this effort will be externally funded by the EU-funded MitoSys project. As the need to quantify and localise molecules extends beyond proteins, researchers at EMBL are also developing new non-disruptive methods to measure small molecules, such as lipids and other metabolites, in living cells.

Box C.2.3 High-throughput functional imaging for Systems Biology

Current and future functional genomics technologies will provide rapid access to individual genomes and the putative functions of their encoded genes. With this information at hand, the next challenge will be to integrate the available information into a quantitative description of biological systems at the protein network level. This will pave the way for a better mechanistic understanding of fundamental cellular processes that, for example, lie behind organism development and disease.

Addressing these challenges, concerted efforts of EMBL groups have resulted in the development of microscopy-based functional and biophysical imaging technologies that not only allow the quantitative mapping of protein concentrations and dynamics but also provide spatially resolved analysis of protein activity and interaction in both single cells and, in favourable examples, in multi-cellular organisms. In parallel, high-throughput high-content microscopy has been developed and is currently being applied to several genome-wide siRNA or small-scale chemical screens to identify the genes involved in, for example, protein secretion or cell division. These screens return comprehensive information on how genes function in the processes under investigation and prepare the ground for the analysis of specific protein functions in living cells by systematic perturbation.

For the next step, the EMBL groups involved now aim to integrate specific functional imaging technologies, involving fluorescent reporters of protein function, with high-throughput microscopy. This is expected to allow for the first time the systematic acquisition of high-quality functional and biophysical imaging data with sufficient statistical power for quantitative analysis and will ultimately have the ability to cover analyses of tens to hundreds of proteins belonging to a specific protein network. To achieve this goal, it will be necessary to automate the whole workflow, including complex image acquisition protocols associated with e.g. photoactivation, fluorescence correlation spectroscopy (FCS and FCCS) or fluorescence resonance energy transfer (FRET) measurements and subsequent data analyses. In parallel, continuous and systematic development and improvement of robust chemical and genetically encoded fluorescence reporters monitoring biochemical reactions or protein activities are required. The development of such sensors is laborious and consequently probes are currently available for only very few enzymes. In future, EMBL scientists will develop strategies to generate more complete sets of sensors in order to validate hypotheses derived from systems biology. Sensor sets based on constructs that place several thousand human genes between a set of fluorescent proteins exhibiting FRET will be developed and made available to researchers throughout EMBL and in the member states. Prime areas for which probes will be produced are enzymatic signalling networks that control key cellular processes such as mitosis, nuclear biogenesis, secretion and endocytosis. Over the next years many of these sensors will also be expressed in transgenic mice to look at enzyme activity under pathological conditions in human disease models.

Advances in the areas of microscopy technology, automation and biosensor development will foster systems-level research in many areas of the life sciences, as they will facilitate functional imaging and make it accessible for the first time to a broad scientific community.

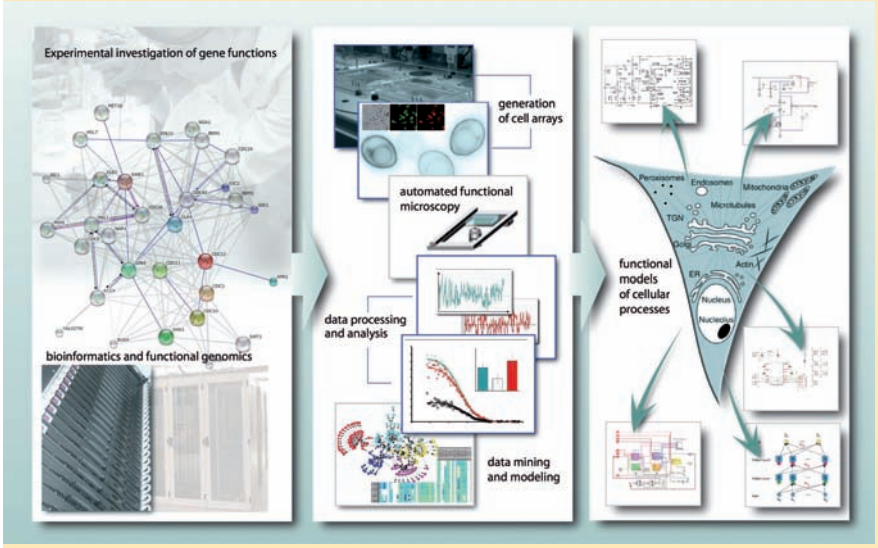


Figure C.2.3: Functional and quantitative analysis of protein networks. Proteomics, genomics and bioinformatics (left) deliver a wealth of information about protein networks. High throughput functional microscopic imaging platforms (middle) provide spatio-temporally resolved information indispensable to understand the properties and behaviour of protein interaction networks. Quantitative mathematical models and computer simulations (right) are then used to integrate the data and to generate testable hypothesis on the functioning of the networks.

The next challenge will be to systematically image protein interactions and activities at the physiologically relevant time and place inside living cells. Again, new imaging methods under development at EMBL will allow us to contribute to meeting these requirements. After tagging potential interaction partners with two different fluorescent proteins, fluorescence cross correlation spectroscopy (FCCS) can determine their interactions with single-molecule precision, giving us access to the percentage of interacting proteins as well as to the stoichiometry of the labelled proteins in the complex. Pioneered at EMBL by gene replacement in yeast to unravel the MAP kinase signalling pathway, this method is currently being automated and will be applied to human cell line collections to analyse interactions between pairs of genes required for mitosis.

In summary, automated imaging methods now allow us to move from genetic networks to localisation and interaction networks in living cells for proteins involved in any cellular function that can be visualised in the light microscope. Combined with computational data processing and mining, integrating the data from these efforts will lead to a next-generation systems biology database covering the core cellular functions that are being studied at EMBL, including cell division, cytoskeletal dynamics and membrane biology.

2.1.2.4 Molecular mechanism in a physiological context: biochemical reconstitution and single molecule activity imaging

When a cellular process has been analysed either with the above imaging pipeline or studied using genetic and/or biochemical systems, detailed knowledge about the relevant pathways and protein complexes will exist (Boxes C.2.2 and C.2.4). This sets the stage for the elucidation of the detailed molecular mechanisms of these processes. True molecular understanding requires biochemical reconstitution of the activity from a minimal set of components *in vitro.* To provide a recent example from work at EMBL, the mystery of how proteins manage to track along the highly dynamic growing tip of a single microtubule was solved by the *in vitro* reconstitution of the so-called plus-tip complex with a mixture of tubulin and (only) three purified proteins. Such experiments require specialised biochemical expertise and we are aware of the need to maintain strength in this area while building up the capacity to generate data via efficient high-throughput large-scale experiments. An example of a problem of this type whose solution seems to be in reach for the time period covered by the next Programme is how two very similar pentameric protein complexes – the cohesins and condensins, which form giant ring-like structures – can mediate linkages either between the DNA molecules of sister chromatids or within a single DNA molecule in an individual chromatid with high specificity.

It is currently impossible to carry out mechanistic biochemistry on tens or hundreds of protein complexes in a high-throughput manner, but these are the numbers that will defined by high-throughput high-content approaches and that will therefore need to be mechanistically analysed. New developments in miniaturised protein purification and parallelised biochemical assays using microarrays as well as microfluidics will therefore be an important future technology development focus (Section C.2.2.1). In addition, microfabrication and the utilisation of novel surface chemistry models will be important in the reconstitution of essential aspects of the physiological environment of the cell *in vitro*, thus permitting more realistic high-throughput assays.

Ultimately, molecular mechanisms elucidated by reconstitution *in vitro* need to be confirmed and validated in the much more complex physiological context of the living cell. Here again, new imaging approaches allow *in vivo* biophysical experiments. Fluorescence redistribution after photoperturbation (FRAP and photoactivation/uncaging) approaches permit determination of binding constants between members of protein complexes in their native environment. Complementing this, single molecule tracking and spectroscopy can be used to probe the mobility and size of protein complexes as well as the rheology (i.e. viscosity, flow, etc.) of the cellular environment they encounter. New super-resolution imaging approaches such as photoactivation localisation microscopy (PALM) and stochastic optical reconstruction microscopy (STORM) allow precise mapping of complexes with a resolution of tens of nanometers.

Such dynamic fluorescence measurements therefore afford single molecule precision and can in principle be applied in a generic manner to the whole proteome one protein at a time. However, conclusions about molecular mechanism based on dynamic analysis of fluorescently tagged proteins remain to some degree indirect. An important ongoing effort is therefore to develop a new generation of fluorescence-based sensors that, rather than simply making the protein of interest fluoresce, report on its specific enzymatic activity or on post-translational modifications that are diagnostic of its active state. So far, development of such probes for a protein of interest has been

Box C.2.4 Structure and dynamics of nuclear pore complexes during assembly and transport

The architectural organisation and function in transport between the nucleus and cytoplasm of nuclear pore complexes (NPC) has been a long-standing challenge for molecular biologists. NPCs are gateways to the nucleus but are not simply static channels in the nuclear envelope that passively control the flow of macromolecules. They are also directly linked to a plethora of other functions, including the regulation of gene expression and the organisation of genome architecture. Furthermore, NPCs are highly dynamic structures. During mitosis, they are broken down into sub-complexes that participate in a variety of regulatory mechanisms that control cell division, some of which are still very poorly understood. Fully assembled NPCs are also not just simple building blocks but highly dynamic entities, a subset of the components of which continually associate or dissociate. The high flux of cargo molecules through NPCs during nuclear transport (thousands each second) and the low degree of temporal and spatial coherence between different transport events pose substantial limits on the information on cargo transport obtainable by the use of any single technology or method. Furthermore, the large size of the NPC impedes X-ray-based structural studies that would provide insight into its architecture.

To help overcome these challenges, EMBL scientists will apply a broad spectrum of complementary techniques covering a wide resolution range to the study of NPC function. As the entire NPC cannot be reconstituted *in vitro,* we will use electron microscopy and especially electron tomography techniques to generate higher resolution three-dimensional structures of individual, active nuclear pores than are currently available. As the resolution of EM tomographic reconstructions is usually too low to reliably place crystal structures of single components into the EM densities, single particle EM reconstructions of sub-complexes at 10 to 20 Å resolution will be undertaken and should provide useful intermediate structural information that can be used to bridge the resolution gap. In parallel, a large experimental repertoire of high- and super-resolution fluorescence microscopy techniques will be employed to study the assembly of NPCs as well as the characteristics of individual transport pathways and events rather than just the average of many events. Fluorescent imaging of single proteins will be used not only to resolve temporal events during transport but also to investigate what happens during breakdown and reassembly of individual NPCs. Molecular and cell biology techniques will then be used to correlate these time-resolved data with individual pore complex electron tomograms at high resolution. To analyse the plasticity of dynamic NPC components, we will employ single molecule interaction techniques that, uniquely, can provide dynamic conformational information. Ultimately, working models of molecular motion in the nuclear transport pathway and NPC disassembly and reassembly will be generated via computer simulation and iterative rounds of mathematical model building and experimental data integration.

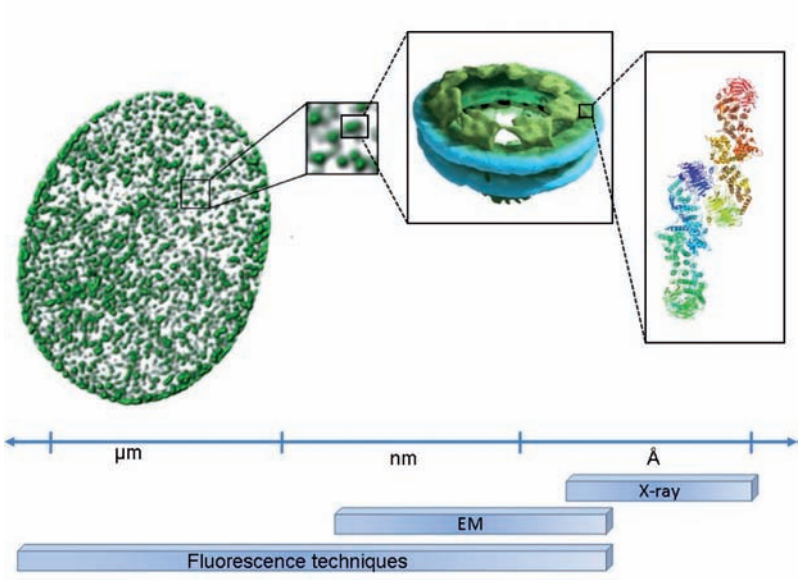


Figure C.2.4: Analysis of NPCs covering different spatial and temporal resolution ranges. Fluorescence microscopy is non-invasive and has an exceptional time resolution to study NPC dynamics unperturbed within the cell. Electron microscopy allows resolution of the molecular architecture of the active transport machinery. Further molecular details will be obtained from crystallographic and single molecule fluorescence approaches.

difficult and time-consuming and new technologies to design and synthesise activity probes in a systematic manner are therefore an important technology development goal for our chemical biology groups (Box C.2.3)

2.1.2.5 Cell Biology provides the physiological context for Structural and Computational Biology

Super-resolution imaging opens the exciting possibility of direct correlation with structural analyses by correlative light and electron microscopy (Box C.2.2). Applying these methods in a generic and robust manner to many cellular complexes will be an important challenge for technology development and will form a strong link between cell biology and structural biology (Section C.2.1). At the same time, it will ultimately be of crucial importance to integrate dynamic data obtained at different levels of cellular and subcellular resolution into computer models of cells carrying out the biological process of interest (Box C.2.1). Such models will not only serve to integrate and visualise the data but will also make them available for the development of predictive models of the function and self-organisation of cellular protein networks (Section C.2.2.3). EMBL is ideally placed to participate in these efforts owing to the strong links between cell biologists, structural biologists and computational biologists. This multi-scale effort of data collection and integration needs to be focused on concrete biological questions, as exemplified by our ongoing and future efforts to understand the assembly and function of the nuclear pore complex (Box C.2.4), the mechanism of endocytosis (Box C.2.2) and the structure of chromatin and chromosomes. Importantly, the methodological framework that must be established to solve these questions will subsequently be available to tackle new challenges. These will include going beyond the level of single cells to study the interactions between cells required for the organisation and function of organs and organisms. Wherever possible, generic aspects of these technologies will be made available to all EMBL groups and scientists from member states via EMBL's Core Facilities.

2.1.2.6 Bridging cellular and organismal resolution

Cells rarely act in isolation. Cues from and interactions with the immediate environment influence and guide the processes that happen inside a cell. Many of the properties that define cells and other living systems thus only become apparent when they are studied in their natural context. For many cells this context is provided by multi-cellularity. Therefore, in order to achieve a comprehensive understanding of the functional and organisational principles of biology, cells need to be studied in the context of the living organism. At a time when biologists aim for a complete understanding of living systems across all levels of biological organisation, understanding how the genotype of an individual is translated into its phenotype remains a major challenge, distinct parts of which are described throughout this document. In this section we will focus primarily on multi-cellular development.

2.1.2.7 Seeing is believing – from visualisation to prediction

As was discussed above for cell biology, developmental biology research has been revolutionised in recent years by the advent of powerful new imaging methodologies. It is fair to say that the ability to view morphogenetic processes in real time in living organisms has changed the manner in which developmental biologists think and formulate their questions. EMBL researchers have collaborated in the pioneering development and application of SPIM and DSLM imaging techniques for the study of embryonic development at cellular and sub-cellular resolution (Section C.1.2). In future, this collaboration will be further supplemented by the use of ever more specific and dynamic imaging reporters and sensors of cellular processes. A major effort to be made during the next Programme that will involve groups from several Heidelberg Units and Monterotondo will be to adapt both imaging and *in vivo* reporter molecules for easy use in mammals, and particularly in mice. Real-time imaging will become the main tool for analysis of the diverse metabolic and signalling events, as well as the cellular behaviours, that underlie morphogenesis and organogenesis in living organisms.

Many aspects of morphogenesis require the accurate development of physical mechanisms, such as force generation or modulation of tension, to proceed correctly. It is therefore important to study and measure these physical parameters when addressing morphogenesis. Precisely coordinated changes in cell behaviour determine the shape and size of organs during embryogenesis. Dynamic changes in tissue organisation underlie wound healing and regeneration and the loss of such organisation is a hallmark of cancer. Hence, studies addressing tissue dynamics in embryos or cultured multi-cellular systems have the potential to profoundly impact our understanding of medically important processes. The combination of new imaging, experimental and computational technologies allows

the analysis of dynamic cellular machines and of biophysical aspects of development, such as the role and distribution of forces in the context of intact developing embryos (Box C.2.9).

With the advent of high-throughput live imaging, developmental biology must, in common with other areas, develop the methods necessary to become more predictive. In future, computer-based simulations will play an increasingly important role in the study of patterning and development. However, the complexity of living systems is such that, even in the "predictive age", unbiased experimental tests involving genetic and other perturbations to cells and organisms will continue to reveal unsuspected molecular and cellular processes. These same experimental approaches will also play a crucial role in the validation of predictions derived from mathematical modelling, high-throughput functional genomic experiments and network analyses.

2.1.2.8 An integrated view of developmental decisions and dynamics

Developmental biologists at EMBL bring the analysis of discrete biochemical reactions underlying cellular function into the context of tissue and organ development, homeostasis and evolution. Taking integrated approaches, and making use of the technologies described in the preceding sections, they will seek answers to outstanding questions regarding the development and homeostasis of living organisms. A few examples follow:

Mammalian genomes have an intricate regulatory architecture that dictates how genomic information is translated into coherent gene-specific expression programmes and phenotypes. Geneticists at EMBL have developed sophisticated tools that allow the precise and controlled manipulation and restructuring of the mouse genome to probe the basis and logic of the regulatory architecture of developmental loci.

An important and poorly understood issue in embryology is the question of how and when cell asymmetries, which lead to the organisation of different primary tissue layers, are first established during the early stages of development. At what stage and by what mechanisms is symmetry broken and totipotency forgone? What is the contribution of molecular information pre-localised in the egg to the early steps of embryonic development in different species and how is its deployment controlled? Future research into these questions will involve an integrated approach combining the imaging of cell-specific read-outs with biochemical and proteomic information.

How tissues are patterned and cells organised into three-dimensional tissues remains one of the central questions in developmental biology. The coordination of cellular behaviour that results in development involves the synchronisation of many distinct patterning processes. Comparative studies of organisms that follow different strategies during development is one approach to this problem, but ultimately it will also be necessary to follow the way in which individual cells within a single organism respond in an integrated way to determine tissue size and shape. A favourable example that will be studied at EMBL is lateral organ formation in the plant model *Arabidopsis thaliana.* Integrated single-cell resolution approaches will be applied to understand organ positioning, differentiation and growth, and how these different processes are coordinated (Box C.2.5).

How spatial signals generated by diffusible molecules are interpreted and translated into dynamic collective cell behaviour will be an important focus of organismal research. Two other experimental systems have been established in zebrafish and will be further pursued. One studies the population of microglia, the phagocytes of the brain, which detect neuronal cell death in the brain and thus ensure that dying cells are disposed of so that brain morphogenesis can proceed in the correct way. The other is the lateral line organ, which is formed when clusters of cells migrate along the developing fish embryo towards the tail and simultaneously differentiate and position mechanosensory organs in the skin. Both are particularly amenable to the combination of the imaging expertise at EMBL and genetic analysis.

Although certain types of biological clocks, like the generation of circadian rhythms, have been heavily studied, how the rate and sequence of developmental events are coordinated and how the timing of individual patterning events are controlled are crucial questions that remain largely unanswered. One such developmental clock is manifest through the periodic waves of gene expression observed during patterning of the embryonic mesoderm in vertebrates. The issue of developmental timing will be a new focus of research at EMBL and, as described above for studies at the cellular level, will exemplify the principles of introducing the fourth dimension to our understanding of developmental processes (Box C.2.6).

Box C.2.5 Mapping the patterning of gene expression at single-cell resolution

A fundamental aspect of multi-cellular organisms is the patterning of gene expression. This phenomenon underlies the distinct ways in which cells or populations of cells behave during the complex process of development and dictates the behaviour of specialised cells and tissues in mature organisms. Independent of spatial patterning, gene expression at the single-cell level also exhibits periodic behaviour (e.g. cell cycle, circadian cycle) as well as stochastic fluctuations. All of these means of generating variation in gene expression play an important role in cell-fate decisions in a range of organisms. Significant progress has been made in documenting the spatial patterns of gene expression on a genome-wide scale. Spatially independent variability, however, remains poorly understood and a complete genome-wide expression map at single-cell resolution for an entire multi-cellular organism has not yet been achieved. The major goal of this multidisciplinary, multi-team project is to leverage the throughput of advanced sequencing technologies and droplet-based microfluidics with the three-dimensional spatial resolution of confocal imaging to gain a comprehensive spatial overview of transcriptional patterning at single-cell resolution, for all cells of a single specimen.

Each cell within a sample will be engineered to express a random combination of distinct fluorescent proteins (FPs). These cell-specific FP "codes" are recorded using both confocal microscopy and, after tissue dissociation, single-cell transcriptome sequencing (RNA-seq). By matching the single-cell FP expression profiles obtained using these two techniques, a preliminary map can be created that assigns cell positions to single-cell transcriptomes. To remove ambiguity caused by the fact that some cells will express the same combination of FPs we will use a probabilistic approach that incorporates prior knowledge of gene expression and transcriptome similarity information. In this way a comprehensive and detailed three-dimensional map of gene expression will be constructed that allows both an analysis of spatial patterns as well as spatially independent variation. The methodology will be initially applied to study organogenesis in the model plant species *Arabidopsis thaliana*. Plants are thought to have evolved multi-cellularity independently of animals and thereby provide the basis for a broad comparative study for understanding how evolution solves similar patterning problems. If successful, the interdisciplinary approach can in future be applied to other organisms to better understand the biological significance of spatial and temporal patterning of gene expression and to analyse which regulatory mechanisms have been conserved across evolution.

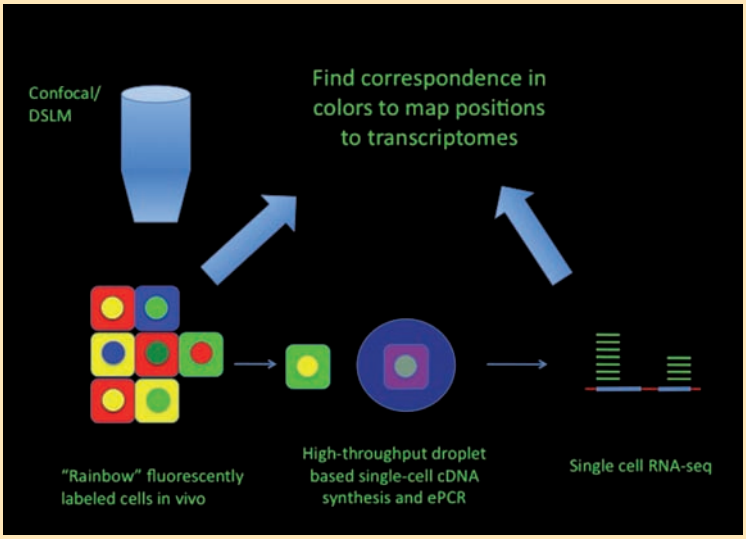


Figure C.2.5: Schematic illustrating the overall strategy for combining high-resolution spatial information with single-cell RNA-seq.

The standardisation of experimental conditions and focus on cells grown under defined culture conditions or animal models kept in a constant laboratory environment has until now largely masked any effect of the environment on differentiation and development. The environment however is known to have a major effect on both metabolism and development, although little is known about either the extent or the mechanisms by which metabolism regulates specific developmental events. Future research in this direction will involve the imaging of metabolites in living organisms that are kept in varying conditions, and the development of sensors of specific metabolic processes during patterning and morphogenesis. The development of such sensors in the cell, developmental and chemical biology laboratories and their successful transfer into animal models will be essential to achieve this goal.

Box C.2.6 The control of developmental timing by biological clocks

Just as it is vital to understand how biological processes operate in three-dimensional space, so it is crucial to learn how they are organised in time. Molecular interactions within cells have to be organised so that they occur in the right order and at the appropriate pace to ensure the correct functional output is achieved. For this reason biological systems have inherent timing devices, so-called biological clocks, which serve as meters and indicators of time. EMBL scientists are aiming to unravel the fundamental principles of such biological clocks: how is time measured during a temporally tightly regulated process like embryonic development? What extrinsic and intrinsic signals control timing? What information is encoded in the temporal profile of signalling activity?

The somite segmentation clock is an example of a biological clock that controls the formation of the pre-vertebrae in early mouse development. Through an as yet unknown mechanism it drives oscillatory activity of three major signalling pathways (Wnt, Fgf and Notch). Conceptually, an oscillatory signalling output can be generated by fluctuating protein levels of key signalling components, as achieved in transcriptional–translational negative-feedback loops. Alternatively, a post-translational mechanism involving regulated protein–protein interactions could likewise underlie dynamic signalling activity.

To distinguish between these two possibilities, we will develop a versatile imaging platform using knock-ins of fluorescent and bioluminescent markers into candidate gene loci in the mouse that will allow us to visualise not only transcriptional activity but also protein dynamics and protein–protein interactions *in vivo*. This approach will reveal at which level – from transcriptional to post-translational – an oscillatory signalling output is generated. With this knowledge we will perform a systematic and in-depth functional analysis of the underlying mechanisms using the powerful toolbox of mouse genetics combined with our ability to visualise the dynamics in real time.

Our approach is designed to provide mechanistic insight into how the period and amplitude of signalling oscillations are controlled by extrinsic and intrinsic signals in development. It can in future be adapted for the analysis of signalling dynamics in developmental processes other than mesoderm formation and for the application to single cells, as well as for the characterisation of mouse models of disease.

How signalling pathways elicit specific, context-dependent responses and activate specific sets of target genes is a challenging question in biology. The temporal profile of signalling pathway activity, such as its oscillation frequency or duration of activity represents an additional, as yet essentially unexplored, level of complexity that can influence the specificity of responses. Our future research will focus on this temporal dimension. Understanding the mechanisms underlying temporal dynamics is essential for deriving general concepts about the importance of time in developmental biology.

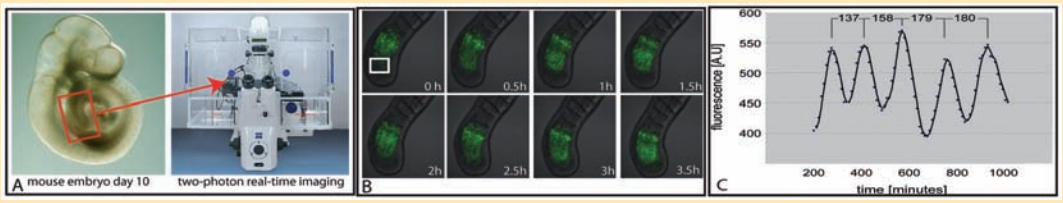


Figure C.2.6: A. The tail region of a mouse embryo is transferred and incubated on a microscope stage (in an environmental control chamber). B. Selected time-points from two-photon real-time imaging experiments are shown (note dynamic changes of fluorescence in tail part of embryo) C. The fluorescence from one region of the embryo in B is quantified (white box) and shows striking oscillations. Fluorescence corresponds to transcriptional activation of the gene encoding the signalling molecule lunatic fringe.

Biology is at a particularly interesting and challenging juncture, with the parameters underlying many biological phenomena being quantified and described with increasing accuracy under different conditions. An important aim of developmental biologists at EMBL is to develop the capacity to model well-defined morphogenetic processes, and it will be important to recruit new group leaders with both theoretical expertise and a combination of interests in modelling and experimental manipulation of developmental processes. Together with other theoretical/experimental modelling groups, these scientists will contribute to creating the network of computational biologists throughout EMBL that will be needed to succeed in the next Programme.

2.2 Unravelling biological complexity

Complexity in living systems stems from the fact that almost every biological function arises from the combined action of multiple molecular or cellular components. Although classical methods available to biologists are extremely powerful in elucidating the identity and function of single components of a system (i.e. genes, RNAs, proteins) they are not good at predicting the often unexpected properties that emerge when several components interact with each other to produce a functional unit. Moreover, living systems behave non-linearly with one input often producing multiple outputs. Examples of this are the coordinated transcription of multiple genes in response to one signal, the context-dependent production of alternative transcripts and therefore alternative proteins (often with different functions) from a single gene, or the many post-translational modifications of proteins that alter function in complex ways. Furthermore, many single molecular species perform different functions as a part of distinct complexes. In this case, it is the ensemble of proteins in the complex that acquire a specific function and this function is frequently hard to predict from knowledge of the individual proteins. At a higher scale, in multi-cellular organisms, interactions between cells have to be taken into account. The collective behaviour of cells greatly influences, and is greatly influenced by, the behaviour of each individual cell in the population through signalling. We need to address this complexity to understand biological systems. This means both embracing new technologies with the capacity to produce datasets that record the behaviour of multiple components of a system and making use of computational methods to analyse and predict the non-intuitive complex behaviours that emerge from the parts of the system.

Much of the current progress in life science research involves the use of unbiased high-throughput technologies. Previously, the only traditional approach that enabled global data generation in the life sciences, in favourable model systems, was genetics. The new "omics" methods allow individual researchers, and EMBL as a whole, to complement detailed mechanistic studies with broader information even when genetic approaches are not available or not adaptable to the problem being addressed. Other innovations provide the opportunity to produce quantitative dynamic imaging information in areas in which previously only qualitative data was obtainable and, for the first time, to gather molecular information in real time from *in vivo* systems. Finally, biocomputing methods and mathematical models associated with computer simulations in combination with experiments are beginning to allow prediction of the emerging properties of these complex systems.

EMBL is leading the way in embracing the challenge involved in understanding complexity. This is reflected in many of the research plans presented in this document, but two areas of the Laboratory have been particularly affected. In order to better meet the needs of functional genomics research we rearranged EMBL's Unit structure, with the creation of Genome Biology in place of the Gene Expression Unit. We have also incorporated new technologies, equipment and personnel into our Core Facilities to provide all EMBL research groups with access to cutting-edge high-throughput approaches for data acquisition and analysis. In addition, we have greatly expanded two areas of computational biology: data analysis and statistical data mining and mathematical modelling and simulation, areas which we see as mutually supportive and essential complementary aspects of our future requirements. The first type of expertise is needed to obtain the maximum benefit from the large datasets being produced by "omics" methods and to build probabilistic interaction and regulatory networks. The second is invaluable for using quantitative experimental data to simulate non-intuitive properties of biological systems, and to allow virtual experiments to be conducted on a scale that is not possible on the bench. Computational modelling generates experimentally testable hypotheses of how complex biological processes function and are regulated and is therefore a driver of progress. We will need more computational biologists in both areas in order to pursue the goals of the next EMBL Programme. However, many of the EMBL researchers who will need to use these methods will not have the required expertise. For this reason we will initiate three new EMBL Centres focused on key aspects of computational biology research and training (Section C.3). They will further strengthen our activities in these areas by providing researchers at all five EMBL sites with intellectual support and a platform for knowledge exchange, training activities and collaboration in computational methods.

2.2.1. Illuminating complexity through experiment

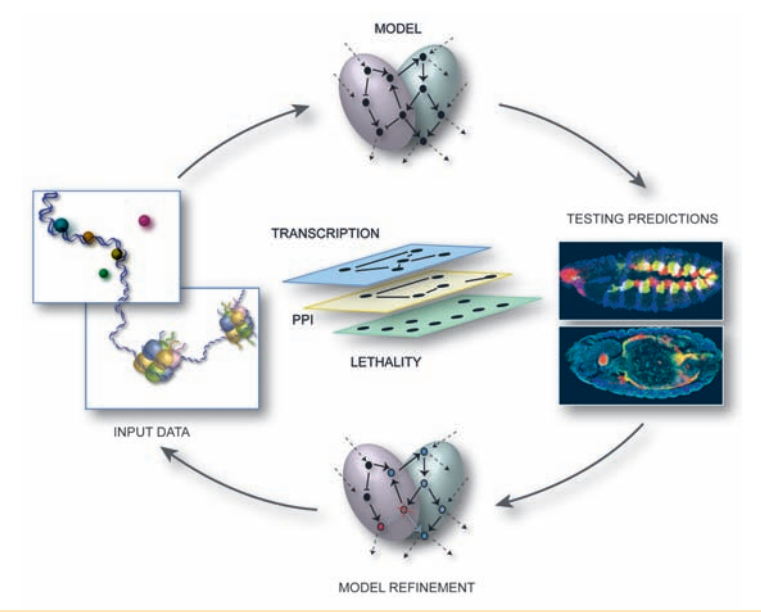
The key challenge for the life sciences is to understand how the information encoded in the genome is interpreted functionally. The genome is the genetic blueprint that ultimately gives rise to form and function in living systems and, through its variations, to long time-scale processes including evolution and short time-scale effects such as

Box C.2.7 Building predictive networks in single cells and multi-cellular systems

Biological processes are controlled by highly interconnected molecular networks. These have tens to thousands of components and exhibit extensive connectivity, cross-talk and redundancy. Substantial progress has been made in generating a molecular parts lists of different biological systems, for example through genome-wide measurements of mRNA, non-coding RNA, gene promoter occupancy by transcription factors, chromatin and epigenetic modifications and protein abundance. Integrating these heterogeneous data leads to the construction of networks that describe particular biological processes and their regulation. The future challenge lies in converting these static descriptive networks into predictive models of network behaviour in order to better understand the inherent complexity of biological processes.

EMBL researchers will tackle the specific problem of regulatory gene networks at two levels in the coming years. The first will use the single-cell model organism, yeast, to profile the effect on the transcriptome of hundreds of different genetic and environmental perturbations. The genetic component will be addressed by studying a subset of a well-characterised collection of single-gene-deletion mutants. Environmental perturbations will be induced through chemicals including drugs and metabolites. Analysing the transcriptional responses with RNA-seq will yield a quantitative readout of the expression levels of all genome regions and provide a genetic response map on a global scale. Bayesian approaches will then be applied to reverse engineer the network with the goal of predicting the expression response to combinatorial perturbations. We expect to gain major insights into the genetic and environmental regulatory mechanisms that control yeast cell behaviour.

Second, at the multi-cellular level, predictive models for cell-fate specification during development will be created. Mesoderm development in the fruit fly Drosophila melanogaster is a well-defined system for which EMBL scientists have generated a transcriptional network that defines the temporal and spatial occupancy of cis-regulatory modules (CRMs) by transcription factors during successive stages of development. These networks can successfully predict spatio-temporal CRM activity, providing the first global approach to transition from descriptive transcriptional binding maps to accurately predicting activity. The next key challenge – making predictive models of the relationship between CRM activity and gene expression – will be addressed through a probabilistic mathematical framework, taking advantage of the large body of in situ hybridisation data available in Drosophila and information on chromatin status. This approach should yield predictions at the level of when and where a gene is expressed and which transcription factors provide the most important regulatory input, which will be tested experimentally through genetic perturbations. For example, homologous recombination can be used to delete CRMs or alter the ability of transcription factors to bind to CRMs within the network. The tight iteration of experimental and mathematical approaches will be greatly facilitated by the complementary genomic, computational and genetic expertise of different EMBL groups.



Model analysis within both the yeast and Drosophila systems will elucidate the principle mechanisms and logic underlying gene regulatory networks, and provide a proof-of-principle that can be applied to model other systems including humans.

Figure C.2.7: Schematic of predictive network model construction.

propensity to health or disease. Functional genomics is moving beyond its initial phase, which encompassed defining a parts lists and a descriptive understanding of how pathways respond, to the understanding of why specific functional responses occur. More and more of the 'single-gene' assays that have been successfully applied to draw mechanistic insights in the past will become feasible on a global scale. Combined with genome-wide data at different levels, these datasets can then be analysed comparatively to gain direct insight into functional mechanisms. Over the past few years, genomics studies have already begun to yield unprecedented insight into the transcriptional response to various developmental, environmental and disease conditions. Unexpected discoveries have been made in our understanding of transcriptional control and how gene expression leads ultimately to a translated protein, areas in which EMBL scientists have made important contributions (Section C.1.2).

In this context, four technological areas strongly influence our planning for the next five-year period: next-generation sequencing, methods to identify molecular interactions, biochemical analysis at the single-cell level and high-throughput microscopy. The first three will be discussed here and the fourth has been topic of Section C.2.1.2.

2.2.1.1 Next-generation sequencing

The development of next-generation sequencing (NGS) is changing the way we tackle biological experimentation. NGS is an accurate and sensitive technique that provides unprecedented coverage of the response of biological systems. This allows us to gain quantitative insights into the molecular mechanisms underpinning a host of biological processes, data that are essential for predictive modelling of complex systems. It is therefore difficult to overstate the potential impact of this technology on biology in the near future. EMBL's interdisciplinary strength in molecular biology and leading expertise in genomics, computational biology and genetics puts us in an excellent position to leverage the power of NGS to understand how the genetic blueprint directs form and function in biological systems.

EMBL scientists will use NGS in many specific areas across almost all Units of the laboratory, as illustrated by the following examples. The comprehensive dissection of the basic mechanisms of gene expression is a central question that impinges on all aspects of biology. Ongoing efforts aim to quantitatively link each regulatory step of gene expression, from transcription factor occupancy, chromatin remodelling, RNA polymerase II recruitment and elongation, mRNA capping and processing, all the way to the recruitment of mRNA by the translational machinery. Each of these layers of regulation will be measured under defined conditions during specific phases of embryonic development in order to obtain a complete picture of the gene regulatory mechanisms and their deployment to direct the formation of an organism from a single cell. In particular, EMBL researchers in different Units will integrate NGS, quantitative single-cell live imaging, and microfluidics to address the robustness of cell-fate decisions during the specification of *Drosophila* mesoderm cells into different types of muscle. A second example will be to follow-up on the discovery of many classes of non-coding yeast RNA discovered during the last Programme by dissecting the functional requirement for these RNAs and the mechanisms by which they regulate biological functions. An important part of these goals will be to comprehensively identify the positions, dynamic changes and function of chromatin modifications during gene expression changes. Thus the projects will be achieved through the integration of large-scale transcriptome, epigenome and genetic perturbation analyses.

2.2.1.2 Identifying molecular interactions

Elucidating interactions between the molecular components of the cell remains a major experimental challenge. Estimates of protein number and complexity can be derived from genome analysis (like measurements of the number of spliced mRNAs) but cataloguing and understanding protein–protein interactions and the post-translational modifications of proteins is lagging behind. Even comparatively simple single-celled organisms such as yeast produce around 30 000 proteins (~3-5 proteins per gene). For humans this number scales up to an estimated 300 000 proteins (~10-15 proteins per gene). EMBL scientists have made many contributions to charting proteins and their interactions, both at the level of technology developments in mass spectrometry and in developing and implementing new methods to purify protein complexes (e.g. the widely used TAP-tag approach). A concerted effort in the coming years lies in obtaining a quantitative understanding of global protein abundance and dynamics in a number of model systems, with particular focus on the nuclear proteome. Extending similar measurement tools to selected metabolites such as lipids and carbohydrates is another exciting challenge that we will tackle, since all classes of biomolecules have crucial roles in connecting genotype with phenotype.

EMBL scientists will identify these molecular interactions by applying a number of different strategies. For example in yeast, studies will measure the transcriptional response to thousands of perturbations and then apply computational approaches to derive the regulatory network driving these responses. Similarly in *Drosophila,* this will be pursued in the context of multi-cellular embryonic development by integrating temporal changes in chromatin modifications, transcription factor occupancy, RNA abundance and RNA translation during consecutive stages of development into a computational framework to predict spatio-temporal gene expression and ultimately cell-fate decisions. Having dynamic measurements at all levels of the network will provide the relational data needed for probabilistic modelling and will feed into future quantitative modelling approaches (Section 2.3).

2.2.1.3 Single-cell analysis

Data generated by large-scale projects such as those described above must integrate with mechanistic studies to achieve the desired level of biological understanding. Increasing the throughput of detailed mechanistic analyses is an ambitious future goal. Following the dramatic progress made in DNA sequencing and protein analysis in the recent past we believe that there is enormous potential for developing methods that can measure biological components at miniaturised scales using microfluidic and nanotechnology devices. At EMBL we will be developing and applying these new technologies to drive progress towards high-throughput biochemistry and single-cell genomics. High-throughput biochemistry will render the biochemical analysis of proteins easier, cheaper and more generic, allowing many proteins to be studied in parallel. The emerging field of single-cell quantitative genomics has unparalleled potential to measure global changes in cellular components in individual cells, rather than obtaining average measurements from hundreds to thousands of cells. This will ultimately provide new insights into a range of diverse topics including the precise transcriptional response of a cell and how this changes in response to genetic variation or developmental signalling. One can already carry out some isolated experimental steps, such as separating and lysing cells, isoelectrically focusing organelles, or detecting single species of RNA or DNA, with lab-on-a-chip devices. The crucial next step will be to develop technologies that allow the integration and parallelisation of different steps of genomics experiments. To actively push technology developments in this area we recently recruited a microfluidics group that will work together with the biological and computational groups in the Genome Biology Unit.

Once systematic measurements of the components and their various interactions have been achieved in the ways described above, two additional challenges remain. First, how can we visualise the complex interactions within and between networks so as to make them meaningful and in such a way that we can extract fundamental principles of biological function (this will be dealt with in Section C.2.2.2)? Second, how can we use the information to generate mathematical models that can be tested experimentally (this will be the topic of Section C.2.2.3)?

2.2.2 Analysing and representing complexity

One of the many challenges arising from the new global methods is how to present the results in an accessible, user-friendly and understandable way to facilitate detailed analysis by experts and comprehension by non-experts. The sheer quantity and heterogeneity of the data produced makes this a difficult exercise. It must nevertheless be tackled by EMBL as we are engaged in many projects, both in-house and with external participation, in which adequate data representation will be required to allow the extraction of information and hypotheses. This requires not only novel data abstraction models but also a re-implementation of existing methods. For example, the classical visualisation of simple sequence alignments needs to be completely redesigned to make them useful for large-scale projects like the 1000 Genomes or the Cancer Genome Projects. Current phylogenetic tree displays also break down when more than 10 000 organisms in a sample need to be classified, as is routinely becoming the case in metagenomics projects (Box C.2.8). Understanding biomolecular networks in space and time, which is a major objective of many EMBL research groups, is hampered by the same problem: how to represent the data so that it can be navigated and understood. Static interaction networks representing only one type of biological node (e.g. proteins) in two dimensions are already difficult to interpret. The problem is further exacerbated when resolution in time or space is required. The analysis and interpretation of relationships between individual molecules, networks and higher order phenotypes is becoming a major bottleneck in systems biology. For these reasons, EMBL computational biologists will engage heavily in this area.

A prerequisite of any integration and representation of data are generalised ontologies and standards that are developed, discussed, agreed and adopted by the community of researchers in the field. This is an ongoing worldwide

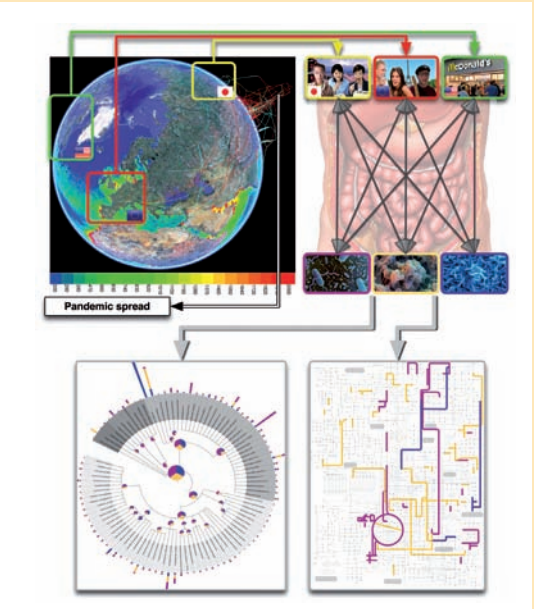
Box C.2.8 Metagenomics

Every biological system is determined by a combination of intrinsic properties and environmental factors. Over the past years great progress has been made in unravelling the genetic and molecular basis of human health and disease. However, a complete understanding of the human body as a biological system also needs to incorporate information on the environmental context. A hitherto hidden aspect of the human environment – the human microbiome, the collective of microorganisms living in and around us – is increasingly becoming available through large-scale metagenomics projects that sequence and generate quantitative data on all organisms present in the human body. However, metagenomics studies will encounter three major challenges in the future: first, the vast amount of sequence data produced, soon expected to exceed a Terabyte of data per study, have to be analysed to extract meaningful information in a timely manner; second, new visualisation tools have to be developed to represent the complex interactions and relationships present in environmental samples; and finally, the complex data have to be made available to the scientific community in a useful and manageable way.

Researchers at EMBL will use their expertise in computational biology to produce tools and pipelines that can manage and annotate the data and metadata (e.g. human properties such as disease state or weight) that come with the microbial sequence information. Using comparative metagenomics, we will further identify microbial functionalities that correlate with human properties, e.g. marker genes or pathways in faecal samples that correlate with diseases and might become useful diagnostic tools.

In the context of the international human microbiome consortium, EMBL researchers have begun to analyse microbial communities in the gut. To assess the impact of microbes on human health and diseases such as cancer or obesity, many different parameters, including diet, disease state, age, gender, seasonal temperature variation and pandemic spreads need to be considered. EMBL scientists will develop new ways to integrate such heterogeneous data, produced by different disciplines and methodologies, with sequence data and to efficiently visualise the complex network of human–microbial interactions. EMBL scientists are currently participating in an EU-funded project called MetaHIT, in which faecal samples from several hundred patients suffering from Crohn's disease or obesity are being subjected to metagenomic RNA-profiling techniques and metabolomic analysis. The goal is to identify abnormalities of the microbiome and ways to diagnose and perhaps even treat the diseases by directed modification of the microbial gut communities and their interactions with the host.

Besides its clear biomedical potential, environmental sequencing and metagenomics are also powerful tools to unravel the complexity of entire ecosystems and to decipher the evolutionary adaptations and relationships between organisms inhabiting an ecological niche. They thereby provide a broader, more global biological context in which human beings and other organisms must be analysed and understood.



To realise the full potential inherent to metagenomics and the ambitious projects discussed above requires a community effort involving teams of biologists, clinicians and bioinformaticians who analyse different aspects of the data to answer specific research questions. To facilitate such emerging global efforts an extension has been added to EMBL-EBI's International UniProt database (UniMES, short for UniProt Metagenomic and Environmental Sequences database) and a portal will be built to make it available to the scientific community.

Figure C.2.8: Schematic illustrating the principles of metagenomic studies. Information on (a) global environmental and geographical aspects of world-wide data collection need to be integrated with data on the metagenomes of these samples at the level of species and molecular functions. The example shows the identification of three 'gut–enterostates' shared by people in distinct geographic location (b) and the assignment of species involved within a tree of life (c) as well as metabolic pathways and genes (d) that seem to cause the differences observed between the three states.

task, the importance of which continues to grow, that has become an integral aspect of the mission of EMBL-EBI. Recent examples are the Systems Biology Markup Language (SBML) and the Systems Biology Graphical Notation (SBGN) that have been designed with strong input from EBI researchers to facilitate the exchange of biological models between different types of software or to allow a standardised representation of biological interactions. These are now being widely used by the systems biology community.

Similarly, in genomics we will develop novel conceptual models of how to represent and compress gene assemblies and splice variants on genomic maps, whilst in transcriptomics, novel graphics for representing gene regulatory networks and expression data are being explored. In cell biology, methods have to be developed to link the presence and activity of a molecule with its function and localisation in the cell. Changes in metabolic activities and signalling networks must also be represented, including the differences in these parameters between individual cells in a developing or mature organ or organism.

For data integration, the genome is a powerful reference onto which various types of data can be mapped. This has already been done, for example, in Ensembl, the collaborative EMBL-EBI/Wellcome Trust Sanger Institute data portal for genome data. Other data, for example quantitative high-resolution dynamic maps of the collective of molecular machines in an organelle or a cell, are not usefully represented in such a linear system. Instead, maps that can incorporate information on the interactions of the diverse components of large macromolecular networks – including proteins, DNA, RNA and metabolites such as carbohydrates and lipids – that will be produced by large-scale application of biochemistry methods, are needed. In addition information about the function of each component generated by large-scale phenotypic screens (e.g. the Mitocheck project, Section C.1.2) can be added as another layer of information to define functional domains within gene or protein networks. Understanding how cells work and how they are organised requires the integration of experimental structural and cell biological data with computational approaches that will enable us to store, visualise, navigate and mine the data in order to arrive at mechanistic and phenotypic interpretations and to derive underlying principles. The four-dimensional context of the living cell is the natural reference frame onto which quantitative structural and functional data can be mapped. Representing this data in a central resource and making it navigable as a "Google Cell" would be a major achievement. It is our intention to work towards this goal.

Although such a global cellular navigation system may seem very ambitious, EMBL is in an excellent position to tackle such a project because all the necessary expertise is present and can readily be integrated. In fact, a pilot project to characterise, in one "simple" unicellular organism, as many cellular complexes as possible in a three-dimensional context, and to begin an analysis of their interactions with metabolites, was begun over the period of the current Indicative Scheme. This project, on *Mycoplasma pneumoniae*, involved many EMBL groups, including most of the Structural and Computational Biology Unit, as well as external collaborators from the EMBL Partnership Unit located at the Centre for Genomic Regulation in Barcelona. The long-term goal of the project is to obtain as comprehensive as possible a three-dimensional picture of one of the simplest free-living organisms (Box C.2.1). The project will continue in the future, as will an even more ambitious effort directed at a similar analysis of the considerably more complex *Mycobacterium tuberculosis*, the single-celled eukaryote yeast, and an important functional ensemble of human cells, the network of components that executes cell division.

2.2.3 Physical modelling and simulation

Even when molecular interactions have been elucidated, we are at present still largely unable to predict how components of living systems function collectively. This holds true for gene expression and protein signalling networks, the generation of functional modules such as organelles or cytoskeletal structures inside a cell, the establishment of cellular functions such as polarity, cell division and motility, as well as for the collective behaviour of cells during embryonic development and adult life or the collective behaviour of microorganisms forming an ecosystem. As stated in the introduction to this section, statistical data analysis and mining (Section C.2.4) and mathematical modelling using the tools of statistical physics are complementary approaches that must both be applied to biology. The first helps us extract information from large and diverse sets of biological data whereas the second, described below, provides the mathematical tools needed to investigate dynamic interactions between the components of a system and to model the properties that emerge from these interactions.

Simulation-based models are based on the expression of fundamental physical and chemical principles in mathematical and computational languages. This involves integrating biological knowledge with the use of equilibrium

theories, conservation laws and ordinary or partial differential equations in which physical parameters that are employed in or generated by the simulations need to be compared against real values obtained by experiment. The increasingly fruitful interplay between modelling and experimentation in biology is supported by technology development. Stochastic simulations of biological processes, for example, often predict the emergence of dynamic behaviour that would not have been intuitively predicted. The ability to quantitatively analyse complex dynamic processes at the cellular and multi-cellular level (Section C.2.1.2) enables the predictions of such models to be checked against robust experimental data.

Pioneering steps in physical modelling of biological function were taken by EMBL's Cell Biology and Biophysics Unit, the members of which modelled key aspects of the assembly and function of the mitotic spindle, the structure that segregates chromosomes (Section C.1.2). Ongoing work on this topic involves the development of new computational tools to achieve more comprehensive, accurate and quantitative models. Importantly, the software platform created during this work allows a mixture of stochastic and deterministic modelling and therefore permits the average behaviour of populations of molecules as well as the action of individual molecules to be investigated. Reaction-diffusion processes of regulatory molecules can be included and are crucial for the emergence of a "system", to be represented in the model. These advances allow other biological functions and different cellular systems to be addressed. For example, EMBL groups are pursuing the molecular, kinetic and ultrastructural dissection of endocytic vesicle formation (Box C.2.2) to reach a point at which the data are precise and quantitative enough to allow the development of predictive models of how a network of regulators interacts with the actin cytoskeleton and the plasma membrane to allow internalisation of proteins from outside the cell. Other cellular functions that we plan to model on the basis of quantitative data that are now being collected include the assembly of the nuclear pore complex (Box C.2.4), the Golgi apparatus and the formation of cytoplasmic actin networks in animal oocytes.

But modelling is not limited to intracellular functions; it can also be used to study and generate hypotheses about multi-cellular processes. An example is the modelling of how adult body parts such as the wing are formed in the fruit fly from their precursor tissues, the imaginal disks, during embryonic development (Box C.2.9). Modelling this complex function with realistic physical forces of cell–cell interactions that give the adult tissue its shape is now becoming possible, driven by advances in imaging technology to quantitatively record development in real time at sub-cellular resolution. Another example is the formation of the lateral line organ in the zebrafish embryo. Here the modelling focuses on predicting how the dynamic interplay of collective cell movement and differentiation from mesenchymal to epithelial cells can give rise to the regular deposition of connected pressure-sensing organs along the lateral line of the fish. New technologies in imaging and laser nanosurgery developed at EMBL, combined with zebrafish genetics, are allowing EMBL groups to quantitatively analyse and test model predictions even in a biological system of this complexity. As the above examples show, the general aim of the current and future physical modelling projects at EMBL is to quantitatively describe the generation of biological form and function at the sub-cellular, cellular and supra-cellular level and to derive predictions about the emerging properties of complex biological systems that can be tested experimentally. This approach will ultimately lead to a true understanding of the system, in the sense of being able to predict how it will behave in response to perturbation.

These ambitious and interdisciplinary goals can be achieved because of the unique collaborative and interdisciplinary research environment of EMBL. In recent years, we have made use of the turnover generated by EMBL's fixed-term contract system to appoint physicists, engineers and mathematicians into positions that are embedded in the biology Units of the Laboratory. Driven by the success of the mitotic spindle modelling efforts, which provide a powerful example of the potential of this approach, EMBL currently has a significant and growing number of collaborations between biologists, physicists and mathematicians. EMBL is committed to continue to develop this "physical biology" spirit across the Laboratory. We feel this is essential because, as stated repeatedly in this document, we cannot understand complex biological systems by using bioinformatics, genetics and structural, cell and developmental biology alone. These disciplines are needed to generate, integrate and represent quantitative data about the components of a system. However, we crucially need the approaches of mathematics, physics and quantitative modelling to study how relatively simple molecular properties and interactions lead to complex collective behaviour, i.e. to understand the function and evolution of living systems. The predictive power that can be derived from quantitative modelling will change the status of biology as a science and will conversely lead to new developments in physics, mathematics, computer science and other disciplines. To foster this new way of doing biology, EMBL has created and will continue to strengthen training, mentoring and collaboration systems to ensure that the

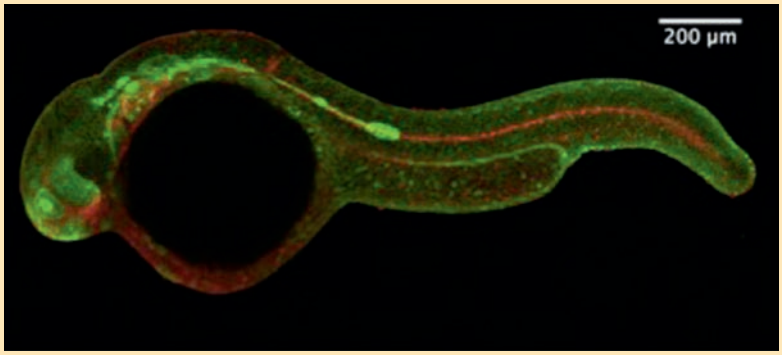
Box C.2.9 Dissecting the forces behind organogenesis

The development of specialised organs, such as the heart, brain or liver, is a crucial aspect of embryonic development. A fine balance of tissue growth and cell patterning and migration makes sure that developing organs achieve the correct size, proportions and identity. These processes are all connected and must be tightly regulated and coordinated in time and space.

Using a combination of state-of-the-art imaging techniques and genetic perturbation together with theoretical modelling approaches EMBL scientists will seek deeper insight into the mechanisms that underpin organogenesis during embryonic development. Novel light microscopy techniques will be used to image developing organs, including the zebrafish lateral line organ and the developing wings of *Drosophila* at sub-cellular resolution. Advanced image analysis techniques will identify the positions of both nuclei and cell membranes in three dimensions. Changes in nuclear position and cell shape will be monitored by novel tracking algorithms and generate a detailed high-resolution map of the collective dynamics of organogenesis from which valuable information on, for example, the reproducibility of organogenesis between embryos or the effect of either mutation or external conditions can be derived.

The main focus of the project will be on the biophysical mechanisms involved in organ development. Addressing the relationship between the mechanical stability of tissues and cellular growth control will require the integration of biochemical and physical determinants on multiple scales: the physical properties of the cytoskeleton and cell membrane that determine the structure and migratory behaviour of individual cells, the adhesion machinery that governs cell-cell interactions in the tissue, and the physical properties of multi-cellular aggregates and whole tissues.

In the future, we will further develop our mechanics models to address two fundamental questions. First, how can we bridge molecular and multi-cellular scales to build a model of the molecular basis of tissue mechanics? Second, what are the collective properties of tissues and how do they influence the fate of individual cells in the tissue? These mathematical models will provide a theoretical framework in which the information collected across multiple temporal and spatial scales, from observations on single cells over cell-cell interactions to population effects, can be integrated and interpreted. On the basis of such models, computer simulations will generate predictions about tissue growth and migration, which will be used to guide further experimental work. The model – and with it our quantitative understanding of organogenesis – will be refined in iterative cycles alternating between experimental and computational phases.



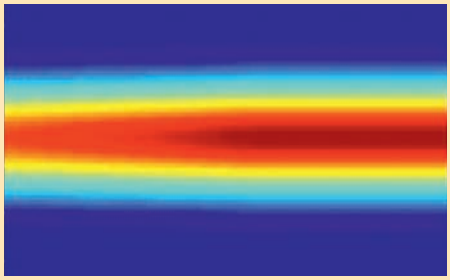


Figure C.2.9: Above: Tissue migration in the zebrafish embryo leads to the formation of the lateral line organ.

Below: Reaction diffusion equations allow dynamic modelling of guidance cue distribution that steers tissue migration during lateral line development.

required high level of interdisciplinarity is achieved. In addition, new lecture series and practical courses will be made available so that EMBL researchers from disciplines outside the life sciences can keep up with scientific progress in their fields. Creating Centres for Computational Biology (Section C.3) and the EMBL EIPOD Programme (Section E.3) are also important steps in this direction.

2.3. Exploring biological variation

Biological variation has multiple facets that are all of enormous interest to research. Inter-species variation reflects evolutionary processes and its study provides a window on the past and insight into the mechanisms of evolutionary change. At the same time, identification of what has been conserved over evolutionary time provides information on what is important for biological function. Divergence between the phenotypes of species reflects the genetic differences that separate them, which have been moulded by natural selection.

Just as species evolution has a genetic and an environmental component – random mutation and natural selection, respectively – variation between individuals of the same species has two distinct sources, genetic differences and environmental differences – or nature versus nurture in the words of behavioural psychologists. In the vast majority of studied cases, both sources of variability contribute to phenotypic differences between individuals but in different proportions depending on the trait under study.

In the broadest sense disease is also a type of biological variation, a variation from the normal, healthy state. The extent to which pathologies are determined by genetic and environmental factors varies from disease to disease. Hereditary diseases range from being completely genetically determined to having a fractional genetic predisposition whereas infectious diseases, caused by external agents such as viruses and bacteria, affect individuals differently in part because of inter-individual genetic variability. Most conditions, including cancer, obesity and neuropsychological disorders, are determined by a combination of genetic and environmental factors, the relative impact of which is often difficult to discern. EMBL's expertise is in the analysis of genetic and epigenetic mechanisms and we will focus on how these might differ between individuals and species and begin to explore, when a specific research question requires this, how the environment influences these mechanisms.

The enormous advance in DNA-sequencing technology over the past few years has transformed the extent to which the genetic variation between species, individuals and healthy and diseased states can be analysed. Complete genome sequences and an analysis of global transcription were previously limited to a few species, and generally to one or a very small number of individuals within a species, but the same data are now being obtained from many species, many individuals of at least some species and even from various individual cells or tissues from a single individual.

This means that the genetic component of variation between individuals or between healthy and pathological cells within a single individual can be studied in unprecedented depth. At the same time, the genetic differences between species at different levels of relatedness can now be explored in a much more systematic way. With the new data, our understanding of biological diversity will reach a new level and many EMBL groups will therefore be involved in studies of variation in the coming period.

2.3.1 Evolution: inter-species variation

The new capacity to perform detailed comparative studies of genetic variation between species has initiated a new era of evolutionary research at all levels of biological complexity. At the gene level, variation between genomes, transcriptomes and proteomes of different species can be used to extract evolutionary information by examining what has changed and what is conserved. Aside from experimental analysis, the conservation of a given sequence during evolution is the only reliable indicator of functional importance. Various facets of comparative sequence analysis will remain a major focus of EMBL groups at the EBI and in the Structural and Computational Biology Unit, who will focus on both method development and their applications in this area. Frontier areas of current and future interest include the detection of short linear motifs in proteins, which represent sites of interaction or modification, of non-coding RNA and their targets, and of DNA and RNA regulatory elements with diverse functional roles by phylogenetic footprinting.

Building on straightforward sequence comparison is the comparative analysis of gene regulatory networks between groups of closely related species, such as different Drosophilid flies, or between remote species such as insects, annelid worms or vertebrates. Determining the inter-species variation in network features will reveal the core architecture of gene regulatory networks and aspects of their flexibility and constraints. EMBL scientists are actively pursuing comparative approaches to gene regulatory network analysis and future research will focus on a deeper understanding of networks that are conserved between distant organisms, such as those controlling heart development in the fly, fish and annelid worm.

The new sequencing technology changes the way variation can be analysed and understood, all the way from the cell to the ecosystem. Transcriptome sequencing and expression profiling of different cells from the same individual with single-cell resolution, or of similar cells taken from different species, will be used to infer the evolutionary path that gave rise to distinct cell types. How did animal cell types as diverse as gut, muscle and nerve, or plant stem and flower cells, evolve from ancient precursor cell types? What triggers the diversification and differentiation of the cells in developing embryos? Comparative analysis should help identify the factors that triggered the evolutionary diversification of cell types and these are likely to retain important roles in cellular differentiation and in its pathological aberration, cancer.

In whole organisms, the availability of an increasing number of fully sequenced animal genomes and their gene inventories, in conjunction with refined reverse genetic knockdown and knockout techniques, sets the stage for the establishment of new model species. This will overcome the current "diversity gap" that complicates the analysis of molecular diversity and evolution, particularly for marine species. Exploring marine species will reveal the origins of key animal and plant innovations because the evolution of both major kingdoms began in the sea. At EMBL we will focus on studying specific marine animal species for this purpose. Moreover, the establishment and study of new model organisms – an area of research in which we have expended significant effort in the current Programme – will help to better understand which aspects of well-established model systems (*Escherichia coli*, yeast, *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Drosophila*, mouse) can be generalised and which are species-specific. Ecosystems represent the highest level of biological complexity. Their investigation has traditionally been descriptive in nature but the advent of metagenomics has fundamentally changed this picture. It is now possible to analyse, for example, the absence or presence of selected gene families in all the species that make up an ecosystem. This will enable us to learn about the physiological challenges presented by ecosystems and the solutions that evolve depending on the species present. EMBL bioinformatics experts are well equipped, in terms of access to computing power and environmental samples, to remain leaders in this exciting new field of biological research (Box C.2.8).

2.3.2 Genetic variation: intra-species variation

DNA sequencing and high-density genotyping of single nucleotide polymorphisms (SNPs) and copy number variant (CNV) regions has provided an unprecedented recent advance in the capture of data on genetic variation between individual human beings. The ongoing drop in the cost of acquiring this information has already led to a systematic advance in the discovery of both natural and somatic variation and its statistical association to disease, and a great deal more such information will be gathered in the coming years. However, this increase in discovery has not yet been matched by an increase in the ability to understand genetic variation. EMBL's next Scientific Programme will include several approaches that should greatly improve our ability to probe and understand the functional consequences of variation.

In the arena of human variation, the HapMap project and the subsequent 1,000 Genomes project have provided progressively more complete catalogues of human variation. The 1000 Genomes project aims to map essentially all genetic variation that occurs at approximately 1% or greater frequency in three major geographic populations – Europeans, Africans and Asians. Initial results have revealed a wide spectrum of variation in the human genome and a higher variation frequency than expected. Other projects plan to sequence more than 20 000 cancer genomes over the coming decade and to target whole genome and whole exome (the protein-coding part of the transcriptome) sequencing for a large number of human diseases.

Understanding the functional relevance of variation is generally difficult, and is pursued via a variety of approaches. Genome-wide association studies (GWAS) have statistically linked 114 human phenotypes to over 150 different genome regions and associate the presence of particular variants with increases or decreases in an individual's risk of disease or of displaying any other specific phenotype. However, GWAS studies have severe limitations: they often

highlight genome regions without any known biological function or connection to a disease, they explain only a fraction of the risk estimated from family studies and they do not allow the identification of genetic interactions or gene–environment interactions. Finally, the components of a pathway that contribute to the risk of developing a disease are not necessarily the same as the components that are amenable to intervention. This was illustrated by model studies carried out at EMBL of the contribution of individual allelic variants to complex genetic traits in yeast that affected its ability to act as a human pathogen. The identification of genes that contribute to the phenotype frequently provided no assistance in developing a mechanistic explanation of how the phenotype arose. All these factors point to the need to increase our mechanistic understanding of genetic variation to match the impressive gains in our power to discover variants.

In the next EMBL Scientific Programme we wish to contribute to a step change in our ability to understand the functional consequences of genetic variation. This will involve synergy between experimental and computational approaches and will make use of both experimentally induced and naturally occurring genetic variation. Our aim is to provide a general methodological framework with specific, cost-effective technologies to analyse the contribution of genetic variants to any phenotype, including disease. We will use methods that will allow individual allelic differences between genetically variable samples to be studied in isolation. High-density functional genomics assays will be applied to this panel and the results of the assays then integrated in a consistent computational/analytical framework. This approach follows the standard scheme of perturb-measure-analyse that is used in modern molecular systems biology, but at each stage there are unique components to the approach to be taken at EMBL.

Using wild variation in experimental studies requires a source of isolated individuals, e.g. isogenic model organism lines. We are already analysing the phenotypic consequences of inter-*Drosophila* sequence variation, but other resources such as cell lines derived from individuals in the population are available. In humans this has usually relied on the isolation of lymphoblastoid (*in vitro*-transformed white blood cell) lines from individuals. The 1000 Genomes project aims to collect lymphoblastoid lines from all sequenced individuals, and to make them available to the research community. We have already performed a variety of experiments (e.g. global chromatin modification analysis) on the first 60 of the expected 500 such cell lines to be produced in Europe. During the course of the Scientific Programme we expect the cell line of choice to change to induced pluripotent stem cells (iPS cells), which in principle allow a wider variety of functional studies. Studying genetic variation in iPS cells will also allow EMBL to contribute its expertise to the development of methods and technologies that will help to explore the feasibility of stem-cell therapy.

To experimentally manipulate and reveal the contribution of genetic variation to intra-species phenotypic diversity and human disease, we have established an allele-specific RNAi technology that functions both in cell lines and, in certain species, in whole animals. This allows the knock-down of one of the two alleles present in a diploid cell coupled to a functional test of the phenotypic consequence. By comparing reciprocal knock-downs of two alleles we can screen many loci in an allele-specific manner whilst holding the genetic background of the test constant. In contrast to the traditional approach of statistical association testing in outbred populations, in which allelic effects can be masked by genetic background variation, allele-specific RNAi is a direct test of allelic contribution to phenotype. Molecular assays to probe these panels of variation can include any assay in use across EMBL. We will, in particular, concentrate on analyses that can easily be performed in medium throughput (100-500 samples) and that can provide high-density readout of many aspects of a biological phenotype. High-throughput cellular imaging, as recently developed at EMBL to study cell division and protein secretion, is one class of assay in which we have particular expertise and which is broadly applicable. By examining live cells with automated high-resolution microscopy many aspects of cell behaviour can be extracted, including attributes associated with the cell cycle, cell morphogenesis and motility. Previous work with these systems has provided robust assay procedures and downstream analysis techniques able to process terabytes of time-resolved images and convert these into informative quantitative measurements amenable to mathematical systems modelling. The second class of assay considers chromatin plasticity and RNA expression coupled with deep sequencing. EMBL has significant experience and resources in these areas in both humans and *Drosophila*, and we will focus on providing higher throughput methods for these techniques in future.

The mouse is widely considered the model organism of choice for studying human disease, first because mice share the majority of their genetic information and physiological systems with humans and second because of the extreme flexibility of genetic methodology in mice. The distinguished history of mouse experimentation has recently been supplemented with powerful new tools to introduce genetic variation into the mouse genome. Large-scale

mutant libraries generated by the international EUCOMM, KOMP and NORCOMM project consortia, to which the EMBL Monterotondo outstation will continue to contribute significantly, can supply embryonic stem (ES) cells with conditional knockouts in many mouse genes. The ease and power of genetic manipulation of ES cells derived from these mice allows the creation of better models of genetic diversity in an otherwise homogeneous genetic background. For example, approaches developed within EMBL will enable the creation of "humanised" mice, harbouring human-specific gene variants or even human-specific haplotype blocks. These can be used for whole-animal studies, and also as a source of cells for cell-based approaches. To study the role of structural variants in mouse models, EMBL research groups have developed an *in vivo* transposon-mediated tagging approach. The collection of tagged mice will constitute a unique resource of distributed sites that can be combined to create genomic deletions, duplications or inversions, thereby reproducing human aneuploidies and structural variants for functional analysis.

All these approaches culminate in the analysis and modelling of the information obtained. The first step is the storage, manipulation and extraction of information from the data. EMBL groups have established approaches to reliably infer genetic variations, including SNPs and structural variants, from large-scale genetic data. Furthermore, assays with high-density phenotypic readouts require a complex data processing pipeline in order to generate high quality information. Production of cell images can involve up to 1000 variables that describe shape, morphology, localisations, and dynamic rates and ChIP-seq analysis can give rise to roughly 10^8 variables that summarise nucleotide enrichment at each position in the sequences analysed. Our experience in these areas will enable the phenotypic variables derived from such analyses to be associated with genetic variants, in effect performing a quantitative trait locus (QTL) analysis of the molecular phenotype.

More interestingly, we can take advantage of the co-measurement of different molecular phenotypes, of already existing mechanistic understanding of their interdependence, and of the causality between genetic and phenotypic variables to derive a new, predictive, model-based understanding of the system. For instance, from the joint distribution of two phenotypic variables such as a chromatin modification state and expression of a transcript and a genetic variable such as a SNP or a CNV, we can distinguish whether the chromatin modification causes the expression change or whether the expression change affects the chromatin.

2.3.3 Disease models and mechanisms

2.3.3.1 Deciphering mammalian genomic function

The international mouse genetics community, in which EMBL has a major role, has set long-term goals for achieving a comprehensive genetics-based view of mammalian physiology and disease. For the foreseeable future, the mouse will remain the premier mammalian model organism for defining regulatory networks, applying large-scale proteomics and mutagenesis strategies and, most importantly, for phenotyping these mutants to generate increasingly more accurate models of human pathologies.

To meet the increasing worldwide demand for conditional mutant models, internationally coordinated initiatives that include EMBL Monterotondo have been established for the systematic generation of mouse mutants on a large scale using various reverse genetics strategies. The majority of these initiatives are committed to the production of mutant mouse ES cell lines, each of which carries an altered or "floxed" allele of a different gene. ES cell mutations can be readily transformed into mice using blastocyst injection, and the mutation activated by crossing with the desired Cre recombinase driver strain. Recent studies have shown that accurately copying human disease-causing variants into mice can lead to the production of more sophisticated and accurate disease models. The increase in human genome sequence that is underway (see above) will therefore lead to many opportunities to produce better and better mouse models of diseases for detailed mechanistic study.

Many phenotypic characteristics vary throughout the lifespan of an individual, and late-onset alterations in gene expression that cause age-related health problems, and the majority of common human diseases, are often not adequately recapitulated through standard methods of gene inactivation or over-expression strategies in the mouse or other models. EMBL is ready to meet the challenge of bringing mouse models closer to human diseases using the extensive infrastructure and expertise within the Mouse Biology Unit, together with access to the neighbouring European Mutant Mouse Archive and prospects for an adjoining national mouse phenotyping clinic. Propelled by a dynamic and interactive team of mouse researchers, the Unit's activities integrate molecular, cellular and organismal information in a systematic approach to uncover genotype–phenotype relationships in physiological and

pathological settings. On the technical side, innovative chromosomal engineering and conditional approaches are being developed to achieve spatial and temporal control of gene expression, to augment the power of large-scale mutagenesis screens and to accurately model the human genomic variation that underlies differential susceptibility to disease. This goal requires the implementation of a number of approaches that can be grouped into three general categories:

Mouse mutagenesis

Together with new initiatives in human genomic analysis with heavy EMBL involvement (e.g. ENCODE), the recent launch of several internationally sponsored projects to analyse mammalian genetics through mouse mutagenesis, molecular analysis and phenotyping will broaden the utility of the mouse for translating the language of the human genome. An ongoing research emphasis on heterozygous, point mutant, and humanised mice, and the collaborative EMBL effort to build a novel mouse gene dosage screen based on specific human gene CNVs described above will greatly augment our ability to generate accurate disease models. Perturbations of metabolic, cardiovascular, immune and neurobehavioural functions that are identified through this screen will be pursued both by mouse biology groups at EMBL and by a network of formal collaborations in the wider mouse research community.

EMBL Monterotondo is spearheading CREATE, an EU-funded project to integrate large-scale imaging initiatives for defining expression patterns and to guide the genome annotation necessary for identification of cell type-specific regulatory landscapes that could be exploited in the construction of new Cre drivers. Available Cre driver strains cover only a small number of the 350 cell types in the mammalian body. Future studies of particular functions in which numerous cells/tissues communicate, such as neurobehaviour, metabolic controls and cancers, will depend on the production of a larger palette of cell-specific Cre driver strains. This project is extremely important, because increased cell-type specificity for Cre-mediated recombination is a crucial unmet need in international mouse mutagenesis initiatives and because it forms a link between EMBL's existing strengths in molecular cell biology and its longer-term goals in molecular medicine and the impact of genetic variation on mammalian biology and human disease.

Mouse phenotyping

Characterisation of mouse mutants in a medical context currently exploits standardised, comprehensive and systematic phenotyping approaches that have been pioneered by European consortia, in which EMBL Monterotondo has played a crucial role. Specialised phenotyping protocols developed in the Mouse Biology Unit for genetic and environmentally sensitised phenotypic screening have identified phenotypes not revealed under baseline conditions, and will be an important future source of disease-relevant annotations of the mouse genome. In the area of psychiatric genetics, EMBL researchers have pioneered the use of pharmacogenetic neural inhibition tools for dissecting neural circuits in behaving mice. These methods reflect the current revolution in genetically encoded neural manipulation tools, which will be combined with *in vivo* electrophysiology to identify and perturb neural circuits controlling anxiety, fear and aggression. A new focus on mechanical, chemical and thermal sensation is determining how activation of specific excitatory ion channels in the mammalian somatosensory system contribute to pain sensation under normal and disease states. A collaborative project established across EMBL will build a bank of sensors and switches to measure and analyse ionic flux in the mouse. This will generate tools for real-time imaging of perturbations in channel activity and their molecular regulators that can be correlated with electrophysiological and behavioural readouts. The relevant expertise and infrastructure for both phenotyping parameters has been recently expanded at EMBL Monterotondo and Heidelberg, providing an optimal environment to undertake these complex comparative analyses.

Mouse informatics

A major challenge facing bioinformatics today is the systematic capture of genotypic and phenotypic descriptions for complex queries and comparative analysis. To date, thousands of human diseases associated with phenotypic and genetic variations have been described, and their underlying genetic causes of disease are being continuously uncovered and recapitulated in mouse mutants. An increasing number of inherited human syndromes have been associated with combinations of different genetic variants in an individual. Encoding this information in standardised descriptive language, or ontologies, allows mining of phenotypic datasets using other integrated data types such as sequence, genomic location and/or biological function of gene products, assists in building appropriate experimental disease models, and promotes discovery of candidate disease genes and molecular signalling pathways.

These developments require collaborative initiatives that are clearly beyond the scope of an individual laboratory or centre, but with its breadth and strength in bioinformatics and computation, EMBL is ideally suited to tackle the formulation of innovative strategies for rationalising the pending explosion of information in mammalian biology. The interdisciplinary approaches fostered at EMBL can contribute to this international effort in a unique way, and will continue to be a central theme of collaborative research amongst computational and mouse biology groups. A core role for EMBL-EBI in hosting a new centralised database for mouse phenotypic information, recently established in collaboration with the International Mouse Phenotyping Consortium, will be crucial for generating and testing hypotheses of human disease aetiology, and will position EMBL to contribute to the future development of these aspects of molecular medicine research in a manner that will serve the whole community.

2.3.3.2 Molecular medicine at EMBL

Medicine has undergone a paradigm shift during the past two decades through its ability and desire to seek a molecular understanding of human pathologies and by rapidly moving toward diagnostic methods and causal therapies that embrace molecular biology. Research at EMBL is firmly rooted in the quest to understand biology mechanistically and multiple current developments make human biology more tractable in molecular terms than ever before. As is evident from the preceding section, disease is partly the result of the phenotypic consequences of genetic variation, a topic that is at the heart of EMBL's research. Thus, the medical research community and science at EMBL are on convergent paths and we intend to further increase our contribution to understanding human biology. The rapid changes in clinical and human analysis that result from some of the technology developments already discussed underline the requirement for experienced input from molecular biology to medicine and we hope this will make EMBL an attractive partner for the medical research community. It is crucial that we take steps to contribute to and benefit from this developing dialogue.

With its focus on fundamental aspects of biology, EMBL research can most strongly impact on medicine through the generation and molecular analysis of specific disease models. Extensive expertise and infrastructure has been established over the past decade to build and study increasingly refined animal and cellular models of pathologies in a spectrum of body systems including brain, blood, muscle, heart, respiratory and reproductive systems. EMBL has developed new research activities that feature innovations in high-throughput technology for genomic, proteomic and imaging analysis of mutant models. In recent years mouse models have allowed insight into a wide range of disorders and conditions including stroke, leukaemia and other blood diseases, anxiety-related disorders, muscle degeneration, sudden infant death syndrome, multiple sclerosis and defects in iron metabolism. EMBL scientists regularly engage in interdisciplinary collaborations with medical scientists to study psychiatric disorders, cardiovascular disease, cancer, disturbances of cholesterol metabolism, haematological diseases and cystic fibrosis. EMBL's repertoire has also recently been expanded to study pain disorders as well as explore non-coding genetic elements in physiology and disease.

Many of our future plans build on groundwork carried out during the current Programme to investigate molecular mechanisms of disease. We have implemented functional genomics as a technqiue to study the genetic basis of disease and these methods have, for example, been applied to shed light on mitochondrial genes involved in disease, to identify genes involved in cholesterol regulation and to construct a high-resolution map of large structural CNVs on human chromosome 21 that are associated with Down's syndrome. Live imaging also bears great potential for molecular medicine research. EMBL scientists are planning to develop fluorescent probes that can monitor different disease states in the living organism and that can be non-invasively applied with the prospect of being used as diagnostic tools in clinical medicine. One success in the current Programme has been the development of a fluorescent marker that monitors enzymatic activity in macrophages and functions as an indicator of chronic inflammatory lung diseases. The project will be continued in the future with a focus on uncovering the multiple functions of macrophages and their roles in disease (Box C.2.10). This research will be carried out with the Medical Faculty of Heidelberg University in the context of the Molecular Medicine Partnership Unit (MMPU, Section G.2.2.3). In the MMPU, integrated teams of clinician researchers and basic scientists collaborate to develop and analyse disease models in a way that couples the clinicians knowledge of disease with EMBL's know-how, scientific infrastructure and Core Facilities.

The integration of clinician scientists into joint interdisciplinary projects will in future be extended and intensified through our scientific interactions with other institutional partners including those of the Nordic EMBL Partnership for Molecular Medicine (G.2.2.7). These collaborators share common goals and scientific interests that build

on collaborative projects, shared conferences and EMBL's participation in staff recruitment. Extension of the successful internal EIPOD scheme for interdisciplinary postdocs will enable us to attract clinician scientists as well as physicists, mathematicians, engineers and computer scientists to EMBL. In addition, EMBL-EBI directly supports medical and clinical communities through efforts to discover human genetic variation. The EBI is the European hub for the collection, maintenance and curation of such data and its scientists annotate human variation and mutation data from external resources that either have performed genome-wide studies or have examined specific regions of the genome including genes used for diagnostic tests. In this effort, active collaborations between the EBI and diagnostic laboratories and clinical researchers are developing. One example is a joint project in targeted

Box C.2.10 Functional imaging in mouse models and in the clinic

EMBL has a long tradition in developing cutting-edge tools for imaging dynamic cell behaviour *in vivo* with high temporal and spatial resolution. Recent advances in optical technology and probe sensitivity have now opened up the possibility to image cellular activity in living mice. Groups in Monterotondo and Heidelberg are at the forefront of these developments and are generating mice that express novel sensor molecules in a tissue-specific manner. Using these transgenic mice, events such as ion flux, signalling, and tissue damage can be monitored to allow precise readouts of cellular status in disease states in living animals. This approach has led EMBL researchers to the recent discovery of a specialised type of macrophage in heart muscle that is crucial for heart regeneration in response to injury. In the coming years the development and application of live imaging probes to study both the basic and pathogenic mechanisms underlying macrophage tissue surveillance will be a focal point of research. In particular, the technique will be applied to investigate microglia, a macrophage-like cell type that is crucial for clearing neuronal debris in chronic pain states and neurodegenerative diseases such as Alzheimer's and Parkinson's disease in the living mouse.

Live cell imaging is also being used in clinical studies to study the destructive activity of proteases in chronically inflamed lungs of patients suffering from cystic fibrosis (CF) and other forms of chronic obstructive pulmonary disease. These patients frequently suffer from lung emphysema caused by proteases secreted by infiltrating macrophages and neutrophils. In close collaboration with the University of Heidelberg, in the context of the Molecular Medicine Partnership Unit, EMBL scientists have developed a set of fluorescence resonance energy transfer (FRET) reporters that have been successfully used to distinguish healthy mice from those suffering from inflammation in a CF mouse model. This technology is being tested on samples from CF patients and can successfully distinguish healthy and CF patient macrophages (see figure). The activity-based assays are non-invasive and therefore well suited for diagnostics and to monitor disease activity in CF and other chronic inflammatory lung diseases. In future, we aim to develop the FRET reporters further and introduce them to the clinic as diagnostic tools.

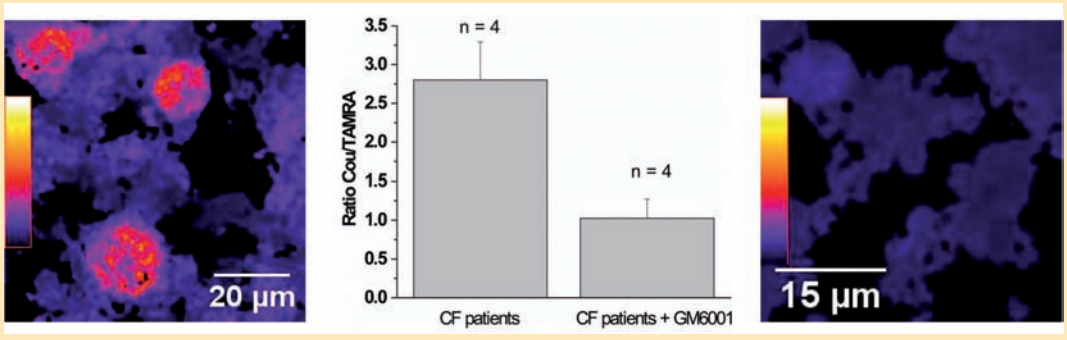


Figure C.2.10: Images of macrophages isolated from bronchoalveolar lavages of cystic fibrosis patients treated with a protease activity probe in the absence or presence of a protease inhibitor. Artificial red and orange colour scales identify activated macrophages.

locus (next-generation) resequencing in the context of MRI scans following heart attacks. A number of known loci with heterogeneous alleles are thought to contribute to the risk of heart attacks. Intensive study of these loci in patients coupled with the extensive public domain information should aid the discovery of new clinical variants and associations with clinical phenotypes.

2.3.3 Infectious diseases

In addition to research on genetic disease, another focus at EMBL will be infectious diseases and their causative viruses, bacteria and parasites. These organisms will both be exploited as model systems for understanding cellular function and studied from the point of view of their medical consequences. At present, several of the structural biology groups in Grenoble, Hamburg and Heidelberg are studying disease microorganisms, taking full advantage of the new technologies of cellular structural biology, systems biology and chemical screening for novel antibiotics.

Tuberculosis has reappeared as one of the deadliest human infectious diseases worldwide through the recent emergence of multi-drug resistant and hyper-virulent *Mycobacterium tuberculosis* strains and the high mortality in HIV-positive patients as a result of *M. tuberculosis* infection. Innovative and novel multi-disciplinary approaches to the physiology of *M. tuberculosis* are required to develop new strategies to combat the disease. A previous EU-funded structural genomics project in which researchers at EMBL Hamburg participated solved the high-resolution structures of close to 50 target proteins of *M.tuberculosis* using X-ray crystallography. These proteins exhibit significant variations in expression pattern in infected patients, and ongoing work is now aiming to elucidate their molecular functions. In this way a completely novel type of enzyme, a cysteine/lysine dyad acyltransferase, has already been discovered, which provides a new starting point for future target-based drug discovery.

Building on these successes and on the experience gained in the systematic analysis of multiple facets of *Mycoplasma pneumoniae* (Box C.2.1), EMBL groups in Hamburg and Heidelberg, together with groups from the EMBL-CRG Systems Biology Partnership Unit in Barcelona, have formed a new consortium aimed at the systematic analysis of several aspects of *M. tuberculosis* during different stages of its life cycle. Transcriptomic, proteomic, metabolic and structural studies will shed light on mechanisms of the bacterium's physiology and pathogenicity. The data produced will provide the basis to model different stages of the *M. tuberculosis* life cycle and its response to external stimuli from the human macrophage host environment. The insights gained might have the potential to inform the design of new antibiotics.

In Grenoble, EMBL has teamed up with the CNRS and the Université Joseph Fourier to form the International Unit for Virus-Host Cell Interactions, where in the coming years various collaborative projects on negative-strand RNA viruses, including influenza virus, rabies virus, HIV and other viruses relevant to human health, will be pursued. The successful recent elucidation of domains of the influenza virus polymerase using X-ray crystallography will in future be extended and complemented with electron microscopy studies and live-cell fluorescence microscopy, which together will form the basis for new efforts in anti-viral drug design (Box C.2.11).

Further future plans in the area of infectious diseases include analysis of HIV assembly in human cells using cross-correlative light and electron microscopy (Box C.2.2), functional genomics approaches to understanding the transmission of malaria and integrated structural biology studies on innate immune responses to viral infection, focusing on intracellular pattern-recognition receptors for viral RNA, the associated signalling pathways that lead to interferon production and on the recognition, activation and signalling of cell-surface receptors. Research on pathological bacteria and viruses has always been an important part of EMBL's research activities and its importance for human health, as well as the usefulness of pathogens as tools to illuminate human host biology, warrant continued activity in this area.

Box C.2.11 Studies of influenza RNA polymerase

In recent years, thanks to the emergence of first avian flu and then swine flu, the potential danger to human populations of new influenza pandemics has been in the headlines. This danger might be mitigated by the development of effective influenza drug treatments. Targets for new drugs have however been difficult to identify and to analyse. The influenza virus RNA-dependent RNA polymerase plays a central role in the virus life cycle and therefore may provide a good target for new antiviral drugs. It is responsible for both replication and transcription of the viral RNA genome and its ability to adapt to interacting cellular host molecules is an important factor in inter-species transmission of the virus, the cause of influenza pandemics. One way to promote drug development would therefore be to obtain high-resolution structural information on the polymerase to aid in drug design. However, the polymerase resisted attempts at structure determination for roughly 25 years. Thanks to EMBL's new ESPRIT technology parts of the polymerase have recently been expressed and their structure successfully solved. In the coming years, EMBL scientists will build on this major advance to pursue the goal of obtaining a high-resolution structure of the complete trimeric functional enzyme to enable a concerted effort in structure-based drug design.

The polymerase is a complex and dynamic machine with multiple functional states that will be studied using a combination of cryo-electron microscopy and crystallography. The insights that will be gained from the structure of the enzyme will be an important step towards a comprehensive understanding of the transcription and replication of the virus at an atomic level.

These structural details will be complemented by cell biology approaches, employing fluorescence microscopy and cross-correlation spectroscopy, to decipher the transport and assembly of the polymerase complex as well as its interaction with host cell molecules in infected cells. A crucial milestone towards this goal is the construction of recombinant influenza viruses with genetically encoded fluorescent viral proteins that will allow their dynamics to be monitored in living, virally infected cells using correlative electron and light microscopy. Combining these interdisciplinary techniques into a cellular structural biology approach will reveal the molecular mechanisms and dynamics that underpin the infection process and might help to identify additional host proteins for drug targets as well as new strategies through which drugs might prevent or interfere with influenza virus infection.

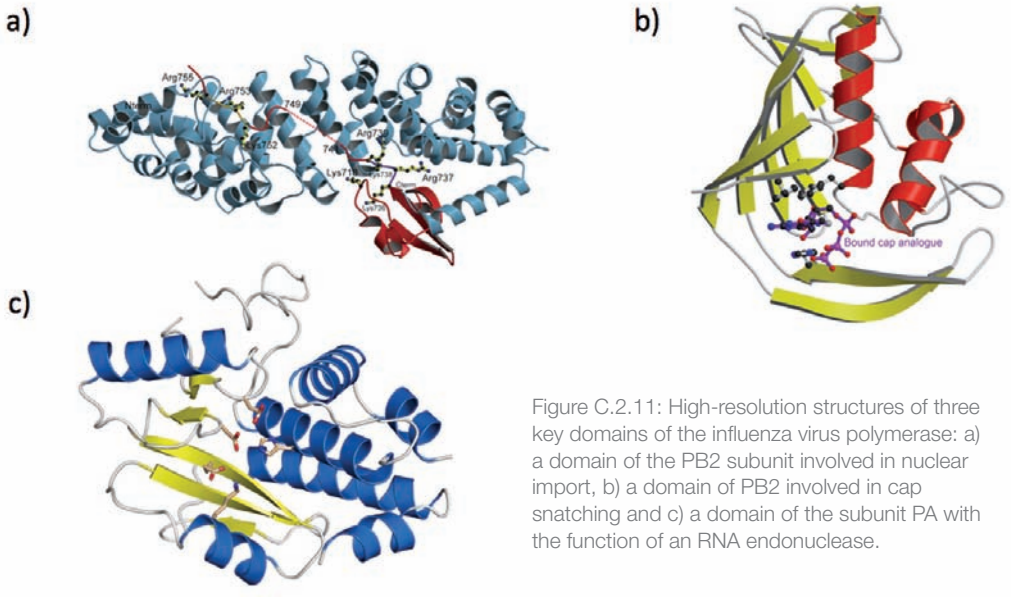


Figure C.2.11: High-resolution structures of three key domains of the influenza virus polymerase: a) a domain of the PB2 subunit involved in nuclear import, b) a domain of PB2 involved in cap snatching and c) a domain of the subunit PA with the function of an RNA endonuclease.

2.4. The need for and use of Bioinformatics and Computational Biology

The complexity inherent to all biological systems reflects the fact that they arose via evolution, as a result of random changes followed by selection, rather than via design and engineering. This complexity makes biology a very information-rich science and computational approaches (bioinformatics, simulation and statistical analysis), which offer ways to disentangle intricate networks of heterogeneous components, are moving to the heart of modern biology. One major reason for EMBL's ongoing success is that it has been a centre of excellence in computational biology for many years, with diverse world-leading research groups in this field based in EMBL-EBI, EMBL Heidelberg and EMBL Hamburg. In addition, roughly 40% of the time spent by EMBL researchers in "wet labs" is devoted to computational work. In the current Scientific Programme we have therefore initiated a policy of embedding groups with a computational focus in all EMBL Units so that expert input and advice is readily available. Some of the fruits of this policy have been referred to in previous sections and readers of this document will quickly note how crucial computational expertise is to almost everything we plan. Here we provide an overview of EMBL's computational strategy and the synergistic benefits gained from the broad expertise within the laboratory across many different areas of bioinformatics, simulation and modelling.

2.4.1 Bioinformatics research is changing fast

The past five years have seen major changes in bioinformatics research, driven by the availability of novel experimental technologies that generate datasets with a size and comprehensiveness that were out of reach even a few years ago. The new DNA-sequencing technologies, which continue to reduce the cost of sequencing, are having a major impact on biological research. Increasingly, sequencing is used to measure molecular concentrations, to map intermolecular interactions, to provide medical diagnoses, to follow the effects of drugs and to measure biodiversity by using metagenomics. Aside from opening up vast new areas of biology to molecular approaches, each of these applications has required the development of new computational tools to manage, integrate and interpret the data. The advent of genome-wide association studies has greatly expanded the link between genotype and phenotype, especially in relation to disease. Other new "omics" technologies, such as proteomics and metabolomics, are continuing to develop and biobanks, which generally store human samples and data, are proliferating. As human biobanks become more organised, this will allow improved access to the data generated using both patient and normal tissues, accelerating the understanding of linkages between model organism and human biology. Imaging technologies, which are allowing quantitative analysis of new problems all the way from cell biology to whole organisms, provide enormous challenges in image analysis and data handling. Unrestricted access to the scientific literature is beginning to be available, making text mining an important tool for the future and the development of better text mining tools a challenge for the present. Parallel changes in computing over the past five years, for example with the development of web services and cloud computing, offer new opportunities for distributed data resources and associated research.

Whilst all these technological developments provide huge challenges for the data service infrastructure, they also present amazing opportunities for *in silico* research and hypothesis generation and testing. Over the next five years we foresee a significant further strengthening of the role of computational approaches throughout biology. In particular we highlight the continued importance of data integration, the rise of chemical biology to explore and understand the role of small molecules in biological function, the increased power of comparative cross-species analyses, the rapid shift from single-molecule analysis to network systems biology, the need to understand the link between genotype and phenotype, and with this the gradual move from basic biology to medicine in translational research. All these developments will lead to an increasing reliance on *in silico* analysis and modelling. To address these upcoming changes, EMBL has recently recruited computational biologists with research interests in genome biology and epigenetics, stem-cell differentiation, cheminformatics and systems biology.

2.4.2 Future vision

During the next five years the scope of our bioinformatics research will be broad, to meet the challenges presented above. The majority of experimental groups will increasingly have some computational expertise and we foresee a further increase in research groups with expertise in both experimental and computational biology. However a core of research groups whose major focus is computational, and who provide the in-depth expertise that is crucial

for the rest of the Laboratory, will remain. These groups will continue to interact closely with both internal and external "wet lab" scientists. They will also play a leading role in providing a link to the growing bioinformatics research community throughout Europe. The interactions between these EMBL groups and the teams that manage the service resources at EMBL-EBI are strong, and not only help to link the data resources and their research communities but also, and very importantly, the research group leaders play a role as expert users to contribute to and influence the development of the data resources. This helps ensure that future trends are spotted early and acted on appropriately. As already described in previous sections, the major biological areas of computational research will reflect both the experimental priorities of the Laboratory and the data provided by EMBL-EBI services.

Research in the bioinformatics-led groups at EMBL will focus on the following areas:

- Understanding genome function and evolution, including the consequences of both natural variation and somatic mutations. Datasets from projects such as the 1000 Genomes Project and the International Cancer Genome Project will form the underlying resources for both human genetic disease correlation and cancer research. EMBL's close collaborations with major data generators in these areas, in particular the Wellcome Trust Sanger Institute in Hinxton and the German Cancer Research Centre in Heidelberg, will allow our bioinformatics research to occur at all levels of analysis, from signal processing methods through sequence alignment and assembly algorithms to functional annotation. The challenges in processing and understanding this data range from the sheer quantity produced, which often requires novel algorithm development, to its integration with the extensive set of molecular, phenotypic and model organism data resources. EMBL bioinformatics researchers will continue to work in close collaboration with high-throughput data producers and develop new algorithms and statistical methods to accelerate the discovery of new biological features, often of disease relevance, in these datasets.
- Applications of new sequencing technologies have led to an unprecedented and unexpected change in our conception of genome organisation and function. To date, this is particularly the case in relation to our understanding of transcriptional regulatory mechanisms. Bioinformatic analysis of functional genomics data have been pivotal to the discovery and quantitation of these processes, which include pervasive transcription, new classes of regulatory RNA, coordinated activity of transcriptional regulators, and a much deeper appreciation of the diverse and widespread roles of epigenetic modifications. Over the next five years, computational biology approaches will continue to analyse these new findings, attempt to reveal their contribution to organism function and in our view will doubtless bring yet more novel information to light.
- Biomolecular networks will be expanded from protein–protein and protein–DNA regulatory networks to capture not only more about RNA but also information on carbohydrates, lipids and metabolites. They will also be developed towards the incorporation of temporal and spatial information in the distribution of these components in order to produce realistic models of aspects of cellular function (Box C.2.1)
- High-throughput imaging is becoming an essential tool for systems biology (Box C.2.3). EMBL is embracing the research challenges in image analysis, computer visualisation and modelling to explore the potential of using image data to help understand the effects of phenotypic variation on function.
- Structural data can be used to understand how molecules perform their functions and predict how individual genetic variation as well as small metabolites and drugs modulate these functions. An important unsolved problem is to understand and predict interactions between molecules, both large and small, and their role in the formation of functional complexes.
- With the advent of databases that integrate chemical and biological data and cheminformatics expertise within EMBL, new challenges will be tackled. For example, developing new methods to provide a measure of "druggability" for pathogenic organisms, to identify alternative uses for known drugs, to discover the molecular basis of drug side effects, to predict an organism's complement of and ability to interconvert small molecules (the metabolome) from its genome sequence and improvement in the prediction and design of small molecule interactions with biological macromolecules based on existing knowledge.
- Every biological system is determined by its intrinsic properties (including the genetic material available) and its environmental context. The latter aspect is increasingly becoming accessible to study via, e.g. generating quantitative data on the organisms present in an environment (metagenomics) or the integration of data from other disciplines (medical or climate data, for example). Incorporation of environmental data into our analy-

ses of biological systems promises to considerably advance our understanding of the properties and functions of individual cells, species and species communities. This is a major long-term challenge that we will begin to pursue in the next five years (Box C.2.8).

- Quantitative modelling and simulation of metabolic networks, signalling pathways and gene regulatory networks will continue to increase in importance and will start to produce useful biological models in significant number. These models will cover larger spatial and temporal scales, reaching to coarse-grain simulations of macromolecular complexes as well as to descriptions of organelles and sub-cellular compartments. To better understand complex functions, such as cell division, "simple" aspects of brain function, morphogenesis or complex diseases such as cancer, models will be developed at EMBL based on the wide diversity of quantitative data generated and distributed by the Laboratory as well as data available from other sources. In future we expect *in silico* experiments involving computational models to become a routine possibility in hypothesis-driven research and that EMBL will be at the forefront of this development. Apart from their predictive power computational models also provide an efficient way to integrate knowledge. They serve as minimal summaries of biological systems or processes and are useful tools to store and exchange information. To facilitate this information sharing we are developing standards for the design, curation and annotation of models, including the widely used Systems Biology Markup Language (SBML) and the more recent Systems Biology Graphical Notation (SBGN).
- Although life science data resources continue to grow in number, size and diversity, the single most useful resource is the scientific literature. However, because of the way the literature has been traditionally organised, access to published information is extremely inefficient. The sustained growth of open-access literature will begin to deliver content that requires integration with the other life sciences data resources. In the years to come, joint efforts between publishers and both research and service-oriented bioinformaticians will need to increase to allow the development of standards and protocols for automatic literature analysis. If this succeeds, it will greatly increase the potential efficiency of scientific research and thus ensure maximum return on the public investment in research.

Finally, a key strength of bioinformatics is its unique ability to integrate and compare data from multiple sources. Therefore, the key feature of EMBL's bioinformatics research will be data integration, interpretation and representation. Such integrative research is described throughout this Scientific Programme and without the contribution of computational biologists it cannot succeed. To ensure a cohesive computational strategy and a coordination of activities the ambitious plans outlined above will be supported by the creation of three new EMBL Centres with a computational focus; the Centre for Mathematical Modelling, the Centre for Statistical Computing and Data Analysis and the Centre for Biomolecular Network Analysis (Section C.3).

3. Initiatives to foster interdisciplinary collaboration and exchange at EMBL

EMBL's research focus and plans for the 2012-2016 period will require an unprecedented degree of interdisciplinary collaboration. For this reason EMBL, over the course of the current Programme, has increased the breadth of scientific expertise available in-house by recruiting physicists, chemists, computer scientists, mathematicians and engineers and has developed collaborations and partnerships with clinicians in. Equally important is the continuous exchange of knowledge and the provision of mechanisms to encourage and facilitate joint projects by groups with complementary skills. EMBL's collaborative spirit has always been a defining characteristic of the Laboratory (Figure C.3.1) and provides an excellent starting point, but institutional mechanisms can also help to overcome the practical difficulties involved.

To foster exchange and interdisciplinary collaboration, we have implemented a variety of mechanisms from regular, thematic seminars and retreats to joint appointments of faculty members between research Units. An instrument that we initiated on a small scale in the current Programme, and that has proven extremely powerful in stimulating collaboration, is the new EIPOD Programme, which recruits postdoctoral fellows to work on interdisciplinary projects jointly coordinated by (at least) two group leaders with different scientific expertise and skill sets. The initial pilot phase of the scheme has been very successful and we intend to build on this during the next Programme (see Section E.3). An application for EU co-funding of the EIPOD Programme was successful and this will support some of the EIPODs from 2009 to 2013. This means that only limited EMBL internal funding will be required in the current Programme but that this support will need to be increased through to 2016 if the scheme is to continue at what we consider a useful level, roughly 25 EIPODs per year. We also plan to further promote cross-disciplinary projects involving co-advisors from different scientific backgrounds in the International PhD Programme (see Section E.2). These

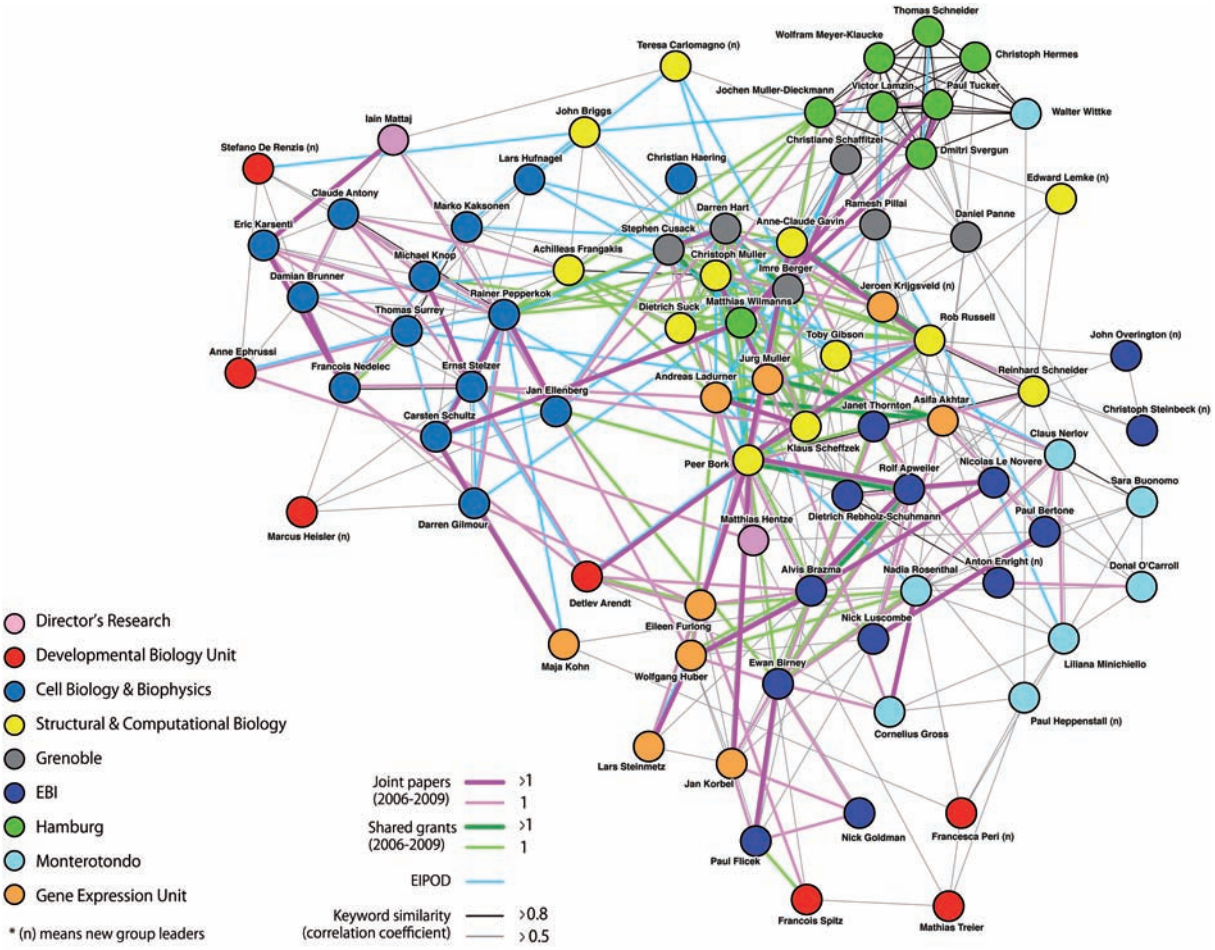


Figure C.3.1: Networks of internal collaboration at EMBL 2006-2009. Different degrees of collaboration are indicated by joint publications (pink), shared grants (green) and EIPODs (2007-2009, blue) affiliated with two or more research groups.

new initiatives for pre- and postdoctoral fellows will encourage a higher level of internal collaboration at EMBL. Very importantly, they will also help to foster the education of a new class of scientists whose training leaves them optimally prepared for the increasing complexity and cross-disciplinary nature of future life science research.

3.1 EMBL Centres

During the current Programme we introduced EMBL Centres to promote collaboration in interdisciplinary areas of technology that span several, and frequently all, EMBL Units. EMBL Centres are transient structures formed bottom-up as ad hoc interest groups to network and encourage the exchange of ideas and the acquisition of expertise. This is achieved through seminars, conferences, training activities and, where applicable, by engaging in collaborative instrumentation development or purchase and interdisciplinary research projects. The EMBL Centres are, in the first instance, tools to promote internal collaboration across disciplines. However, they also provide a platform through which EMBL scientists can interact with like-minded external communities and raise EMBL's visibility to these groups.

Initially, in 2007, we created four Centres: the Centre for Molecular and Cellular Imaging, the Centre for Computational Biology, the Centre for High-Throughput Functional Genomics, and the Centre for Disease Mechanisms. The Centres are organised by small committees of two to four senior scientists or group leaders and participation in Centre activities is open to all EMBL scientists.

All four Centres have built up substantial participant communities and have initiated a variety of new interdisciplinary activities at EMBL. Most notably, the new EIPOD Programme grew out of the Centres and its pilot phase was partly funded using Centre resources. As intended for such flexible bottom-up structures, the individual Centres naturally evolved to fulfill different roles and to address the specific needs of their respective user communities. The Centres are meant to address acute requirements rather than to become permanent structures. This means that although there is a sustained need to continue the activities of some of the Centres into the next EMBL Programme, the original Centres themselves are often no longer the most useful means of supporting those activities that need a stable framework. For this reason, successful Centre activities have been integrated into other organisational structures. Aspects of training, technology development and equipment provision have often been taken over by the relevant Core Facilities and joint research projects have been completed or transferred into EMBL's Research Units. As described below, the Centre model has proved extremely useful to EMBL and will be continued in the form of new Centres in the 2012-2016 period.

3.1.1 New Centres for 2012-2016

Biology is a data- and information-rich science, and computational approaches are important enabling technologies in modern molecular biology. One of EMBL's major strengths has been its emergence as a centre of excellence in computational biology, owing to significant recruitment in this area over a number of years. In this Scientific Programme, we have pursued a policy of embedding computational research within all EMBL Units, to ensure that this expertise is in place to tackle the scientific challenges we address. A measure of the need for expertise and support was provided by a survey of the time spent on computation (data analysis, bioinformatic analysis and data modelling) by members of EMBL Heidelberg experimental research groups in late 2009. To our surprise the resulting figure was 40%! Expert support for these groups is essential if they are to operate at the cutting edge computationally as well as experimentally. To facilitate the use of computational biology resources across the EMBL scientific community and provide focussed support to experimental groups, we will reorganise the broadly based Centre for Computational Biology into three new Centres for the 2012-2016 Programme: the Centres for Statistical Data Analysis, Biomolecular Network Analysis and Mathematical Modelling.

3.1.1.1 Centre for Statistical Data Analysis

Biology is becoming a discipline in which most hypothesis generation, inference and interpretation relies on computational and quantitative methods. Biologists are faced with the challenge of not only analysing their own primary data, including data from high-throughput imaging, sequencing, and proteomics, but also of interpreting these data in the context of a seemingly overwhelming amount of published data and public database content. Often, the efficient mining of available data sources is crucial for the quality of the hypotheses that underlie the design of the next experiment. The unprecedented size and complexity of this task requires expertise and software with which many researchers are unfamiliar – for example, for the automation of analysis workflows, machine learning, statistical discovery and probabilistic inference of facts in order to derive biological insights.

To aid EMBL research groups with this task, the new Centre for Statistical Data Analysis will:

- Organise courses and workshops on statistical and data-mining methodology for biological applications and an EMBL internal seminar series on technical aspects of data analysis.
- Foster contacts and synergies with other institutes with complementary expertise in statistical data analysis, to engage in method and technology exchange and organise short visits.
- Provide an intellectual home and technical support for computational and statistically working scientists embedded in experimental groups and assist with their recruitment and professional development.
- Advise experimental groups on data-analysis methodologies with the initiation of collaborations.
- Provide hands-on collaborations with flagship projects of general or methodical significance.

- Develop methods for recurrent or generally important tasks.
- Be a hub for the build-up and dissemination of expertise across EMBL.
- Maintain an up-to-date installation of computational statistics software tools on EMBL's server cluster; write and maintain online tutorials on how to perform common analytical tasks; and provide user inductions for new EMBL staff.
- Contribute to the maintenance of the "R" and "Bioconductor" project software repositories through package reviewing and editorial processing, and by providing technical help to contributing authors. A European mirror of the Bioconductor software repository will also be maintained. (R is a computational statistics platform and Bioconductor is a large international software project for genomic data analysis.)

3.1.1.2 Centre for Biomolecular Network Analysis

To achieve a holistic understanding of biological processes, systems biology frequently applies a range of complementary large-scale methods to characterise the relationship between the biomolecules that comprise living systems. The construction of networks is a common approach to record and communicate large biological datasets. Networks display molecular parts lists in the context of their interactions and help to infer new functions for genes or proteins. Currently, thousands of networks that document millions of interactions between proteins, genes and small molecules are available. This vast quantity of highly heterogeneous data has to be analysed and integrated in order to generate meaningful new hypotheses about the system. Tools for interpreting and exploring large biological networks have been developed and some have been integrated into public databases. However, data integration and the analysis of complex biological networks require high-level computational expertise. To make expertise in network analysis available throughout EMBL, we will create a Centre for Biomolecular Network Analysis. This Centre will focus on a level of complexity beyond that of the analysis of raw data, instead integrating data into existing networks based on the combination of available heterogeneous datasets. It therefore takes up its work at the point which the support provided by the Centre for Statistical Data Analysis reaches its limits. As already discussed in detail in Sections C.2.2.2 and C.2.4, there is an urgent need to develop new methods for the representation and visualisation of complex datasets. The Centre will be involved in both pursuing such developments and coaching non-experts at EMBL in the use of the best tools that have been developed elsewhere.

The mission of the new Centre is therefore to disseminate expertise and provide guidance in the field of biological network integration and analysis throughout EMBL to computational biologists and experimentalists alike. For expert users, it will serve as a platform to share resources, exchange expertise and learn about advances and new approaches in the computational analysis of networks. At the same time the Centre will offer support to experimentalists with less computational expertise who wish to pursue large-scale biology or to place results obtained in small-scale mechanistic experiments in the context of existing networks. The education and training of EMBL researchers in data integration and network biology is a central mission of the Centre.

In particular, the Centre for Biomolecular Network Analysis will engage in the following activities:

- A biweekly meeting series featuring presentations by EMBL scientists using or developing approaches to analyse large datasets.
- Training in the use of databases and software for biological network representation and analysis.
- An annual Network Biology Retreat for scientific exchange and strategic discussions on scientific proposals, suggestions for educational activities, the organisation of symposia and the invitation of external speakers.
- Symposia and courses featuring invited external speakers and trainers.
- A help desk run by computational scientists to provide assistance with individual projects requiring network analysis, for example helping large-scale biologists with the analysis and integration of datasets and advising on methods and software for data integration.

3.1.1.3 Centre for Mathematical Modelling

A major focus of our research plans for the period of the next Programme is to develop more predictive networks, models and simulations of biological processes. Good models are defined by the fact that they can generate testable

hypotheses, predict the quantitative outcome of perturbations and serve to guide future experiments. Models are the ultimate way to test whether we understand biological processes and those that consistently provide accurate predictions of the outcome of perturbations provide the best way to describe biology.

Generating computer models and simulating biological processes using fundamental physical and chemical principles expressed in mathematical language requires a high degree of specialist expertise in mathematics and physics. During the current Programme we have started to recruit this expertise to EMBL and have embedded physicists, mathematicians and computer scientists who pursue this form of analysis into several research groups. As a result, a variety of activities in mathematical modelling have been initiated at EMBL in the past two or three years:

- Training courses in relevant software packages from resources such as "Matlab" and "R" and a computational modelling course organised by EMBL-EBI.
- Efforts to improve the standardisation, exchange and dissemination of biological models: EMBL-EBI's BioModels database and contributions to the Systems Biology Markup Language (SBML) and Systems Biology Graphical Notation (SBGN).
- Regular physical biology meetings with other institutes in the member states.
- Courses for physics undergraduates who wish to be exposed to biological research.

In future, we aim to bundle these activities under the umbrella of a dedicated Centre for Mathematical Modelling. The Centre will provide a home to the specialists in fields other than biology, who require further intellectual support structures for guidance, exchange and specialised training. Additionally, new lectures and practical courses will be made available so that those coming to EMBL from disciplines outside the life sciences can keep up with the progress in their fields. A seminar series dedicated to theory will be held biweekly for all EMBL scientists interested in keeping up-to-date with activities in mathematical modelling and a regular retreat will provide a platform for exchange, the identification of common research interests and the initiation of collaborations. The Centre for Mathematical Modelling will serve to cultivate a "physical biology spirit" at EMBL, which will increase EMBL's visibility outside the life sciences and help recruit talented physicists and mathematicians and foster more interdisciplinary collaborations in future.

3.1.1.4 Centre for Chemical Biology

Although the Centre for Chemical Biology was initiated in 2010, we describe it here because it will be active throughout the 2012-2016 period. Chemical biology is a crucial enabling technique for cell biology, systems biology and synthetic biology. Its major purpose at EMBL is the development of new tools to perturb and manipulate biological systems including, but not limited to, small molecule modulators. Chemical biology also provides reporter molecules that allow the visualisation of molecule locations, dynamic biological processes and enzyme activities in cells and organisms. In addition, EMBL-EBI's new initiatives in chemoinformatics provide a means to integrate chemical information with biomolecular data.

EMBL has substantially expanded its expertise in various branches of chemical biology from one research group in organic chemistry in 2004 to seven groups that cover a broad spectrum of chemical biology and cheminformatics in 2010. In addition, the Chemical Biology Core Facility provides complementary expertise and technology for chemical screening and the development of assays and inhibitors. EMBL-EBI maintains two major databases for chemical biology: ChEBI for small molecules of biological interest, and ChEMBL for data relevant to drug discovery. These are complemented by in-house virtual compound collection databases and software used for virtual and physical screening studies. Taken together, our in-house chemistry expertise currently encompasses preparative organic chemistry, peptide, nucleotide, lipid and carbohydrate chemistry, microarray preparation, *in vitro* and *in vivo* labelling, lipid manipulation and analysis, chemoinformatics including database creation, native chemical ligation, the incorporation of artificial amino acids (expanded genetic code), genetically encoded fluorescent reporters, small compound screening, molecule docking and virtual screening, NMR analysis of small molecules and small molecule–protein interactions, and proteomics. This spectrum provides EMBL scientists with an unusually broad range of molecular and computational techniques that are useful for cell, structural and developmental biology, which we are planning to build on in future.

To ensure maximal contact and integration with wet-lab biology, the chemical biology groups at EMBL are embedded in the Research Units. Although this structure has the advantage of close interdisciplinary collaboration and application-oriented development of tools, it means that special efforts are needed to enable subject-specific intellectual exchange and training for those involved in chemistry. To address this issue, we will establish a Centre for Chemical Biology at EMBL, which will foster exchange and interactions among the chemistry-oriented groups and inform the EMBL research community about new developments within the groups at EMBL and in the entire field. We also intend to make supervised chemical synthesis facilities available to scientists from groups that lack expertise in chemistry.

The Centre will organise the following activities:

- An annual Chemical Biology Retreat for EMBL group leaders. Apart from scientific exchange, this event will be instrumental for the Centre to receive input from group leaders and to coordinate educational and other activities.
- Symposia and courses related to research in chemical biology.
- There is a growing demand for access to synthetic chemistry throughout the Laboratory, some of it driven by interdisciplinary EIPOD projects. We therefore propose to install dedicated visitor laboratory space attached to the existing chemistry laboratories where EMBL group leaders may place co-workers to perform all kinds of chemistry experiments under the supervision of expert chemists.
- The Centre will contribute to the organisation of the biannual EMBO conference 'Chemical Biology', currently the largest event of its kind in the world.
- Internal education in chemical biology. The Centre will provide advice on how to i) use and set-up compound screens, ii) study ligand binding by NMR, iii) perform fluorescent labelling, iv) outsource synthetic needs, v) use mass spectroscopy to analyse artificial biomolecules from cell sources, vi) perform radiolabelling, and vii) when to seek intellectual property protection. Finally, the bioinformatics groups of the Centre will coordinate training in software and databases and provide a constant overview on availability and license requirements.

3.1.2. The mature Centres

As already mentioned, four EMBL Centres were created in 2007. Apart from the Centre for Computational Biology, whose evolution and redirection has been described above, these were the Centre for High-Throughput Functional Genomics, the Centre for Molecular and Cellular Imaging and the Centre for Disease Mechanisms. Their current and future forms will be discussed below.

3.1.2.1 Centre for High-Throughput Functional Genomics

Founded in 2004, its aim was to promote functional genomic experiments at EMBL through:

- Building the infrastructure to physically carry out large-scale experiments at EMBL.
- Providing a forum to promote exchange of technical expertise and scientific knowledge on genomics approaches.
- Increasing the awareness and synergy of high-throughput projects being carried out at EMBL.

The Centre has been instrumental in establishing functional genomics as a commonly used research tool in many EMBL Research Units. High-throughput approaches have been applied to a variety of biological projects, including studies of the regulation of muscle and eye development, the definition of sites of genetic recombination in meiosis, the discovery of components required for reassembly of the nucleus after cell division, the identification of networks of genes whose regulation is crucial for stem-cell renewal, and the association of functional data with large fractions of the yeast and human genomes. The Centre also established a microarray data analysis and storage database (EmBASE) that allows biologists to analyse their data, keep records of all analysis steps, share and integrate data and transmit it to ArrayExpress, EMBL-EBI's database for gene expression data. The Centre hosts a biweekly seminar series for computational and wet-lab biologists, to bridge the gap between the laboratory bench and computational data analysis. It also provides the genomics community at EMBL with a forum for informal exchange of methods, protocols and ideas, and serves as a breeding ground for collaborative projects.

The activities of this Centre are related to the services provided by the Genomics and Proteomics Core Facilities (Sections D.3.1.1 and D.3.1.2). Our intention was that, once technologies had been introduced and tested by the research groups that formed the Centre, those that were found to be broadly useful, and robust enough to be offered as a service, would be incorporated into the Core Facility. This has now happened or will soon happen for much of the robotics and other instrumentation (microarray, embryo sorter, etc) and the technology experts who operate them.

Another activity of the Centre that has been extremely useful and very popular is support for the analysis and interpretation of high-throughput sequence information. In recent years the performance of nucleotide sequencing technologies has increased by several orders of magnitude, while costs per nucleotide sequenced have been dropping. In the light of new products now being launched, this process is predicted to continue in at least the near future. Increasingly, data processing and analysis rather than data collection is the bottleneck in experimental projects. A large and crucial component of life science research is therefore informatics support for elementary tasks such as data quality control, alignment, genome and transcriptome assembly, DNA-polymorphism discovery and evaluation, alternative transcript discovery and evaluation, multiplexing and deconvolution and abundance quantitation. The second mission of the Centre for High-Throughput Functional Genomics that will be maintained in the next Programme, in the context of the Genome Biology Unit, is the support of these tasks by:

- Testing, evaluating and providing competent methodology advice on rapidly emerging new technologies and software.
- The expansion of EmBASE (the EMBL microarray analysis database) to facilitate Solexa next-generation sequencing data storage and analysis
- Maintaining an up-to-date repository of software on EMBL server resources, and curation of problem-centric documentation ("how to" instructions).
- Customised, project-based support of individual groups and researchers in data processing and analysis.
- Being a lively discussion forum on challenges associated with next-generation sequencing and the development of common solutions.

Thus, although the Centre will cease to exist as such, it has created activities that have continuing importance and value for EMBL's research. We will therefore build these activities into the appropriate EMBL Core Facilities and Genome Biology Unit, and redeploy the Centre funding towards this goal.

3.1.2.2 Centre for Molecular and Cellular Imaging

The Centre was founded with the aim of bringing together communities involved in various aspects of biological imaging and to promote the use of imaging technologies at EMBL. Data processing and analysis forms the biggest bottleneck in most imaging projects. Therefore the Centre adopted an important role in providing various levels of support for image processing:

- The Centre is run by a dedicated member of staff with expertise in data processing. He offers hands-on support and helps develop solutions on the basis of individual projects. His activity is supported by several image processing experts in different groups (e.g. the Cell Biology and Biophysics Unit, Mitocheck and the Advanced Light Microscopy Facility) who spend a significant part of their time on Centre tasks.
- The Centre engages users from different Units in joint tool and software development for image processing and its webpage provides a platform where such solutions are shared and exchanged. Currently the focus is on a systematic collection of Image J macros and plugins as well as high-performance image processing solutions for cluster computing in Python/C++. In addition, the web platform provides documentation and online tutorials for image analysis methods, many links to information resources on the internet and a list of contacts for EMBL researchers and descriptions of their skills and expertise in this field.
- Regular internal training courses on image processing have been organised that address different aspects of image processing and data analysis. These courses are held several times a year and cover basic-, advanced- and developer-level image processing as well as courses in MatLab and Image J programming. The courses have always been oversubscribed. In total, almost 200 EMBL researchers, mainly PhD students and postdocs,

have taken these courses. In addition, the Centre now routinely offers image processing modules at EMBL-organised EMBO practical courses that are related to imaging.

- A biweekly seminar series and regular symposia to bring together internal and external experts in imaging.
- A LIMS infrastructure (CellBase) has been put into place to facilitate the management and joint analysis of high-throughput imaging data (in collaboration with software developers in the Cell Biology and Biophysics Unit and the Centre for High-Throughput Functional Genomics). The LIMS is used by both internal and external users of the Advanced Light Microscopy Facility to handle data obtained using its high-content screening microscopy service.

With its focus on image processing and data analysis the Centre for Molecular and Cellular Imaging is complementary to the image acquisition technology-based services provided by the Advanced Light Microscopy Facility, and addresses a need that, particularly with the increasing use of high-content functional imaging methods, will only grow at EMBL in future.

The Centre will continue to flexibly adapt to any newly emerging needs of the imaging community at EMBL. Particular developments that we foresee include:

- To systematically engage in open source software development communities (such as the Image J project FIJI) to make shared image analysis resources developed at EMBL available to the wider community.
- The development of robust methods and tools for coupling image analysis and computer simulations.
- The development of integrated solutions for storing, managing and analysing the vast data quantities produced by high-throughput imaging and 4D atlas imaging projects of cells and embryos.
- The development of data processing tools for better integration of imaging at different scales of resolution, in particular correlative light and electron microscopy, and the integration of imaging with fluorescence (cross) correlation spectroscopy approaches (see Box C.2.2).

Thus, similar to the Centre for High-Throughput Functional Genomics, the tasks taken on by the Centre for Molecular and Cellular Imaging have matured to the point where it is clear which of them must be continued at EMBL. We have decided to carry these mature activities on in the context of the Cell Biology and Biophysics Unit and the Advanced Light Microscopy Facility (Section D.3.1.6), to which we will transfer the Centre's funding.

3.1.2.3 Centre for Disease Mechanisms

A major challenge for the life sciences in the next decades will be to develop new reagents and methods for the treatment of a large number of human diseases. The understanding of the mechanisms leading to disease pathogenesis will be fundamental to this effort by providing new, personalised, targets and methods for therapeutic intervention. This task will require an interdisciplinary approach bringing together different expertise from basic research and clinical medical science. As a result, the borders between basic and applied medical research will become less clear and will give rise to a new generation of biomedical research laboratories. EMBL, with its unique structure and position in European research, can play a pivotal role as a collaborating partner in this effort.

The EMBL Centre for Disease Mechanisms was established in 2004 to support medically relevant research at EMBL by expanding the interface between EMBL groups and clinical research activities at Europe's medical institutions. The Centre aimed to promote the increasingly important application of basic research to the understanding and treatment of human disorders through molecular and biomedical exploration. To achieve its goal, the Centre organised and sponsored a number of activities:

-The Centre hosted popular EMBL-organised symposia and workshops on medically oriented topics including the "Minisymposia on Molecular Medicine", which foster contacts between clinicians and basic researchers. The Centre hosted one or two such symposia each year at different EMBL sites. They provided external visibility to the preclinical research within EMBL, and sparked collaborations on medically relevant research topics.

- The Centre established institutional partnerships with European clinical institutes to amplify the positive interactions between EMBL and the external research community. Partnerships that entail substantive involvement of EMBL are highly desirable in strategic terms as they are fully consistent with the pan-European mission of EMBL.

Through these partnerships, EMBL contributes its expertise in research, training and institution building, while in turn these partnerships reinforce EMBL's strengths and provide invaluable entrees into clinical research.

- The Molecular Medicine Partnership Unit between EMBL and the Medical Faculty of Heidelberg University is an example of such successful integrating activities (see Section G.2.2.3). Joint training of MD and PhD students, as well as postdoctoral training for MDs and PhDs in the project teams, aims to understand the molecular basis of human diseases, lead to applications in diagnosis and therapy, and provide feedback from 'bedside medicine' into the research laboratory.
- A collaboration with the Heart Science Centre at Harefield Hospital London, part of Imperial College London, was established in 2005 (see Section G.2.3). It is devoted to cardiovascular disease research, broadening the scope of effective links by utilising the complementary experience in cardiac morphogenesis, congenital heart disease, cardiac transplantation and transgenic models.
- The Nordic EMBL Partnership for Molecular Medicine, established in 2007, constitutes three Nodes (Norway, Sweden, Finland) that collaborate closely in the area of molecular medicine (See Section G.2.2.7). It capitalises on Norway's strength in molecular mechanisms of disease, Sweden's strength in microbial pathogenicity and molecular infection medicine, and Finland's strength in integrated clinical and basic biomedical research coupled with population-wide genetic epidemiology and well-characterised databases.

The involvement of EMBL in medically oriented molecular biology, human genetics, and research on the mechanisms of disease pathogenesis will be strongly enhanced through continued partnerships with these and other national clinical institutions.

The foundation of these partnerships has elevated the work of the Centre for Disease Mechanisms to a different level and our intention for the 2012-2016 Programme is to continue these activities in the context of our molecular medicine partnerships.

D. Infrastructure and services to the member states

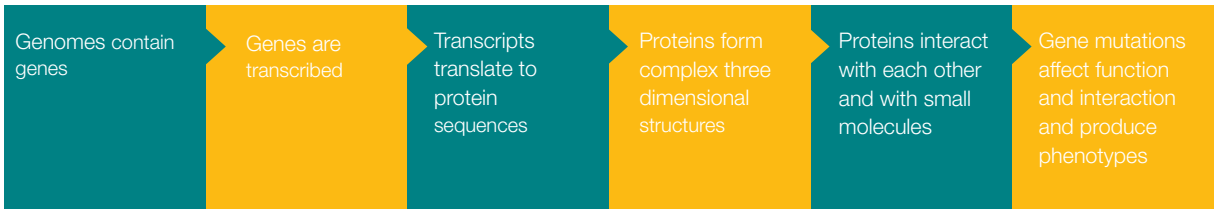
1. Bioinformatics Services and Databases

1.1 The EMBL-EBI service mission

Enabling optimal exploitation of biomolecular information is at the heart of EMBL's mission. EMBL-EBI provides a range of complementary databases that collect, organise, annotate and store core biomolecular information and make it available to researchers in Europe and worldwide (Figure D.1).

In addition, there are 25-30 small specialist databases and supporting resources such as controlled vocabularies. According to a recent survey run in the context of ELIXIR, the emerging European Life-Science Infrastructure for Biological Information, there are roughly 200 databases offered by 100 institutions throughout Europe. The EBI, perhaps not surprisingly, holds the largest data collection and serves by far the largest number of users. EMBL-EBI also forms a European node in several global data-sharing collaborations.

The key databases are organised around what we sometimes refer to as the "central dogma" of molecular biology, a simplifying conceptual framework that goes:



Genomes and Genes	Ensembl	A joint project with the Sanger Institute presenting high-quality, integrated annotation on vertebrate genomes
	Ensembl Genomes	Ensembl-based environment for genome data from other branches of the tree of life
	1000 Genomes	Catalogue of human variation from major world populations
	EGA	European Genome-phenome Archive – Genotype, phenotype and sequence data – particularly those of relevance to human disease
	European Nucleotide Archive	Submitted assembled DNA and RNA sequences. Reads from next generation sequencing (SRA). Traces from next generation sequencing (Trace Archive).
Transcription	ArrayExpress	Archive of transcriptomics and other functional genomics data
	Expression Atlas	Differentially expressed genes in tissue or cell types, disease or under treatment
Protein	UniProt	Protein sequences and functional annotation
	InterPro	Integrated resource for protein families, motifs and domains
	PRIDE	Public data repository for proteomics data
Structure	PDBe	Resource for biomacromolecular structure and function
Small molecules	ChEBI	Chemical Entities of Biological Interest
	ChEMBL	Bioactive compounds (drugs and drug-like molecules), their quantitative properties and bioactivities
Processes	IntAct	Public repository for molecular interaction data
	Reactome	A curated resource of core pathways and reactions in human biology
	BioModels	Mathematical models of cellular processes
Ontology	GO	Gene Ontology for consistent descriptions of gene products
Literature	CiteXplore/UKPMC	Bibliographic query system

Figure D.1. EMBL-EBI's range of complementary databases

EMBL-EBI also offers a range of computational tools to explore the data available in the repositories, for example tools for sequence alignment or pattern searching. These services are used by academic and commercial researchers throughout Europe and the world. Industry users receive targeted support through a dedicated Industry Programme (see Section F.3.2.1).

Usage of the EBI services is monitored in several ways – the most obvious metrics come from web hits at the EMBL-EBI website, the portal to its resources (Figure D.2). In 2010 the average web hits per day reached four million and the services were accessed from about 300 000 unique IP addresses per month. This figure is an indication of the number of user sites, not the number of individual users, which would be much higher. Sampling over a year, we see usage from about a million sites.

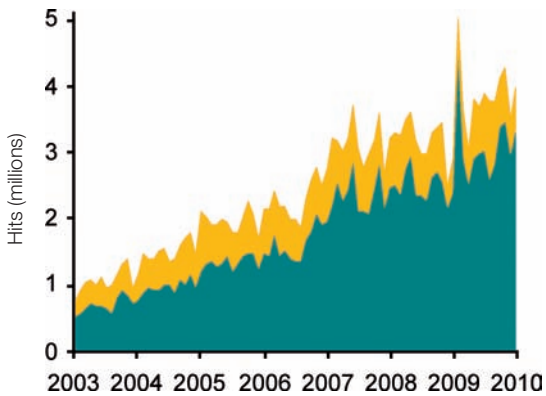


Figure D.2. Use of the EBI website The orange graph includes Ensembl hits (a joint project with the Wellcome Trust Sanger Institute).

Running databases is a resource-intensive task, and scientific advances mean that new methods frequently appear that could benefit from shared data collections. Even though the EBI offers a wide range of data, it must scrutinise suggested additions to its palette and accept only those whose value is beyond doubt. Similarly, the value of existing databases and services must be re-examined routinely. To this end we have developed a framework for reviewing services that considers a range of issues and metrics. These include:

Scientific demand — from the user community, data producers, funding agencies and journals

Appropriateness to the EBI — considering EMBL's mission, any competition, and collaborative context

Usage — actual or predicted

Costs and value for money — in staff and information technology (IT) resources

This means that new projects are not embarked upon lightly, and that existing projects can be discontinued when necessary. For example, in 2010 four databases will be discontinued:

- The Integr8 genome portal
- Genome reviews
- The Alternative Splicing and Transcript Diversity database (ASTD)
- The International Protein Index (IPI)

Outreach and training activities, which are discussed elsewhere in this document (Section E, Box E.3 and Section F.4), help users to make optimal use of EBI resources. The people who provide the EBI services, including outreach and training, as well as those who provide systems and other technical support account for about three-quarters of the EBI staff.

These services enjoy a synergistic relationship with the EBI Research Groups and with the research activities within many of the Service Teams. They benefit from in-house research in several ways:

- Researchers are in-house expert users who exercise tools to the limit, and whose feedback leads to constant improvement.
- The result of research work is often incorporated in new or existing services.
- The scientific dialogue with researchers influences our long-term vision and planning by giving early insight into scientific trends and advances.

The motivation behind, and the commitment to, the service mission will continue undiminished in the 2012-16 Scientific Programme. However, this constancy of mission is accompanied by the need to ensure that our services are continuously evolving to keep pace with changing science:

Ever-increasing data flow rates have perpetuated exponential database growth curves during the past years (Figure D.3.). However, the surge from next generation DNA sequencing has changed the EBI's task in this domain qualitatively as well as quantitatively. For example, molecular biologists no longer just sequence single genomes, but also sets of related genomes in structured studies.

New information resources are necessary in response to new scientific trends. For example, during the current Programme new resources in chemistry and in genotype-phenotype relationships became necessary, and existing resources had to adapt to incorporate data from studies of biodiversity and metagenomics. We can expect similar innovation in this Scientific Programme.

The diversity of 'omics data and the holistic emphasis of systems biology mean that information must be integrated to be of value.

Expanding user groups are emerging from new domains adopting high-throughput biomolecular methods, for example in medicine, agriculture and environmental studies.

Changing information technology increases expectations, for example for the ease of use of scientific user interfaces, and brings challenges and opportunities. Advances in distributed informatics (such as Grid and Cloud

computing, Service Oriented Architectures, and Virtualisation) challenge traditional methodologies. For example, instead of running programs locally users are increasingly utilising algorithms made available via web services to run their computational tasks on EBI machines.

In addition to these foreseeable challenges, unpredictable scientific developments will certainly impose new demands on the EBI during the Scientific Programme. The EBI's mission must combine the maintenance of long-standing, valuable resources, which fulfil an essential role as historical archives for the life sciences, and services for biomedical research with an ability to react to new trends flexibly and quickly. We must not only keep up with science by recording its findings, but also support exploration by evoking 'the next question', thereby contributing to tomorrow's research as much as today's.

1.2 EMBL-EBI's role in ELIXIR

As will become clear below, Europe's future requires a bioinformatics infrastructure whose magnitude and scope must extend beyond the EBI. For these reasons ELIXIR's developing vision depends increasingly on the involvement of infrastructural nodes throughout Europe. This meshes well with the conclusions of the European Strategy Forum for Research Infrastructures (ESFRI), which have enabled this vision to be cast in the form of the ELIXIR project.

ELIXIR, which is coordinated by the EBI, is one of ten pan-European Biomedical Research Infrastructure projects selected by ESFRI for inclusion in its Roadmap (see Section H.1). Its mission is to construct and operate a sustainable infrastructure for biological information in Europe, and to support life science research and its translation to benefits in medicine, the environment, the bio-industries and society. The preparatory phase, which has been funded by the European Commission, involves 32 partners from 13 countries, including most of EMBL's European member states. It will develop a plan for the provision of biomolecular data services across Europe for the next 10-15 years.

The suggested structure of ELIXIR is a network of specialised, national nodes that will be connected to EMBL-EBI, which will host the central hub of ELIXIR. EMBL-EBI will be responsible for holding the core data collections while ELIXIR will be responsible for their connection to and integration with the data held at the nodes, thereby enabling the development, interconnection and integration of ELIXIR into a European-wide distributed infrastructure. Due to its existing resources and expertise, that are unique in Europe, the EBI is the only viable option for the ELIXIR hub. In future, the ELIXIR hub should also take over responsibility for the delivery of the data resources to users, including storage, distribution and provision of access. EMBL-EBI will still be responsible for the production of the databases, data collection and integration of the resources held at EBI.

The prior existence of EMBL as an international scientific organisation has enabled good progress towards a legal framework for ELIXIR, potentially as an EMBL special project. Also, several member states have already made financial commitments to their national ELIXIR nodes and the UK Research Councils have recently provided capital funding for a substantial upgrade to the physical facilities of the EBI to prepare the way for EMBL-EBI to adopt the role proposed for it by the ELIXIR stakeholders, i.e. to act as the European hub of ELIXIR. This is a critical issue for EMBL, and in particular for the EBI, and so it will be discussed in detail separately, in Section H.1.2.1. There is however no doubt that this is a critical task for the life sciences in Europe, and that carrying it out successfully will depend on significant additional funding being made available to the EBI via the EMBL budget. While the creation of ELIXIR requires significant investment to create a hub at the EBI, at the same time it offers a sustainable solution for the problem of how to fund the collection and provision of biomolecular data in future. Decentralisation and distribution of data resources will make the provision of bioinformatics services to European users feasible in spite of the ongoing rapid increase in data production and diversification of data types that will be required.

1.3 Rationalisation, restructuring and integration of services

In the current EMBL Programme we have rationalised and restructured our data resources to better reflect our 'central dogma' from genome to molecular function.

The table of resources (Figure D.1) reflects this structure, as does our web site, which aims to address the problem of cross-database exploitation. The website provides unified access to all EBI resources and a powerful search engine that allows instant searches of all EBI databases via a single query. It generates consistent, up-to-date results, and enables navigation between individual biological entities. Users can browse from genomes to genes, proteins to structures and biological functions from a single, simple interface.

A further reorganisation in 2008 combined our DNA- and protein-sequence activities into a single team with several sub-teams. This has enabled more robust links between the UniProt protein sequence database and the DNA sequence databases, which are exploited in increasingly integrative user interfaces.

We have also restructured the nucleotide resources to better serve data in the genomic era. We distinguish between:

- **The Ensembl family of databases,** which includes Ensembl for higher animals (especially model organisms) and Ensembl Genomes for the rest. They provide genomic data organised and annotated by the in-house Ensembl pipeline, with Ensembl Genomes drawing extensively on community help for the annotation. Ensembl is a joint project with the Wellcome Trust Sanger Institute whereas Ensembl Genomes is an EBI project.
- **The European Nucleotide Archive (ENA),** which includes nucleotide sequence data submitted by the scientific community along with submitter annotation where available. It includes the nucleotide sequence database EMBL-Bank, the Trace Archive for raw capillary sequencing data, and the Sequence Read Archive for unassembled data from next generation sequencing.

Aside from this reorganisation within the sequence domain, the global collaboration in the collection of macromolecular structures was cemented by an agreement forming the world-wide Protein Data Bank (wwPDB) in 2007. In early 2008 wwPDB announced the completion of the "remediation" project – a massive collaborative effort to correct historic errors and inconsistencies in the database.

The developments in the current indicative scheme have been made possible not only through the contributions of the EMBL member states, but also through substantial grant and contract support from the European Union, the Wellcome Trust, the US National Institutes of Health, who have provided 20.6%, 14.4% and 7.7% of the cumulative funding for the EBI in the current EMBL programme to date (2007-2009) and the UK Research Councils, as well as by contributions from our industry partners. We currently predict that these very generous external funders, without whom the EBI would not be functional, will provide roughly 50% of total EMBL-EBI funding in the 2007-2011 period.

1.4 Growth of the EBI

EMBL-EBI's data repositories are growing exponentially. Figure D.3. shows the growth of six mature databases, for which we have longitudinal data. With the advent of various high-throughput technologies and the increasing focus on large-scale systems biology studies we can expect a continuation of the rate of expansion of these databases in the near future.

Despite this growth, the overall staffing for these mature projects has remained fairly stable. As an example we show the recent development of UniProt, the universal protein sequence database (Figure D.4.). In spite of the enormous growth in data flow into this database, staffing levels have been fairly stable over the four years up to 2009. Steady or falling staff levels are common in the mature databases such as those in Figure D.3., meaning that increases in input to established databases do not increase staff costs (although we are changing the source of funding for these costs to increase stability, see below). However, we still must tackle rising IT demands in both compute and storage and the requirements for new data resources such as the new DNA sequence collections in the 1000 Genomes and Trace Archive resources.

The decision to award increased funding to support EBI services in the 2007-2011 Indicative Scheme was largely based on a desire to allow consolidation, using new funds to support existing activities rather than starting new ones, thereby diminishing the risks of over-dependence on insecure funding. The increased staff funding from the internal budget has therefore been mostly dedicated to such consolidation in the core databases and services (Figure

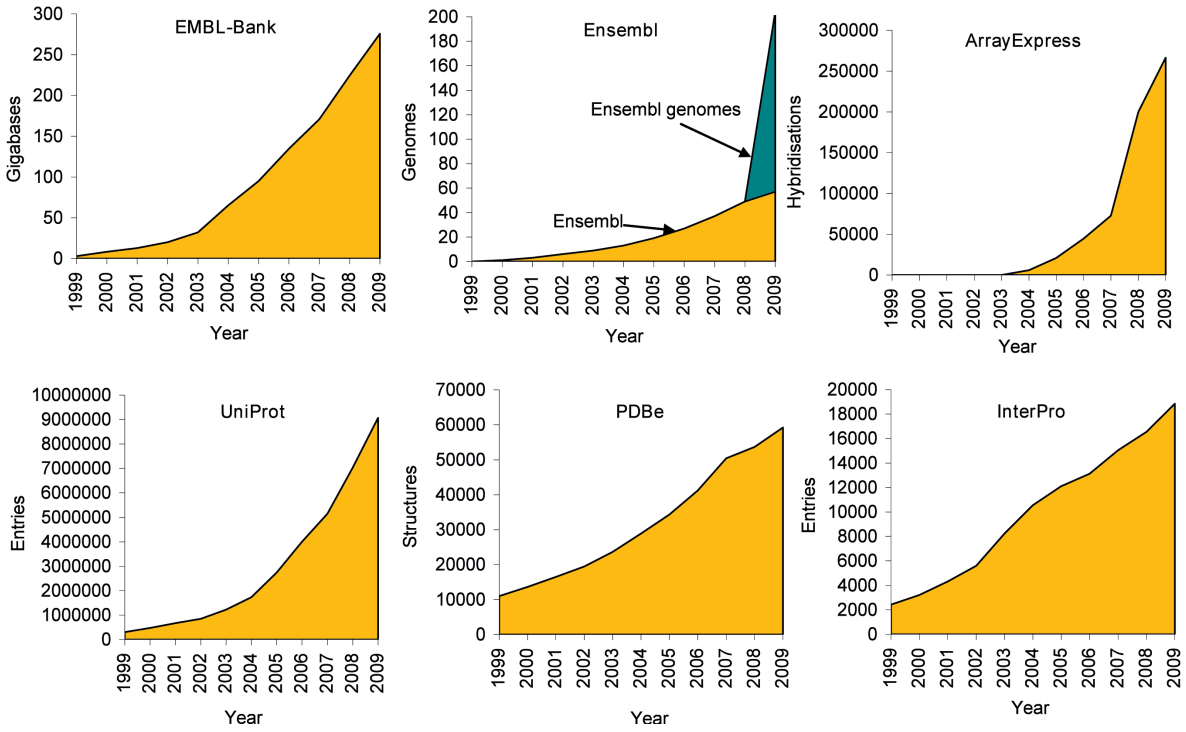


Figure D.3. Database growth for six mature databases

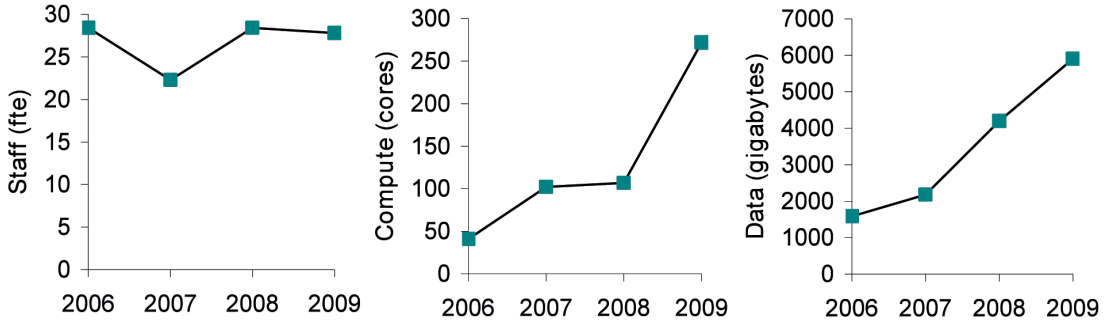


Figure D.4. Total staff, compute and storage trends over four years for UniProt

D.5.). The previously described new activities are reflected in the increment in the number of staff at the EBI that, at the end of 2009, comprised 407 employees. The restructuring of the core database funding decreases our reliance on external support but does not by any means eliminate it. The unpredictability of external support therefore remains an issue for EMBL-EBI future planning. The budget increases secured during the 2007-2011 Indicative Scheme have been mainly owing to generous increments in the EMBL contribution rather than in external sources of funding. Looking to our future needs, we will include projections for the next EMBL Programme in the draft Indicative Scheme to be presented to EMBL Council in 2010-2011. We will include projections for external funding based on cautious estimates derived from our past experience. During the Indicative Scheme 2012-2016 we will also include the projected cost of the ELIXIR hub. Funding for the construction of the new facilities has been requested from the UK government as their contribution to creating the central ELIXIR hub, but the running costs for the hub will be most logically provided by the EMBL member states.

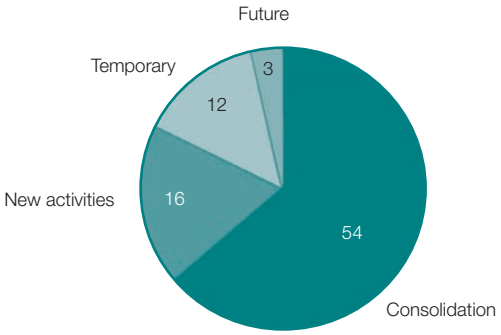


Figure D.5. Posts funded by EMBL budget increments 2007 to 2009 and their allocation to consolidation of existing activities and new activities.

Total of 85 new EMBL funded posts 2007 to 2009

Aside from these staffing increases, the data growth has led to a significant rise in the IT demands at the EBI (Figure D.6.). Part of the increase in storage results from the creation of an offsite replication centre to provide data security and give continuity of service in the event of a major failure on the Hinxton Campus. The first solution to this problem was implemented, following urgent recommendations from SAC, during the current EMBL Programme. A comprehensive solution will be implemented before the end of the current Indicative Scheme but will be costly in terms of storage servers.

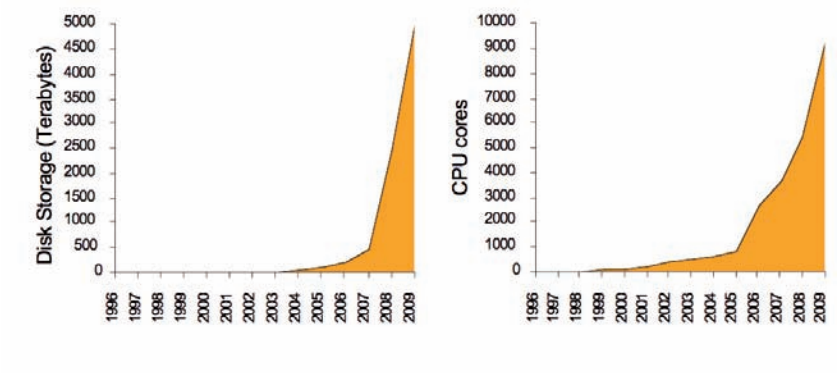


Figure D.6. Growth in EBI IT resources: total disk storage and number of CPU cores

1.5 New challenges and future developments

As a discipline, bioinformatics is undergoing a transition. We have moved from a situation in which many research groups had an embedded bioinformatics expert, to one in which many of today's scientists are trained in bioinformatics and are able to explore biological information space. This changes the task of service providers profoundly. It is no longer sufficient to serve information on different kinds of molecules and processes in the form of separate databases and tools, leaving the local bioinformatician to assemble them. We must offer integrated services that reflect the integrated networks that form biological systems.

Our future is characterised by:

- rapid advances in basic science
- improved high-throughput laboratory methodologies
- expanding user-domains
- increasing applied uptake

These features combine to cause massive increases in the size and complexity of our service mission. Fortunately, IT – the source of solutions to these challenges – is also changing rapidly and is keeping pace with the developing needs of bioinformatics.

1.5.1 Challenges

For the period of the Indicative Scheme 2012-2016 we see the following challenges:

More data — Data volumes are growing fast. Owing to ever-improving high-throughput technologies it is easier to sequence a genome today than it was to sequence a gene a couple of decades ago, resulting in the production of Terabases of sequence.

Diverse data — The interdisciplinary and multi-scale approaches of systems biology produce a huge diversity of data. Underpinning this is the automation at all levels of information gathering from genome to functional product. Supra-molecular information, such as that captured by high-resolution electron microscopic images, is following fast.

Chemistry and metabolism — studying the 'genome to product' cascade in isolation is not enough. Information on small molecules – whether metabolites, effectors or drugs – must also be documented.

Integration — Traditional biomolecular databases have sliced information orthogonally to biological processes, with unified access to data of a given type, but poor connectivity to precursors or products. This must be resolved to support the understanding of systems rather than just molecules.

Towards human data — Increasingly the human will move towards the centre of the life sciences and will in future become biology's best studied organism with human data production growing even more rapidly than other organism data. The power of bioinformatics for cross-organism comparisons and inferences will translate to a better understanding of human biology. Our services will need to cater for and reflect this trend.

The personal genome — Sequencing individual genomes, as well as collecting data on individuals using other biomolecular assays, is both possible and desirable. Such data will inform diagnoses and influence strategies for disease prevention and treatment. We must find ways to reap the benefits from such data while respecting constraints associated with personal consent, privacy and confidentiality.**Variation** — The ease of genome sequencing changes the kinds of problems that we tackle. The 1000 Genomes Project is only the beginning. Projects that sequence large cohorts of individuals with phenotypes of interest, e.g. cancer sufferers or strains of model organisms, will follow. The quantitative problems associated with this are formidable and they are accompanied by the equally challenging qualitative problem of conceptualising the structure and understanding the functional importance of all this variation, as discussed in Section C.2.3.

Genome to phenome — the phenotypic correlates of genetic variation are often the driving force for its study. The effects of variation are studied at all scales, from subcellular to organs, organisms and populations. The task of describing phenotypes in a standardised way is a daunting challenge.

Multidimensionality — the ease with which a range of biomolecular assays can be carried out encourages structured studies with many variables. On the one hand, multiple interrelated samples (multiple tissues from a single individual, multiple individuals with the same disease state, a single tissue sampled along a time course) can be subjected to assay under a single platform, on the other hand a given sample (or set of samples) can be subjected to multiple assays using many different platforms. Careful capture of structured metadata for such studies will be essential for their reconstruction and interpretation.

Beyond biomolecular — new holistic approaches to the biology of tissues, organs, organisms and populations require the integration of data beyond the traditional domain of molecular biology. Bridges need to be built to information from other domains, including for example, medicine and agriculture, biodiversity, and environmental science.

1.5.2 Solutions

Scalability — The scale of the bioinformatics task is outstripping the rate of advances in IT. To meet our rising demands we will have to engineer solutions at reasonable costs. Apart from sharing of the workload, pooling of effort and removal of duplication as outlined above, this will involve:

- Clever engineering to reduce the computational load, for example by storing precomputed results that contribute to the needs of multiple users, or casting the data in efficient structures that reduce computation loads.

Box D.1.1. Data security and replication

In the past our method of safeguarding our data, like that of most academic organisations, was based on tape backups. However, the EBI holds such a vast amount of data that, by our estimates, a recovery of all EBI data would cause a six month break in our services, on top of which the recovery process itself is error prone. A new approach was clearly necessary. In late 2007 we adopted a strategy involving two methods:

Weekly disk snapshots enable us to recover data that have accidentally been deleted from functional disks.

Data are replicated to an external site once a day, creating an off-site clone of the EBI data.

Data replication will be increasingly important in our future IT architecture. Cloned data can provide continuity of service in the case of disaster, with the replicated data source becoming the primary source until the problem is solved. Our current data replication centre is modest by comparison with the main data centre. Its disk storage is of a lower specification (slower), and the number of CPU cores is much smaller. This asymmetry means that only a degraded service could be offered from it, and also renders the task of synchronising the two centres more complicated and risky. Plans are underway to create an off-site data service capability based on two separate physical locations with exactly symmetrical configurations similar to those used for mission-critical systems in the commercial world.

Even the existing replication centre has already proved its worth. In two minor instances of failure, use of the replicated data has allowed speedy recovery.

- Recognising that informatics is the new bottleneck and bringing it under market control. This could mean, for example, serving the shared data and tools from a third party IT resource from which user organisations could buy the compute cycles necessary to their research. This does not imply commercialisation of the public data but rather making all the data available through a provider such as Amazon with the user paying the cost of the computation used in making the enquiry. Currently the cost and performance of services from, for example, the Amazon cloud make it unviable to us as a publicly funded resource. It is, however, conceivable that market changes will render this the preferred style of solution.
- Working with strategic partners with whom we have data exchange collaborations to explore global reductions in data volumes by removal of redundancy.
- In response to exponential data increases, we already have emerging strategies to compress data. This will involve some controlled and acceptable loss of precision and reduce storage requirements by as much as two or three orders of magnitude over raw data.

Integration of data — Integrative access to data must be provided, but the underlying architecture of the stored data must be modular. To achieve this, databases must be constructed following the same standards so that they can be exploited by relatively simple integration software. Thanks to "Service Oriented Architectures", now in widespread use, the EBI data resources will reach a new level of integration during the course of the next Indicative Scheme.

Sharing the curation task — The rate of data generation, and the diversity of the data collected, encourage a model in which responsibility for the infrastructural task of curating and making available the data is shared. This is the model envisaged for Europe by ELIXIR, and its success depends on ELIXIR. It should be stressed that any distributed solution must not jeopardise integrative access to the data, and this will pose constraints and challenges on the information engineering. Careful modularisation along natural lines of cleavage will be essential.

Sharing the service task — Because of the sheer size of the datasets held by the EBI, the previous model by which users downloaded whole databases to their local computers to carry out analyses is changing. The present tendency to run algorithms on the computers that host the data rather than on the user's own machines will rapidly create prohibitive compute costs for a single centre. This means that we must find methods to allow services from multiple

centres. Historically, replicating services at diverse locations has been complicated by differences in local computer configurations. Solutions to this are emerging based on the concept of "virtual machines", which essentially define the properties of a standard computer that physical machines at various locations then emulate to run the programs. Solutions of this type will be explored and implemented in the course of the next EMBL Programme.

1.5.3 Future themes

Below we outline in more detail our vision on a few future themes for the period of 2012-2016. The examples are chosen to be illustrative rather than exhaustive.

1.5.3.1 Genetic variation

As described in Sections C.2.2 and C.2.3 of this document, developments in DNA sequencing and genotyping technology have led to an explosion in genetic (DNA sequence) variation data for human and many other species. The EBI is the European hub for the collection, maintenance and curation of human genetic variation data. The European Genome-phenome Archive (EGA) at the EBI will collect and make available this human genetic variation information along with its phenotypic correlates. These connections will be a major focus as we seek to understand disease susceptibility, differential responses to treatment and the applicability of model organisms to human diseases. This raises a number of bioinformatics challenges:

- Adequate annotation of variations is crucial to make the archive useful to researchers. At present only a few thousand of the nearly twenty million human single nucleotide polymorphisms (SNPs) available in the most recent release of the SNP database are functionally annotated.
- A second challenge is data storage and providing access for researchers. It is in the context of populations that we are able to appreciate the difference between rare and common genetic variation. Thus, archival databases must store the raw variation data from thousands (and foreseeably millions) of individuals. Doing so will require large, well-designed databases that are constructed differently from existing resources in order to be maximally compressed but nevertheless user-accessible.
- Third, issues of consent and privacy mean that most human variation data will not be open for unlimited public access. We must respect original consent agreements from study participants. At present the EBI allows such data to be downloaded only after approval by an appropriate data access committee. This will be challenging for future data volume and usage levels. Analyses, for example, of gene interactions will access data from vast numbers of individuals, straining the one-by-one consent model. If they draw on distributed (e.g., cloud) compute systems, security constraints must also be propagated through the IT infrastructure.

1.5.3.2 Cheminformatics

Computer methods for the storage and exploitation of chemical information – now called cheminformatics – pre-date the existence of bioinformatics, but chemical data have primarily been held in proprietary systems and have used proprietary formats. This inaccessible legacy data and associated inertia have significantly delayed the impact of chemical data on biological sciences in general and healthcare in particular.

A number of recent trends have changed this, in particular the establishment of public databases, such as PubChem (2004), ZINC (2004), and the EBI's ChEBI database (2004). These have provided unrestricted access to chemical information. In 2009, the ChEMBL database of bioactivities was established at EMBL-EBI. This resource links chemical compounds to their biomolecular targets and their phenotypic effects on cells and organisms.

Open access to these data has been combined with the development of Open Source cheminformatics tools and algorithms whose development is led by the EBI. Examples are CMLSpect for encoding spectral data, and toolkits such as the Chemistry Development Kit (CDK).

EMBL-EBI's commitment to these chemical resources will continue, and we expect diverse sets of users to make use of them. Understanding the role of small molecules in biology depends on understanding the relationships and interactions between all biologically active molecules. Thus, cross-database integration, enabling examination of a chemical's association with biological targets, is a pressing priority for this Scientific Programme. This will promote progress, for example in drug development, enabling us to understand and avoid adverse effects and toxicity.

Chemical information is also crucial to our understanding of metabolites – the naturally occurring chemicals in living systems. Our long-term goal is to create comprehensive information on metabolites and chemical entities of interest to life science. The information will include:

- chemical structure
- biological properties
- physicochemical properties
- NMR data
- mass spectra
- retention coefficients of metabolites in the organism of interest
- role in various metabolic states

These are ambitious goals, and substantial effort will be dedicated to automating our methods, for example by collecting the enormous library of information already available in the literature using text mining and image recognition tools.

At the EBI, biomolecular databases such as ArrayExpress, Reactome and UniProt provide the scientific context for ChEBI and ChEMBL, and strategic collaborative partnerships, for example with PubChem at NCBI, provide the global context. This means that we are uniquely placed to set the pace in future scientific developments.

1.5.3.3 Samples and phenotypes

Today's ability to repeat biomolecular assays across different samples and conditions, or carry out multiple assays on the same sample will revolutionise biomolecular science. For instance, the same individual can be genotyped and profiled for gene expression and the results compared across individuals. We must capture this information systematically and securely. Currently it is often given low priority and buried among poorly structured metadata.

We use the term 'sample data' to refer to a range of such metadata including:

- The material sampled – organs, tissues, cell types etc.
- Phenotypic information – including disease states
- The experimental conditions – drug dosage, treatments etc.
- Where the sample was taken (important for metagenomics experiments).

The need to record data on the identity and properties of samples occurs in many databases including ArrayExpress, the EGA, PRIDE, Ensembl, and the ENA. Sample data can potentially be used to link different datasets from the same sample that are held in separate databases, and thus can be used as a tool for integrated access.

A new EBI Sample Database (ESD) will store and standardise all sample data submitted to the EBI. These data and their identifiers will be shared across databases. Prototypes already built at the EBI will be the basis for a production database during the next Scientific Programme. We expect that around a million samples will be included and we will work with strategic partners such NCBI to produce a common identifier scheme.

The ESD will be more than a book-keeping resource unifying sample information across databases. It will also be an entry point for queries from users who wish to see whether data have been collected from particular sources, specific disease states or under conditions of interest.

The ESD is well adapted to capture information in a medical setting. This is important as molecular profiling will become increasingly common in hospitals as well as in research laboratories. In medicine, assays and sample information are often coupled to phenotype data such as disease state, patient clinical histories, environmental conditions, and therapeutic treatments. We do not, however, underestimate the challenge of making them available in a manner that respects patients' consent and privacy constraints.

"Molecular matrices" representing phenotype-molecule associations, such as the connections between the biomolecular characteristics of a tissue and a disease, will be linked to all existing EBI reference molecular databases

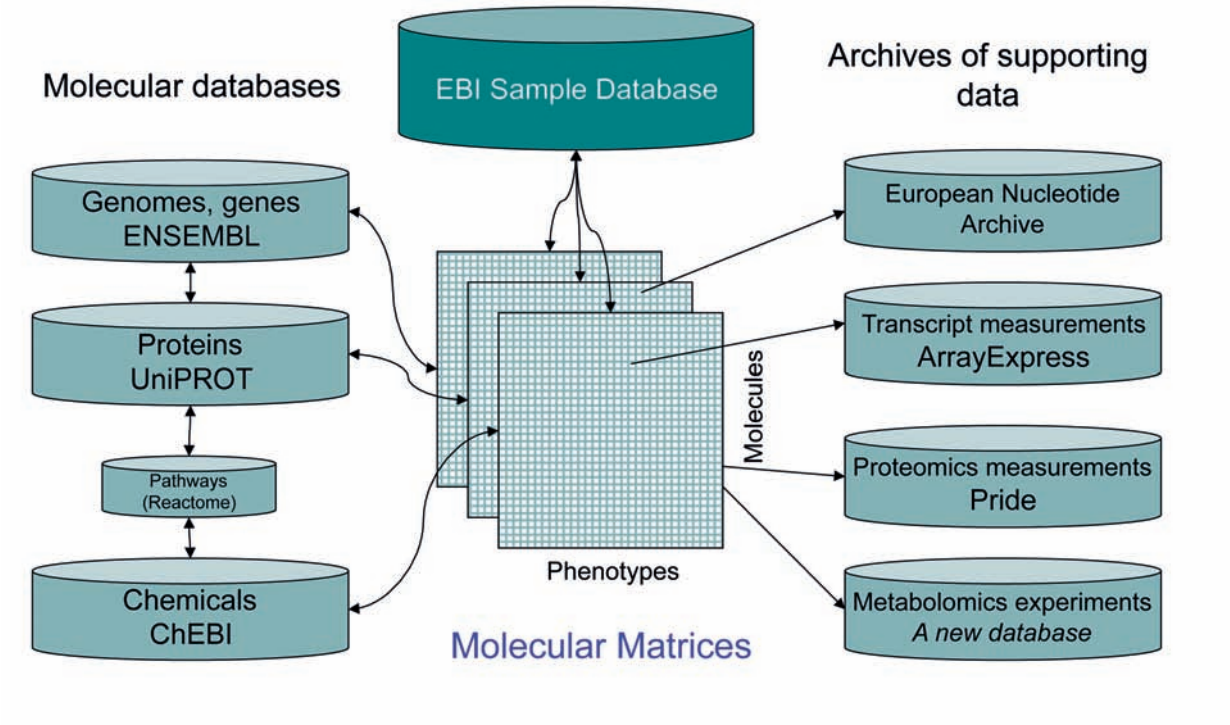


Figure D.7. EMBL-EBI Sample Database and Molecular Matrices

of genes, proteins and their behaviour. We have already developed a database that allows the user to find which genes are expressed under a particular phenotype and, for a gene of interest, to find phenotypes accompanied by changes in expression. Matrices for proteins and metabolites will be developed during the next Scientific Programme, along with an interface that allows the combination of different matrices, some possibly run externally by ELIXIR partners, into a single view. Figure D.7. shows the envisaged relationship between the sample database, matrices and the existing core biomolecular databases.

1.5.3.4 Electronic Literature

The use of the scientific literature is a key component of any scientific workflow. Recent technological and social advances in scientific publishing have made it more feasible than ever to integrate the literature with biomedical database content such as genes, proteins, gene expression patterns and macromolecular structures. In particular, the availability of full-text repositories such as PubMed Central (PMC) and open-access publishing promise greater flexibility for manipulating the knowledge found in research papers. In the long term, this will enable new approaches to extracting biomedical information from the scientific literature and integrating it with the databases. Conversely, the literature should be capable of adding greater functional context to biological databases as it becomes increasingly more challenging to navigate the volume of available information.

The EBI currently hosts a searchable database of biomedical abstracts called CiteXplore, which includes PubMed, Agricola, and patent information. This provides a good foundation for further development of the literature resources. There are five main opportunities that should be seized in the next five years:

- Improve existing resources, for example by enhancing the search function and developing citation network capability
- Expand the scope of the searches to provide full text search and retrieval capability
- Create deeper integration with other public databases

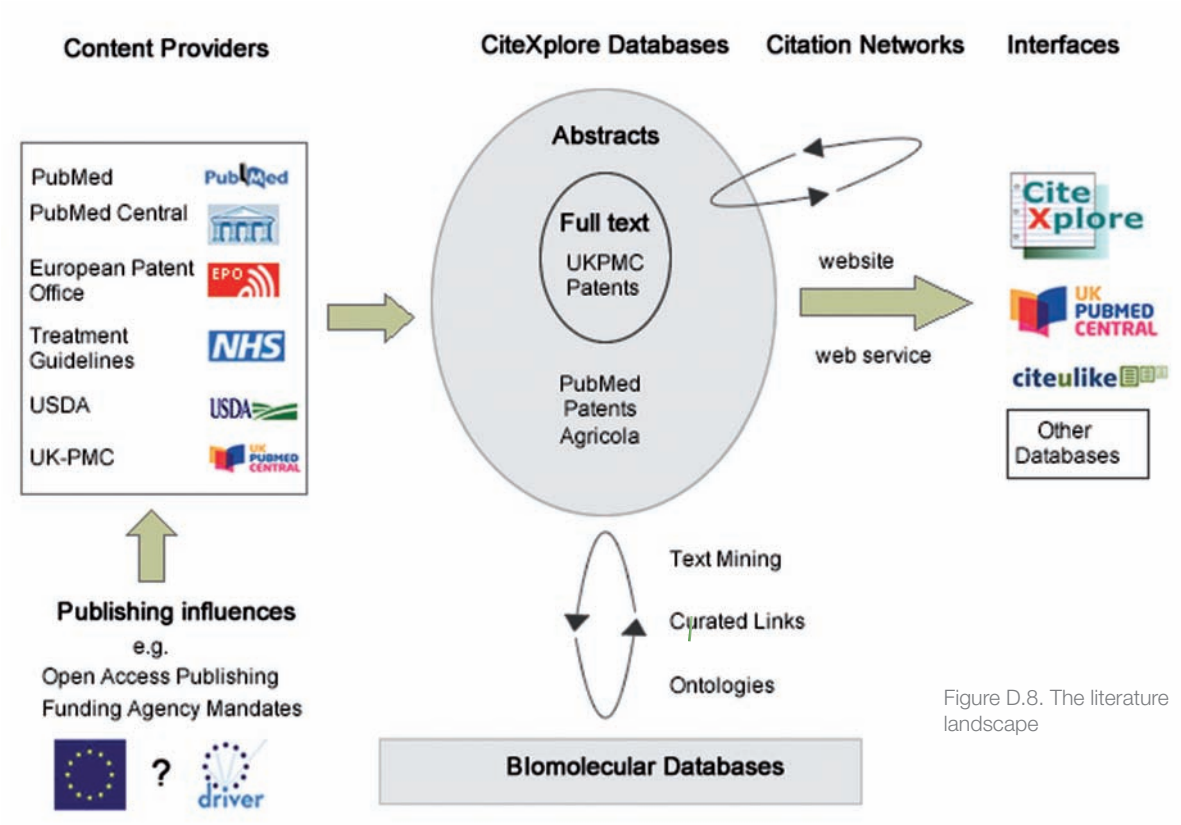


Figure D.8. The literature landscape

- Leverage text mining techniques for search and discovery
- Engage the scientific and publishing communities as partners and contributors

An important aspect to making a valuable contribution to an international public literature resource is to encourage the deposition of new content. As a partner in UKPMC (the UK component of PubMed Central), the EBI has an opportunity to participate in this effort and, ideally, to extend the scope of the database into a European setting.

In addition to the data infrastructure at the EBI, there is in-house expertise in both text mining and the engineering of ontology-driven applications. Although applying these techniques to the scientific literature clearly has a significant experimental element, they also hold promise for making novel and useful search, retrieval and data integration features.

In short, we expect the scientific literature to be explored alongside the existing databases in an integrated computer environment, not only as a stand-alone resource, but also as a key contributor to the connectivity between datasets. We expect the functionality of the literature to be enriched by seamless access to relevant databases through the automatic insertion of many more "clickable links" into electronic publications.

As discussed throughout this document, the accelerating flood of biomolecular data will be a dominant influence in the next Scientific Programme. This will be reflected both in the rate of acquisition of familiar data (sequences, structures etc.) and data generated by new methods. The resulting information engineering task must therefore expand and increase in complexity to support the changing face of the life sciences. In addition, we will face challenges of a different nature as the size, composition, and expectations of our user community evolve.

Changing methodologies at the EBI – related both to software and hardware – anticipate some of these developments, and IT advances will continue to go some way to mitigating the effects of increasing scale and complexity. However, providing the scientific community with the resources and services they need will depend on collaborations with other European service centres to share the workload and pool expertise. For example, as has become apparent in previous sections of this document, particularly Section C.2.1., high-throughput imaging will play a major role in future biological analyses of many kinds. Such imaging techniques produce vast amounts of quantitative digital data on dynamic biological processes. Based on current capacity and resources we do not anticipate that EMBL-EBI will be the European centre of gravity for services associated with imaging. Rather, we are planning to collaborate with other appropriate centres of expertise. This *modus operandi* depends heavily on the future success of the ELIXIR concept as an integrative framework for Europe's data resources.

As indicated in our projections, the EBI must expand during the next Indicative Scheme in order to fulfil its existing role, although this expansion is predicted to be less rapid than in the Programme that ends in 2011. Computer, storage and network costs are destined to rise more than staff costs in the next period, and the profile of funding envisaged in both ELIXIR and at the EBI is heavily targeted to these needs. The ELIXIR hub at the EBI is a key component of our future vision. Infrastructural support for the hub is essential to create the stability of context for both the EBI services and for the ELIXIR network. The mechanism of funding for the hub will be a crucial aspect of our discussions with the member states.

2. Structural Biology Services

2.1. EMBL's mission in the provision of scientific infrastructures in structural biology

EMBL operates three units with complementary activities in experimental structural biology in Heidelberg, Hamburg and Grenoble. The Structural and Computational Biology Unit in Heidelberg is highly integrated into the interdisciplinary research environment of the EMBL headquarters, whereas the two outstations are closely associated with world-class large-scale research infrastructures operated by the European Synchrotron Radiation Facility (ESRF) and the Institut Laue-Langevin (ILL) in Grenoble, France and the German Electron Synchrotron Research Centre (DESY) in Hamburg, Germany. The Hamburg and Grenoble Units jointly provide state-of-the-art structural biology infrastructures and technologies for the international scientific community (Figure D.9.). EMBL Heidelberg offers more limited, focussed services for sample production and imaging techniques. Transnational access to most of these activities is supported by Integrated I3 Projects from the European Commission.

Service Activity	HD	HH	GR	2010 support for transnational access
Sample Production	X		X	P-CUBE
Sample Characterisation		X a)	X	P-CUBE
Crystallisation		X	X	P-CUBE
Synchrotron Radiation Beamlines		X b)	X c)	ELISA
Light microscopy imaging	X			P-CUBE
Software Services		X		

Figure D.9. Structural Biology Services provided by EMBL

a) Under construction as part of "EMBL@PETRA3". b) DORIS-III (until 2012), PETRA-III (starting 2011). c) ESRF public MX and SAXS beamlines, plus BM14 (CRG)

2.1.1 Synchrotron radiation services

By far the most important scientific service in Hamburg and Grenoble is the provision of synchrotron radiation beamline facilities for applications in macromolecular crystallography (MX) and small angle X-ray scattering (SAXS) of biological samples in solution. Owing to the infrastructures provided by the synchrotron radiation sources, both Units have made world-leading developments in methods and instrumentation that make it possible to optimally exploit high-energy X-rays for life science research. The technology developed is available to a broad, international structural biology community through research services and via technology transfer to thousands of external users and projects. Each year several thousand users take advantage of the structural biology services offered at EMBL Grenoble and Hamburg. A substantial number of seminal projects in structural biology, including some that have recently been awarded Nobel prizes, have been made possible with the support of EMBL structural biology infrastructures (for recent highlights, see Annex D.1.).

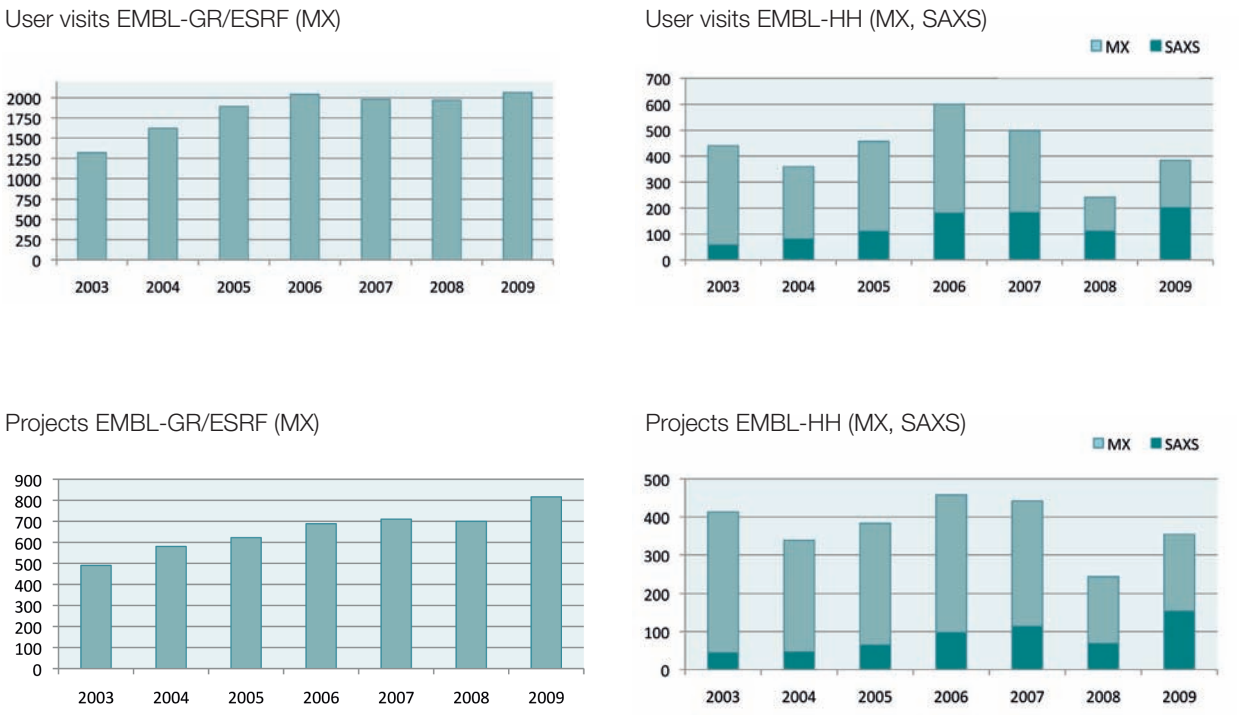


Figure D.10. User and project statistics for EMBL Grenoble and Hamburg. Note that because of the construction work for PETRA-III, only very limited services were operational at EMBL Hamburg for most of 2008.

In terms of absolute numbers, the two Units have generated an impressive record for increasingly diverse external services (Figure D.10.). In recent years (2003-2009) about 60% of all Protein Data Bank (PDB) entries from Europe have originated from X-ray data that were collected at EMBL Grenoble/ESRF and EMBL Hamburg synchrotron radiation beamlines, and in some years the numbers reached 1,000 entries (Figure D.11).

We anticipate that once the ESRF Upgrade Programme becomes effective and the EMBL@PETRA3 beamlines are in routine operation, there will be a substantial increase in usage of EMBL beamlines, arising mainly from two types of project. First, challenging projects on multi-component protein complexes, which often yield small crystals of variable quality, will require high-throughput crystal screening facilities and sophisticated software and logistics to identify the best crystals. Second, we expect that more beam-time will be used for protein–inhibitor complexes, both by academic groups and by pharmaceutical companies, to add value to medically relevant structures.

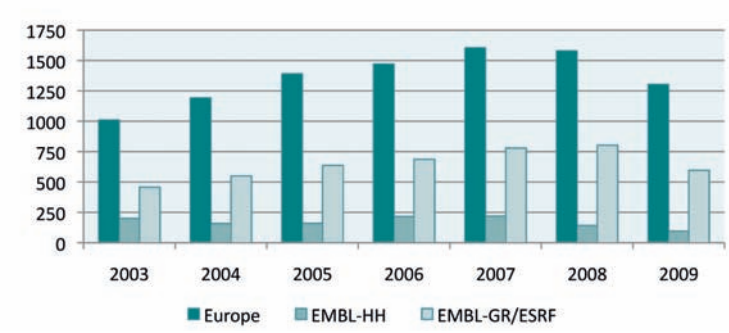


Figure D.11. Number of Protein Data Bank (PDB) entries originating in Europe, EMBL Hamburg and EMBL Grenoble/ ESRF 2003-2009.

The number of SAXS projects and publications resulting from EMBL beamlines in Hamburg has greatly increased during recent years, as the outstation has become a world-leading facility in this field (Figure D.10.). We expect that the same will be true for the recently established SAXS beamline in Grenoble, designed and built collaboratively with the experts in Hamburg, and that the additional high-throughput capacities of the future PETRA-III beamline in Hamburg will lead to a significant further boost in biological SAXS applications. The demand for the recently established automated crystallisation facilities, both in Hamburg and Grenoble, is increasing by more than 10% each year with recent numbers reaching about 1800 projects per year (Figure D.12.).

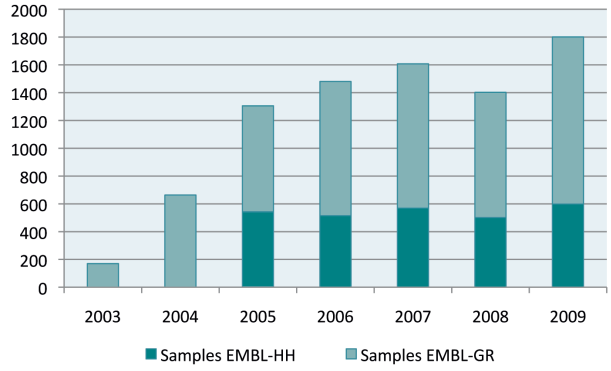


Figure D.12. Samples processed at the high-throughput crystallisation facilities at EMBL Hamburg and Grenoble.

Incorporation of these facilities into local synchrotron radiation beamlines in Hamburg and Grenoble should allow the automated assessment of diffraction properties and further boost the demand. Finally, EMBL-Hamburg has recently established several remote software services for the external user community. The two leading packages for MX (ARP/wARP) and SAXS (ATSAS) have already been used by thousands of users (Figure D.12.). We expect that the demand for these software services will further increase, scaling with future requests for access to beamline infrastructures in structural biology.

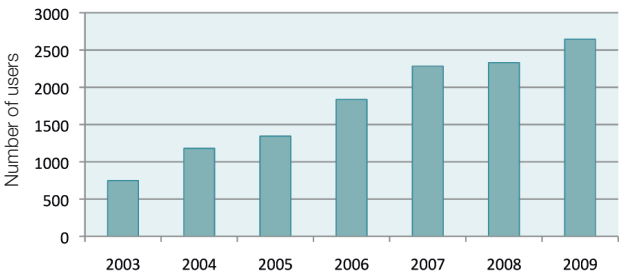


Figure D.13. Number of users of EMBL Hamburg's remote software services.

2.1.2 Integrated structural biology services

Future structure-oriented research projects will require combined applications from different structural biology techniques, complemented by imaging, cell biology and large-scale systems biology ('omics'-) methods. In the past, the three structural biology Units at EMBL have pioneered the concept of integrated structural biology and are therefore excellently positioned for future developments. To address the associated challenges, both EMBL Hamburg and Grenoble have started to extend into upstream and downstream aspects of structure determination, including protein production techniques, high-throughput crystallisation and automated data interpretation. Many of the technological achievements have been shared and disseminated through participation in European initiatives, including SPINE, BIOXHIT, P-CUBE and ELISA, that are aimed both at providing structural biologists with access to existing techniques and the development of new technologies. In future, it is likely that many of these activities will be carried out in the context of the structural biology ESFRI project, INSTRUCT (see Section H.1.3.1). Stimulated by various intramural integration activities and joint participation in several large-scale European Commission projects, the three Units have achieved a high level of collaboration across the entire range of their service activities. In addition, regular bilateral meetings between scientists and engineers from the two EMBL sites in Grenoble and Hamburg make sure that synergies are leveraged and collaborative projects promoted.

2.2 Unit-specific services

2.2.1 EMBL Grenoble Services

In Grenoble, EMBL scientists are continuing to make important contributions in the design, construction and deployment of equipment to create state-of-the-art, automated and remotely accessible experimental stations at a total of eight ESRF synchrotron beamlines, seven for MX and one for SAXS. These beamlines are operated by the EMBL-ESRF Joint Structural Biology Group and four of them are under the direct responsibility of EMBL scientists. From 2010-2014, the Collaborative Research Group (CRG) beamline, BM14, previously run jointly by the UK Medical Research Council and EMBL, will be operated by a consortium comprising EMBL, ESRF and research organisations from India, funded by the Indian Ministry of Science and Technology. In addition to facilities for X-ray crystallography, the powerful neutron source provided by ILL in Grenoble allows neutron diffraction experiments on crystals of biological macromolecules.

For EMBL Grenoble's services we envisage the following additional objectives for the period 2012-2016:

- **Towards nano-MX.** The optimal exploitation of very small crystals is a major goal in crystallography. During the past Indicative Scheme, EMBL contributed significantly to the construction and operation of a new dedicated micro-focus beamline, ID23-2, at the ESRF synchrotron. At less than 10 micrometres in diameter, the ID23-2 beam is close to ten times smaller than conventional MX beamlines and allows the analysis of very small crystals, down to a mere 5 microns. A new, highly precise, vertically mounted goniometer spindle under development at the Unit aims to lower the threshold below 1 micron.
- **Experiment automation**. Scientists and engineers at EMBL Grenoble are internationally recognised for their innovative contributions to the automation of the sample environment of MX beamlines. During the past few years, a standard sample holder, now universally used in Europe, an automatic frozen crystal sample changer, crystal centring and reorienting software (used in conjunction with a mini-kappa goniometer) and a new on-line humidifier system, have been developed, which are all compatible with the pioneering micro-diffractometer. These technologies are installed on all beamlines at the ESRF and many have also been licensed to manufacturers and sold to synchrotron providers all over the world.
- **SAXS**. EMBL Grenoble has played a key role in building and operating the first dedicated SAXS beamline in Grenoble, in collaboration with the ESRF and with the help of the substantial SAXS expertise at EMBL Hamburg. ID14-3 is part of a new platform that began to offer users joint access to SAXS and SANS (small-angle neutron scattering provided in collaboration with ILL) at the end of 2008. It is now equipped with an innovative, small volume automatic sample changer, which was designed and built in a trilateral collaboration between EMBL Hamburg, EMBL Grenoble and the ESRF, as well as a pixel detector.

- **Sample preparation**. Some of the most crucial bottlenecks of modern structural biology occur in the preparatory steps before a sample even enters a beamline. Many proteins, especially multi-component complexes, are difficult to express, let alone crystallise. In Grenoble, the Partnership for Structural Biology (PSB, see Section G.2) offers a broad range of complementary services, including robotics for high-throughput protein expression and crystallisation, facilities for isotope labelling for neutron scattering and nuclear magnetic resonance (NMR), and instrumentation for NMR, mass spectrometry and cryo-electron microscopy. In the current Programme period, EMBL scientists have made important innovations in prokaryotic (screening for soluble domains in proteins that are difficult to express) and eukaryotic systems (automatic assembly of multi-protein expression vectors for baculovirus expression) for the production of recombinant proteins and complexes. These are made available to external users through the P-CUBE programme.
- **ESRF Upgrade Programme**. The ESRF infrastructure will undergo a substantial upgrade in the near future, which will ultimately allow world-class structural biology research at a higher throughput than currently possible. A 'structural biology village' will be built to house future MX and SAXS beamlines. The core of the new system will be built at beamline ID30 and complemented by the existing beamlines ID29, ID23-1 and ID23-2. These facilities will be highly automated and will provide a complete pipeline for crystal screening, sorting and data acquisition. The aim is to screen more than 400 000 crystals and collect X-ray data from up to 50 000 crystals per year. The recently inaugurated BioSAXS beamline will also be moved into the structural biology village. EMBL Grenoble will be a key player in the establishment of these future facilities through continued design and development of new hardware for MX and SAXS and through direct contributions of EMBL beamline scientists to the design and operation of the beams. EMBL Grenoble also aims to build significantly on the new methods it has introduced for producing high-quality samples of proteins and supra-molecular complexes. In combination, these plans will allow us to offer enhanced and better services to structural biologists in future.

2.2.2 EMBL Hamburg Services

EMBL Hamburg has a long record in the independent operation of seven synchrotron radiation beamlines using the central synchrotron ring infrastructures at DESY. The activities have been focussed on service provision to structural biologists interested in carrying out applications in MX, SAXS and, until 2007, X-ray absorption spectroscopy. Although the DORIS-III beamlines are less brilliant than those of third generation synchrotron sources such as ESRF, the quality of service support provided by EMBL Hamburg nevertheless means that the number of projects has remained stable during recent years at around 500 per annum. However, there has been a significant increase in biological SAXS applications (Figure D.10.), in which Hamburg has gained world leadership. In 2007, a major transition began with the construction of three new beamlines with applications in structural biology at the new PETRA-III storage ring. Because of the need to focus resources on this project, the number of accessible DORIS-III beamlines was reduced to four. Three MX beamlines (X11, X12, X13) and the SAXS beamline (X33) will be kept in use during the remaining operation of the DORIS-III ring (until the end of 2012).

EMBL@PETRA3. The new PETRA-III storage ring is one of the world's strongest synchrotron rings, with unique optical parameters that provide photons with unprecedented flux and collimation. EMBL Hamburg has designed and is currently constructing three state-of-the-art beamlines for biological MX and SAXS applications. They will form the core of the "EMBL@PETRA3" Integrated Facility for Structural Biology, which is expected to start user operation in 2011. One of the central tasks during the next Indicative Scheme will be to further develop and to optimize these beamlines for the envisaged experimental challenges (see below). The developments will require additional investments into state-of-the-art equipment and tools to design and create an optimized experiment requirement, which fully exploits the specific beam properties of PETRA-III. The two MX beamlines will be geared towards micro-crystallography, ultimately aiming for the provision of an experimental environment that allows routine X-ray data acquisition of crystals of less than 5 micrometres. The beamlines will also be tuneable across a broad energy range and highly automated to allow exploration of new methods for experimental phase determination in MX. The new BioSAXS beamline will employ microfluidics and high-level experiment automation for large-scale screening of highly dilute biological samples in small volumes. The high brilliance will be used for time-resolved SAXS studies in solution approaching the millisecond range, thus allowing biochemical processes to be monitored *in situ*.

EMBL@PETRA3 will also house infrastructure for high-throughput protein crystallisation, sample preparation and characterisation as well as data processing and evaluation, thus providing a complete set of tools for structural biology experiments using synchrotron radiation. The new facilities at PETRA-III will therefore render X-ray crystallography studies more efficient and help speed up the investigation of molecules relevant to life and disease. They will also allow the investigation of protein machines of unprecedented size and complexity. Some specific research plans have been outlined elsewhere (see Section C.2.1). EMBL@PETRA3 and its services will be accessible to structural biologists from the member states and, to some extent, from the rest of the world. User time at all future EMBL@PETRA3 facilities will be allocated on the basis of scientific merit as assessed by external panels appointed by EMBL, a system that functioned very well for experiments at the DORIS-III beamlines.

Further developments are:

- **Automation of the BioSAXS experiment**. Various methods and software developments and the leading expertise in BioSAXS experiments at EMBL Hamburg have tremendously increased the demand for biological SAXS services (Figure D.10.). To meet these requests EMBL Hamburg has completely refurbished the optics of the SAXS X33 beamline and equipped it with an in-vacuum measurement cell, a 1M Pilatus pixel detector, and an automated sample changer. A beamline meta-server allows the integration of the data acquisition system with an automated data analysis pipeline. These developments have dramatically improved data quality, speed, user friendliness and efficiency of the beamline, and allow for remote access to the infrastructure.
- **Remote software services**. EMBL Hamburg has become a world-leader in providing software that allows automated interpretation of data collected on X-ray beamlines, in particular through the widely used software packages ARP/wARP and ATSAS for MX and SAXS data analysis, respectively. Further packages that have been made available to the community include BEST (advanced data collection strategy), XREC (automated crystal centring), Auto-Rickshaw (automated structure determination), HKL2MAP (crystallographic phasing) and RAPIDO (structure superposition, domain identification). In addition, EMBL Hamburg has started to offer some of these services by remote software provision. During the past five years the use of these services has tripled and exceeded 2500 users from almost 1500 independent domains in 2009 (Figure D.13.).
- **FEL-based structural biology**. DESY is building a new generation of Free Electron Laser (FEL) facilities. The laser test facility FLASH has been in operation since 2005. FLASH generates laser light in the soft X-ray range (around 10 nm) and is thus well suited to biological specimens at a resolution of a few hundred Ångstroms. EMBL Hamburg, in collaboration with DESY, is participating in a number of preliminary experiments that indicate great potential for using coherent laser radiation for biological samples. FLASH will be followed by the European X-FEL facility, which is expected to start operation in 2014. The light emitted by the FEL facilities has unprecedented properties in terms of energy, collimation, time structure and coherence, and therefore could allow biological experiments that have been impossible to date. These could include X-ray diffractive imaging of single cells or cellular systems in the nanometre-range resolution, depicting time structures and associated structural dynamics of biological processes.

2.3. Integration of services for future demands in structural biology of complex biological systems

To meet the dynamic demands imposed by system-wide and cell biology-driven projects, the services in structural biology need to extend beyond their traditional boundaries and link up with complementary disciplines including imaging, genomics and computational biology. EMBL Grenoble and Hamburg are aiming, in the course of the upcoming Indicative Scheme period, to further develop their services in the direction of multi-component complexes and even cell-based interactomes using expertise present at the other EMBL sites. One of the key challenges in the coming years will be to develop approaches that allow the mapping of complex high-resolution structural information into quantitative three-dimensional network models of complete biological systems. These efforts will require pushing the limits of synchrotron instrumentation and sample handling technologies and, on a wider scale, the integration of the methods currently offered in Hamburg and Grenoble (X-ray crystallography and SAXS) with rapidly evolving complementary methods, including cryo-electron microscopy (EM), cryo-tomography and advanced light microscopy, which are mostly located at EMBL Heidelberg. Applied in conjunction with SAXS, MX

and modelling approaches, these lower resolution techniques will allow structural biology-oriented research within complete cellular systems. In this respect a major trend will be to provide integrated access to multiple platforms, for example protein production/nano-volume crystallisation and *in situ* synchrotron crystal screening (Grenoble and Hamburg), production of fluorescently tagged proteins coupled to advanced light microscopy (Heidelberg) and SAXS, EM and crystallography (Grenoble and Hamburg).

Both synchrotron-associated EMBL Units are developing highly integrated facilities: EMBL@PETRA3 and the Centre for Structural Systems Biology (CSSB) in Hamburg, and the Partnership for Structural Biology (PSB) in Grenoble. The PSB was created in 2002 and includes the EMBL Grenoble Unit, the ESRF Structural Biology group, the ILL Neutron/Biology group, the UJF-EMBL-CNRS Unit of Virus Host Cell Interactions (UVHCI) and the Institut de Biologie Structurale. Collectively, the PSB already offers a broad variety of technical platforms for structural biology services. The aim of the EMBL Hamburg Unit is to provide a fully integrated service package at the future EMBL@PETRA3 facility. It will include facilities for state-of-the-art sample characterisation, MX and SAXS beamlines at PETRA-III, plus facilities for automated data evaluation, with on-site and remote options, all within one building. These services will be complemented by the future CSSB, which will be operated by a consortium of leading research organisations from northern Germany. The central aim of the future CSSB is to match Hamburg's internationally leading synchrotron and future laser infrastructures by an equally ambitious research programme, connecting structural biology and systems biology approaches with key projects of biomedical relevance.

The planned future infrastructures and services outlined above will not be possible without further expanding the expertise of the closely collaborating instrumentation and service groups at EMBL Hamburg and Grenoble. Envisaged key projects during the next Indicative Scheme are:

1) **Integration of sample quality assessment, data acquisition and data interpretation**. The aim is to develop an integrated platform for crystal screening, sorting and X-ray data collection, on the basis of a new crystallisation support and associated robotics system, to automatically harvest crystals on supports that are compatible with data-collection at cryogenic temperature. To meet these demands, there will be a requirement for a future generation of sample changers both for biological MX and SAXS applications, using industrial standards to maximise flexible applications in future integrated experiment pipelines. The EMBL Units in Hamburg and Grenoble will jointly take advantage of their recent achievements in synchrotron instrumentation and automated sample handling (see above).

2) **Increase experiment efficiency by optimisation of beamline optics and detectors.** Cutting-edge X-ray optics and detectors will be deployed for optimal use of highly brilliant synchrotron beams for biological research. It will be essential to equip all future beamline facilities with state-of-the-art detectors with the most advanced technology, which is currently the hybrid pixel detector technology. This technology features very short read-out and dead times, high sensitivity and spatial resolution, and an unprecedented dynamic range, allowing shutter-less X-ray data collection.

3) **Low-resolution applications in MX**. Novel experimental and software tools will be developed to determine experimental crystallographic phases of large systems at low resolution (i.e. beyond ~ 4 Å, the general limit of present applications). This will require experimental stations (ESRF, PETRA-III) to collect the highest quality X-ray data from poorly diffracting, small and inhomogeneous crystals. Phasing software will need to be established to maximally exploit the raw data and procedures created for identifying and placing structural fragments. These will need to be converted into non-expert protocols that will allow broad applications by the external scientific community.

4) **Dynamics of structures.** Although high-resolution crystal structures are unmatched in terms of overall precision of structural information at the level of single atoms, they are static snapshots and combining them to understand dynamic processes remains a challenging task. With the available infrastructures in the three Units in Heidelberg, Grenoble and Hamburg, EMBL is in an excellent position to complement atomic structures with additional data, which should provide insight into their associated dynamics. These specifically include future facilities for time-resolved SAXS (in Hamburg and Grenoble), NMR spectroscopy (Heidelberg) and electron tomography and real-time live imaging approaches (Heidelberg). A key future task will be the development and implementation of protocols for the combined use of these methods, to gain complementary information on the dynamics and larger-scale organisation associated with the three-dimensional structures of complex macromolecular networks.

5) **Future FEL-based structural biology**. Applications of the new X-ray Free Electron Laser facilities in Hamburg (FLASH, X-FEL) will be explored. Intense coherent X-rays allow the measurement of a diffraction image from a single nanoscale particle. The unprecedented time structure and intensity of the X-ray pulses, as well as their coherence properties, will open entirely new frontiers for science and technology. Given the expected properties of the X-FEL, diffractive imaging to 10-20Å resolution can be anticipated. To achieve this goal we aim to develop optimised methods for phase retrieval and three-dimensional reconstruction of objects from such diffraction data and for interpretation of the structural content. Collectively, such developments may permit the study of large biological macromolecules *in vivo* and will provide vital experience in the design of future experiments on the DESY X-FEL source.

Core Facilities and IT infrastructure and services

3.1 Core Facilities

As previously discussed, EMBL's policy is to hire young group and team leaders directly after their postdoctoral period and to provide them with their first independent position, for a maximum period of nine years. In order to maximise the potential of these young principle investigators (PIs), EMBL supports them in numerous ways by providing a highly collaborative environment and support that combine to allow all our groups to operate at the cutting edge. Achieving ambitious goals at the forefront of today's life sciences often requires focusing diverse sets of expertise and multiple expensive technologies on a specific biological problem and EMBL's Core Facilities play a critical role in enabling this to occur. Organising EMBL's equipment into shared facilities is also an extremely cost-effective way for the Laboratory to invest its limited resources.

EMBL's Core Facilities are designed to support internal and external research projects, to interact with industry, often as technology developers or Beta-testers, to obtain optimum early access to new technologies, and to enable research scientists to operate at and beyond normal state of the art. They are staffed by technology experts who entirely focus on service provision delivering technologies to be used in research projects designed and run by others. By creating Core Facilities that are open to all EMBL staff we not only avoid unnecessary duplication of expensive equipment but also ensure that the equipment is always supervised and maintained under expert guidance. The Core Facilities are also very attractive collaborative partners for companies who realise that the large number of visiting users and training courses organised at EMBL means that their products, if placed in the Core Facilities, have very good exposure to the researchers who form their target group.

The Core Facilities greatly improve the efficiency of research groups at EMBL. To ensure the support activities are tailored to the demands of the research community all Core Facilities have a user committee consisting of representatives of EMBL's research Units. The committee advises the facilities on new services and technologies needed, defines future strategies and also provides invaluable feedback concerning current operations. Like our research Units, the Core Facilities are also subject to stringent, regular review by EMBL's Scientific Advisory Committee. Core facility reviews took place in 2006 and 2010. In the 2010 review the overall performance of the Core Facility was perceived as excellent and the technologies and services offered were described as 'of the highest quality'. Further, regular internal user surveys of the services provided have revealed a high degree of satisfaction with the technology, services and support provided by the Core Facilities as summarized in Figure D.14. In part, we use the results of these surveys to determine the needs of the EMBL research community. Decisions as to what to provide internally and what to outsource are discussed regularly with respect to each of the Core Facilities. In a nutshell, internal services have to be either technologically or financially superior to those offered by external providers in order for EMBL to invest in them.

The main function of the Core Facilities is to meet a need that is internal to EMBL. Where capacity makes this possible, we however also open our Core Facilities to external users from the member states, with particular emphasis on EMBL Partnerships and the recipients of EMBO Young Investigator awards (YIPs). In addition, EMBL is always prepared to provide advice and practical help to external users who wish to see how our facilities work, obtain advice on the best equipment for their specific requirements, or to duplicate aspects of one or more of our core facilities in their home institute.

The Core Facilities play an important role in training. They organise both internal and external courses and workshops, often in collaboration with industrial partners, spend a significant amount of their time training individual users, and provide many institutes in the member states with advice on setting up and maintaining their own core facilities and services.

The EMBL Core Facilities provide the technological backdrop for many of the ambitious research projects outlined elsewhere in this document and have to evolve in parallel with our research activities. To illustrate their usefulness to EMBL's research, we provide an Appendix (Appendix D.2.) where some of the projects enabled by the Core Facilities in the current Programme period are listed. As the services offered develop in response to changing demands and requirements, detailed changes in the services to be provided in the course of the next Programme can only be partly foreseen at this point. In the following we try to outline broad trends and developments that

User satisfaction index for the core Facilities and IT services

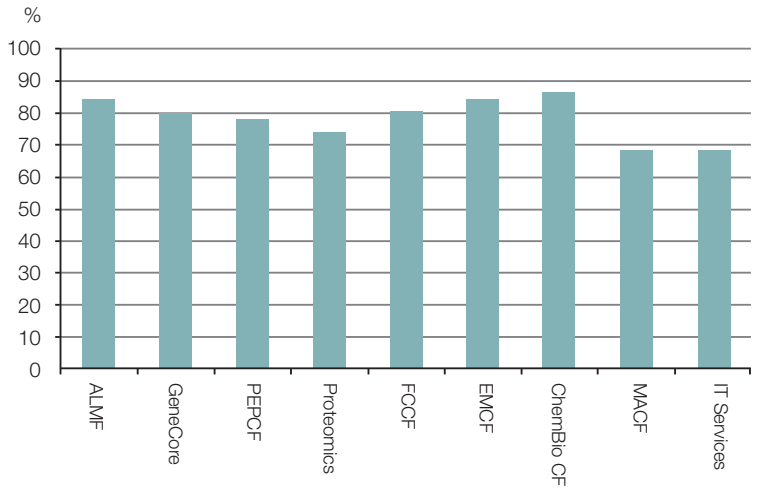


Figure D.14: Results of the 2009 internal user survey addressing the quality of services provided and general user satisfaction with the Core Facilities and central IT Services in Heidelberg.

we see emerging within individual core facilities. These should be considered in the context of the research plans described in Section C.2.

3.1.1 Genomics Core Facility

The Genomics Core Facility offers technologies and specialist support for a variety of genomics techniques, including DNA sequencing, quantitative PCR and microarrays. It also provides services in the area of automation by supporting researchers in miniaturizing and automating experiments and adapting them for robotic handling. In relation to technology development, a recent addition is locked nucleic acid-based oligonucleotide arrays (miChip) that allow the fast and accurate profiling of mature microRNAs from human, mouse and other organisms. miChip was developed jointly by EMBL research groups and the Genomics Core Facility and is now heavily used by researchers at EMBL and its Molecular Medicine Partnership Unit in Heidelberg.

The most important development in recent years, which has already been broadly discussed in other parts of this document, was the introduction of next-generation DNA sequencing (NGS). The use of this technology, first available at EMBL in 2007, has blossomed and continues to expand rapidly in areas such as genome sequencing, chromatin immunoprecipitation analysis (ChIP-seq) and gene expression analysis (RNA-seq). NGS has resulted in a complete restructuring of the Core Facility, which now operates four NGS sequencers plus the related computational hardware and provides initial data analysis support. To free the resources required for preparation of sequencing libraries and operation of next-generation instruments single-pass DNA sequencing has been outsourced to a commercial provider.

Services and support in NGS technology will continue to be the major activity of the Genomics Core Facility over the course of the next Programme. In order to handle and analyse the volume and complexity of the resulting data we will further expand the bioinformatics capacity of the Core Facility by recruiting specialist staff and a powerful software suite for mapping and analysis of NGS data obtained from various sample types (for example, ChIP-Seq, RNA-Seq, Methyl-Seq and ncRNA-Seq) as well as by upgrading the computational hardware. These developments serve as a preparation for many of the research projects outlined in C.2.2 and C.2.3.

The Genomics Core Facility will also support microarray-based methods of gene expression profiling and its use in the verification of NGS data while demand for this from EMBL researchers remains high. Training is an integral part of the Genomics Core Facility. Aside from project-based tutoring of individual researchers the core facility also organises practical courses for the external community on quantitative real-time RT-PCR (qPCR), gene or miRNA expression profiling and next-generation sequencing and data analysis.

3.1.2. Proteomics Core Facility

The Proteomics Core Facility offers a wide range of services for the characterization of peptides and proteins mainly using a number of mass spectrometers. Since its inception in 2001 the facility has been supported by BioRad, who provide instrumentation for gel-based protein analysis in exchange for access. The services offered by the facility fall into two categories, both in terms of application and the number of samples that are processed annually:

- Analysis of intact proteins, serving structural biologists throughout EMBL
- Proteomic applications for identification of proteins and protein mixtures

To enhance the capabilities of the facility and to address the more challenging questions of modern biology, a number of changes have been made since 2009. To embed the facility better in EMBL's research activities it has been relocated adjacent to the proteomics research group in the Genome Biology Unit and it is now led by the Team leader who is also responsible for mass spectrometry research and development. This has improved exchange between the research and service activities. With the new recruitment, a high-end mass spectrometer was acquired for the analysis of complex protein mixtures. These improvements have enabled the facility to address the urgent and growing need for the identification of proteins that can only be acquired in small quantities or in samples that contain many proteins in very different relative amounts.

Support for many of the research projects planned in the new research Programme will require additional investment. In the next 5 years, the following developments are foreseen:

- With the available infrastructure, the facility needs to support the routine identification of phosphorylation sites. This will involve the use of dedicated enrichment techniques based on phospho-affinity methods.
- A major issue is the implementation of quantitative analysis by stable isotope labeling. The facility needs to become proficient in the techniques required for the incorporation of stable isotopes into proteins and the bioinformatic tools required for data analysis.
- Many EMBL structural biologists require detailed characterization of proteins subjected to limited proteolysis. Meeting this need requires that chemical modification of N- and C-termini followed by directed identification by mass spectrometry is offered as a routine service, and it is our intention to do so.
- The Waters Q-tof mass spectrometer, used to analyse intact proteins, is 8 years old and will need to be replaced. Depending on the evolution of technology, we will also consider technologies for the analysis of non-covalent protein interactions. These methods are in demand but are not yet at the stage where they can be provided routinely by a service facility.

3.1.3 Protein Expression and Purification Core Facility (PEPCF)

The expression and purification of proteins in recombinant form is frequently the rate-limiting step in projects that require biochemical or structural analysis. Since protein expression is ubiquitously used in molecular biology, the PEPCF's user community encompasses most EMBL wet-lab groups. Unfortunately, although experience in this area helps, each protein essentially presents a new challenge. Further developments and improvements in the tools and techniques used for protein expression and purification are constantly evaluated and, if they are found to be beneficial for the users, implemented by the PEPCF.

The PEPCF provides technology, support and advice to its users at each step of the protein expression process. In addition to its EMBL internal service function, the PEPCF, together with the Advanced Light Microscopy Facility (ALMF), participates in the EU-funded initiative P-CUBE to provide trans-national access that allows external visitors to perform protein labeling and advanced light microscopy experiments at EMBL.

PEPCF offers a large collection of vectors and host strains for expression of proteins in bacterial and eukaryotic systems, and provides advice on which techniques to use under which conditions depending on the properties of the specific protein. This service is extremely important, because finding conditions to express a protein at reasonable yield in soluble form is often extremely time-consuming. At a small scale, the PEPCF has automated assays to explore a limited set of standard expression conditions and it can provide further hands-on help in difficult cases.

The facility also produces commonly requested proteins, like enzymes in common use, thereby helping researchers to save time and costs.

Finally, PEPCF is responsible for the maintenance of the equipment required for a diverse set of biophysical analysis techniques, and for training and advising EMBL scientists on their use. In future, we are planning to automate the screening for optimal expression constructs and conditions in collaboration with the Genomics Core Facility. This has become possible thanks to the availability of new column-based small-scale purification tools.

3.1.4 Monoclonal Antibody Core Facility (MACF)

Antibodies are in widespread use in all aspects of experimental molecular biology and, unlike polyclonal antisera, monoclonal antibodies have the advantages of being essentially infinitely renewable and of stable quality. The MACF produces monoclonal antibodies (mAbs) for internal use by EMBL research groups and also offers its services to external users (currently the internal:external ratio is close to 1:1). Our mAbs are used to detect and quantify the abundance of a given protein in a sample or a cell, to localize the protein through light or electron microscopy, to purify proteins together with interaction partners and to determine the conformational or modification state of a protein. In addition mAbs have become the tool of choice for investigating the interaction of proteins with DNA in the context of chromatin via ChIP and ChIPSeq.

We have broadened the services provided by the MACF over the course of the current Programme. The access to scale-up and purification of mAbs has been increased and a greater proportion of projects have been taken to the point where the hybridoma lines are cloned to purity. Both services are now part of the facility's standard repertoire. Immunosuppression has also been successfully used in projects, in order to improve the specificity of the mAbs obtained, and can now be routinely included in the service on request.

The future of the MACF lies in further improving the primary antibody screening, so as to gain more information about the antibody's properties at an early stage. One improvement suggested by users that we will implement is to perform limited epitope mapping using Western blots of antigen that has undergone limited proteolysis, so-called "Cleveland" maps. This identifies the region of the protein recognized by the mAb and is frequently very useful in interpretation of the data obtained using the mAb. In addition, we will investigate the use of pooled sets of monoclonals derived from each fusion experiment. In some cases this can provide a superior reagent with increased signal/noise. On a practical note, it will be essential in the next Programme that the ten year-old facility robotics are updated and renewed.

3.1.5 Chemical Biology Core Facility (CBCF)

Established in 2004, the CBCF is one of the newer additions to EMBL's Core Facilities but it has quickly turned into an important node in EMBL's research and innovation network. The facility develops and carries out screens to identify small molecules that modulate biological pathways and enzymatic activities and can serve as biotools to help dissect cellular processes. If compounds disrupt processes involved in disease they can also serve as a starting point for drug development. EMBL groups wish to identify small-molecule inhibitors for both these reasons.

The CBCF provides the infrastructure needed for chemical screening. It provides state-of-the-art technology, an extensive diversity-oriented compound library (that must however be renewed for the upcoming Programme) and the expertise of six highly trained staff to guide researchers through the challenges of identifying and characterising small molecules. The facility also develops bench-top assays to formats more appropriate for medium to high-thoughput screening for both biochemical and cell based assay systems.

The CBCF, originally a joint venture between the DKFZ and EMBL, was expanded in 2007 with the University of Heidelberg becoming a full member. The capacity for projects remains at 10-12 screens per year. In 2008 the compound library was renewed with the purchase of a high quality collection of 80,000 lead-like molecules. One of the main limitations of the facility remains the lack of access to chemical optimisation of active compounds. This is used to improve the efficiency or the chemical properties, such as solubility, stability, or penetrance into cells, of an active "hit" compound. However, the rationale used to construct the screening library allows commercially available structural analogues to be purchased for most projects. In favourable cases, these can provide a "second best" alternative to building up a structure activity relationship for the compound under study.

When the facility started most projects involved screens for inhibitors of protein activity. Now, over 30% of the screens run in the facility instead address protein-protein interactions and are searches for compounds that disrupt these interactions. If this trend continues, the properties of the small molecules in the screening libraries will have to be altered to give a better chance of success against this challenging target class. Library selection will involve collaboration with chemistry groups and through the use of *in silico* approaches.

The most recent addition to the service portfolio of the CBCF is computational chemistry expertise and infrastructure, which offers a complementary approach for identifying active small molecules. This will be opened to research groups as a full service in 2010. A pilot project using this approach has been very successful.

Various productive collaborations with EMBL Research Units and other Core Facilities have led to a number of publications and patent applications involving the CBCF. In addition, two successful spinout companies have been founded to further optimise and commercialise small molecules identified in the CBCF (Elara Pharmaceuticals GmbH, Heidelberg and Savira Pharmaceuticals GmbH, Vienna).

3.1.6 Advanced Light Microscopy Facility (ALMF)

The increased use of imaging in biology has been one of the major themes running through this document. The ALMF provides researchers at EMBL and in the member states with state-of-the-art light microscopy and image analysis technology as well as expertise in their application to biological problems. The equipment provided represents the outcome of ongoing collaboration between industrial partners, EMBL scientists and the ALMF. In this way, the latest technologies are made available to EMBL scientists and visitors as early as possible after their development.

In the past five years the ALMF capacity has increased to 19 advanced light microscope systems and five workstations for both state of the art and in-house custom-developed image analysis. They represent collaborations with 12 leading companies in light microscopy technology development. Annual use of the microscopes has exceeded 30,000 hours by over 180 users. About 30% of this time is used by external users who visit EMBL in small groups to carry out their own experiments in the ALMF, evaluate microscopy equipment or to obtain advice on setting up comparable facilities in their home institutes.

The organisation and running of in-house training courses on the basics of advanced light microscopy and image analysis has become an important task for the ALMF. These courses are organised jointly with the EMBL Centre for Molecular and Cellular Imaging and are aimed at PhD and postdoctoral fellows at early stages of their stay at EMBL. The aim is to improve their expertise in both light microscopy and computational image analysis and thus allow them to make the best use of the ALMF equipment.

The ALMF service has been extended to include high-content screening microscopy. Currently, 9 state-of-the-art high-content screening microscope systems are available allowing large-scale projects such as genome wide microscopy-based siRNA screens. Presently ten large-scale projects are ongoing involving EMBL researchers and seven more involving external users.

Apart from high-content screening microscopy, we have implemented a range of other new technologies over the past five years including: diffraction limited laser nano-surgery for the ablation of organelles and structures in cells or tissues, time domain fluorescence lifetime imaging and fluorescence correlation spectroscopy, which allow the quantitative imaging of molecular interactions in living cells, TIRF-based single molecule detection microscopy and ultrafast laser scanning confocal microscopy. In collaboration with the Electron Microscopy facility protocols for correlative electron and light microscopy were also developed. Many of these developments form the basis for EMBL's future research plans, that rely even more heavily than in the past on imaging technologies (see Section C.2.1)

Future plans:

- We are planning to extend high-content screening microscopy from current 3D time-lapse applications to include applications involving fluorescence recovery after photobleaching, fluorescence resonance energy transfer or correlation spectroscopy. Increasing the throughput by automation of the technology will not only allow large-scale analyses in systems biology applications but also facilitate these complex technologies considerably.

- As soon as they become available, the ALMF and EMBL researchers will investigate and test new super-resolution microscopy technologies that extend light microscopy imaging into the nanometre range with the aim to offer these enabling techniques to ALMF users in the future.
- Efficient and long-term imaging of large objects, such as developing embryos or thick tissue sections, have been made possible by single plane illumination microscopy (SPIM), developed at EMBL in recent years. Commercial versions of this technology, together with powerful data handling storage and analysis infrastructure, will be one of the key implementations for the ALMF in future.
- Data handling and storage, in particular for high-throughput microscopy applications, represents a very important aspect of our future requirements. This is planned together with both users and the EMBL IT Services support unit and will represent a major financial investment in the next Programme.

3.1.7 Electron Microscopy Core Facility (EMCF)

Although light microscopy methods are in a phase of explosive development, many problems in cell and organismal biology benefit enormously from the access to higher resolution imaging provided by electron microscopy. The EMCF provides EMBL researchers with access to electron microscopes (EMs) and offers assistance in optimising protocols to be adapted to a variety of samples subjected to EM studies. The services further include plastic embedding methods, cryo-sectioning, high-pressure cryo-fixation and cryo-substitution.

The EMCF significantly expanded its capacity recently by introducing an electron tomography setup for cell structure investigation, which allows three-dimensional analysis of cellular structures at EM resolution. The setup is centred on a new electron tomography microscope (a 300kV Tecnai F30) equipped with a field emission gun and a high performance CCD camera. The newly available technology, which is mostly devoted to the electron tomography of plastic embedded cells, was in great demand among EMBL researchers. It has necessitated a remodelling of the microscope and cryofixation rooms of the EMFC. To make the facility as user-friendly as possible a tomography engineer supports the new microscope setup, software packages for cellular tomography have been improved and the acquisition process automated to facilitate the analysis of large cellular volumes.

The second major focus of the EMCF is developing Correlative Light and Electron Microscopy (CLEM) in collaboration with the ALMF in order to make it available to users. In particular, the EMCF has been developing two specific CLEM techniques:

- CLEM using laser etching on glass coverslips with cultured cells, and conventional Transmission EM after chemical fixation.
- CLEM using carbon landmarks deposited on sapphire disks, followed by cell monolayer cryo-fixation using high-pressure freezing. This method allows optimal cell morphology preservation. The approach will be developed further to optimise it for different types of biological samples studied at EMBL.

In the future, we are aiming to extend our present methods and make correlative microscopy possible for a larger number of cells and organisms using cryo-fixation for processing samples for electron tomography.

The EMCF also engages in user training. In future, regular basic courses are planned to introduce newcomers at EMBL to EM techniques and their potential use in cellular and developmental biology.

3.1.8 Flow Cytometry Core Facility (FCCF)

The Flow Cytometry Core Facility was established in 2004 to provide flow cytometry services to scientists at EMBL. The facility's equipment includes a high-speed sorter, an analyzer and an embryo sorter. Typical assays offered by the FCCF on these instruments include single and bulk cell sorting, embryo sorting, algal life cycle analysis, FRET, immunophenotyping, phosphorylation cascades and cell cycle analysis.

Due to increasing demand for flow cytometry for the characterization of mouse models an additional facility was established at EMBL Monterotondo in 2007, tailored to the specific needs of mouse biologists. Services offered include multi-colour cell immunophenotyping, hematological analysis, cell cycle analysis, intracellular signalling analysis, cytokine production assays, gene reporter assays, single-cell cloning and gene expression analysis. The Heidelberg and Monterotondo facilities collaborate closely.

In recent years a need for new FCS assays with higher sensitivity has evolved at EMBL. The FCCF staff has been working closely with facility users to develop new techniques adapting the equipment and identifying new components that can be fitted onto existing platforms to meet this need. New assays recently introduced include; purifying cell lines transfected with multiple fluorescent proteins/markers and preparing them for further analysis such as microscopy or microarray; and chromosome sorting and analysis for the identification of chromosome translocations. There has also been a great increase in demand for multi-colour sorting and analysis, in particular in immunophenotyping.

In 2009 the FCCF totaled over 1400 hours on the instruments representing 70% instrument occupation. The facility is also engaged in training researchers at EMBL and from the member states in flow cytometry and has been involved in setting up similar facilities in member state institutes.

Future developments that we foresee for the FCCF include:

- Increases in demand for sorting and analyzing, which will translate in a need for more staff and equipment, such as a multi-laser flow cytometer.
- Replacement of the aging embryo sorter in conjunction with Developmental Biology and Genome Biology Units.
- Creation of basic courses in flow cytometry aimed at pre- and postdoctoral fellows in the early phases of their stay at EMBL.

3.1.9 Mouse Transgenic Core Facilities

Transgenic mice are a critical part of the mouse research being carried out at EMBL and Core Facilities services are located both in Heidelberg (HD) and Monterotondo (MR). These facilities are involved in the routine production of genetically modified mice as well as in the implementation and development of new technologies associated with mouse embryology and transgenesis. With a team comprising one staff member and 7 technicians at the MR site, and one staff member at the HD site, these facilities offer a wide range of services to EMBL researchers – and, as capacity permits, to the European scientific community – including standard pronuclear injection for the production of transgenic animals (including the use of large BAC-based genomic fragments), and blastocyst injection of genetically modified embryonic stem (ES) cells for the production of mice with targeted mutations. Associated technologies such as rederivation by embryo transfer (required for entry of lines into both the HD and MR mouse facilities) and cryopreservation of embryos and sperm are also carried out on a routine basis.

In addition to these standard approaches, EMBL is continuously implementing and developing novel technologies to accelerate and improve the use of genetically modified mice, notably as models of human disease and genetic variation. During the present indicative scheme, we established tetraploid complementation and piezo-drill assisted 8-cell injection of hybrid ES cells that allow for the production of essentially 100% ES cell-derived chimeric mice and thus reduce costs associated with transgenic mouse production. In recent years, the demand for transgenic lines carrying large genomic fragments (propagated as BACs, PACs or YACs) has been increasing rapidly since they enable more accurate tissue and cell-type specific expression of genes in transgenic animals. BAC recombineering technology was developed at EMBL in the laboratory of Francis Stewart in the 1990s and is now in wide and routine use across EMBL. Both the HD and MR facilities are proficient in the production of BAC transgenics, notably through intra-cytoplasmic sperm injection (ICSI). ICSI is now routine at the MR site for the incorporation of BACs into the mouse genome as well as for the rederivation of cryo-preserved, infertile, and/or senescent lines. Several high profile publications during the present indicative scheme have depended directly on these technologies.

The development and implementation of novel transgenic technologies as well as the streamlining of existing services will continue during the next indicative scheme. The availability of thousands of conditional knockouts generated via the EUCOMM, KOMP and NORCOMM programs in C57BL/6N ES cells will require the adaptation of existing transgenic technologies to this strain. The growing number of mouse lines in use and the development of sophisticated genetic approaches will increase the need for efficient archiving and rederivation procedures based on sperm, rather than embryo cryopreservation. The progressive transfer to the Transgenic Facilities of novel approaches and expertise developed in EMBL groups, notably in lentivirus, Cre-RMCE, and PhiC31 integrase-mediated transgenesis, as well as in chromosomal engineering, will keep EMBL at the forefront of mouse genetics.

In addition, the proposed establishment of an ESFRI-funded Mouse Clinic at the MR site will place new demands on and offer new opportunities to the Transgenic Facilities, who will become a reference technology development center in this partnership. Finally, EMBL expertise in the production of transgenic animals using cutting-edge technologies is and will continue to be shared and disseminated in the European scientific community with annual training workshops and on-demand individual courses.

3.2 Central IT infrastructure and services

The advent of high-throughput technologies and the shift towards large-scale and interdisciplinary systems biology approaches have had a massive impact on the IT infrastructure required to support basic molecular biology research. The data quantities generated by next-generation sequencing, high-throughput microscopy, bioinformatics and modelling and simulation efforts are vast. The data collected by EMBL scientists are expected to enter the Petabyte range during the course of 2010 and then rapidly grow to multi-Petabyte dimensions. EMBL's internationally distributed nature adds another challenge to the provision of IT infrastructure to both internal and external users of EMBL services. The ability of EMBL to pursue the plans described throughout the Programme document will continue to be totally dependent on its IT infrastructure.

EMBL-EBI is facing the biggest challenge in terms of IT infrastructure because of its data resource service functions, but these will be described in the context of the bioinformatics services (see Section D.1.). In Heidelberg, IT Services and the Structural and Computational Biology Unit operate the largest IT infrastructure that supports the main laboratory and EMBL Monterotondo. Although the outstations in Grenoble and Hamburg run their own IT infrastructures that are optimized for the needs of beamline data collection by structural biologists, IT specialists at all EMBL sites work towards a unified approach whenever possible. At the centre of this collaboration are efforts to standardize and simplify existing infrastructures and to establish fast and reliable connections between them.

Although the specific requirements for IT infrastructure vary between the different EMBL stations, the common challenge of exponentially growing data quantities poses enormous demands on storage and compute power across all the EMBL sites. Given the current growth rates at EMBL Heidelberg, the storage capacity doubles every 18 months and a doubling of computing power is necessary each 18-24 months. By the end of 2016, the total data volume at EMBL Heidelberg is expected to exceed 15 Petabytes, which will require a central compute power of 20,000 CPU cores. Scalability, reliability and cost-efficiency will therefore continue to be the top priorities for IT infrastructures.

During the past five years, EMBL has succeeded in greatly improving the performance and stability of its high-performance computing (HPC) and storage systems in Heidelberg by consolidating previously heterogeneous systems. The compute infrastructure has been converged onto a single HPC cluster, which is accessible to scientists through a single batch queuing system, and 50 previously existing fileservers have been replaced by a standard high-performance storage solution now based on five systems. This strategy has allowed the IT infrastructure to be scaled up to a current size of ~2000 CPU cores in the HPC cluster and a total storage capacity of 700 Terabytes of primary online storage – compared with 70 Terabytes in 2005 – and 2 Petabytes of secondary storage on tape. These developments exhausted the capacity of the previous datacentres, which were in any case poorly located in relation to the maintenance of cooling. To accommodate future essential expansions of IT infrastructures and to ensure sufficient power and cooling capacity, the server facility has been moved to the periphery of the Heidelberg campus where it is located below ground level.

However, adding new hardware is not a sustainable solution to handling data quantities in the multi-10 Petabyte range. In future, alternative strategies that rationalize space, energy consumption and cost of data storage have to be implemented. One important approach is the virtualisation of large parts of the server environment in the EMBL Heidelberg datacentre, which promises additional positive effects on manageability and resilience of the server landscape. 100 physical servers have already been migrated to virtual structures that currently run on six physical host servers. 100 further physical servers and the majority of all future servers will also be virtualised. In addition, IT Services are exploring new strategies to customise the provision of storage according to specific needs, such as the use of different types and quality of hardware to store intermediate and final datasets and new policies regarding the life-span of storage of raw data from high-throughput experiments.

Data backup and recovery systems will also have to be adapted to the new data dimensions. The current tape media system, with a capacity of 2 Petabytes for backup and archives, is too slow to be a practical solution for the expected future data volumes. Moreover, quick recovery after substantial data loss is not possible. It will therefore be necessary to replace tape storage with a full disk-to-disk backup and archive strategy in the coming years, which will add a factor of 2 to 2.5 to the expected storage volume sizes. In the context of a future strategy for backup and resilience of EMBL's IT operation, the integration of external datacentres and cloud computing approaches is also being explored. On-demand and cloud-based services could cover peak demands for compute and storage resources in the future. EMBL-EBI is currently investigating and testing cloud computing and storage services of commercial providers and will advise the other EMBL sites on how to proceed.

Growing data quantities are not the only challenge systems biology poses to IT infrastructure. Interdisciplinarity is an integral feature of systems biology, which frequently requires many research groups to work together on a problem. EMBL, being distributed over five sites and cultivating a highly collaborative culture, needs an IT environment that allows collaborative access to data for researchers inside EMBL as well as external collaborators. To address this need, IT Services have begun to implement a new collaboration portal. Based on an EMBL-wide identity management system, the portal provides access to shared work spaces to securely manage project documents within teams, shared team calendars and real-time desktop conferencing, which allows teams to meet virtually from a user's desktop. In future, this infrastructure will be extended to allow the exchange of bigger datasets. The system will be constantly upgraded to follow future IT trends. A fully functional intranet that works across all five EMBL sites, and can be customized to suit the needs of each site individually, was installed in spring 2009 and will continue to be developed during the coming period.

A prerequisite for optimal connectivity between the different EMBL sites and with external data repositories or collaborators is a stable internet connection that allows fast and reliable data transmission. Scientists from different groups currently transfer multi-Terabyte datasets between the EMBL outstations – especially to and from the EBI – as well as to and from external collaborators or EMBL service customers. The speed and volume of such data transfers are severely limited in Heidelberg, Hamburg, Grenoble and Monterotondo by the existing 10 or 100 Megabit internet links, so that scientists currently have to resort to the manual transport of hard disks to transfer large datasets. This approach is not sustainable and an upgrade to large-bandwidth and low-latency internet connections will be an inevitable requirement at each site to provide high-speed access to data repositories and allow shared data processing on local clusters.

E. Training

1. EMBL's training mission and the EMBL International Centre for Advanced Training (EICAT)

Advanced training has always been one of EMBL's core missions. EMBL's unique fixed-term contract system means that we are constantly training high-quality staff members on both the scientific and administrative sides of the organisation. The group and team leaders who leave EMBL to enter the national systems are particularly appreciated by our member states. The EMBL International PhD Programme (EIPP), the new EMBL Interdisciplinary Postdoc (EIPOD) Programme, the Visitor Programme as well as the courses and conferences held at EMBL have also greatly contributed to EMBL's reputation for excellence in training. In 2004, EMBL created the EMBL International Centre for Advanced Training (EICAT), adding professional leadership, to integrate and better co-ordinate our scientific training activities at all five sites, to complement existing programmes by much-needed additional ones – for example, the Postdoctoral Programme – and to expand our ability to serve scientists across Europe with an attractive, cutting-edge courses and conferences programme. EICAT ensures that our limited resources are optimally used, redundant features are avoided, fruitful synergies are exploited and a single point-of-call is established for the training needs of scientists in the EMBL community.

EICAT has a dual mission: it coordinates training for scientists working at EMBL and serves as a hub that offers advanced training opportunities to scientists from EMBL's member states. EICAT aims to support scientists at all career stages so that they acquire the international, interdisciplinary, collaborative "EMBL culture", which should prepare them comprehensively for their careers after leaving EMBL and help them to succeed. We are increasingly complementing research training with opportunities to acquire managerial and leadership skills. Within the next Indicative Scheme, we hope to provide additional training opportunities for scientists at all levels, including preparing our more senior group leaders for

their future roles as heads of departments and institutes. EICAT also seeks ways to export training opportunities, for example through the PhD Programme's "Shared Applicant Pool", the EBI training roadshows, the support offered to other institutions in setting up their own training programmes, and by sharing content through distance learning and videostreaming.

EICAT acts on five levels to meet the training challenge posed by the increasing requirement for interdisciplinarity in modern life science research:

- Individual training programme initiatives such as the EIPP and the EMBL Postdoctoral Programme are implementing new training schemes specifically geared to broaden and make more interdisciplinary the skill-sets of young scientists at EMBL through structured research training activities. The EIPOD initiative and PhD students working in an interdisciplinary fashion under dual mentorship are examples, as are the many internal courses organised by EMBL scientists for both PhD students and other staff members.
- Complementary skills training is being offered to prepare young scientists to meet the complex requirements for a successful future career in either academia or industry. This is pursued in collaboration with the General Training and Development Programme (see Section I.1.2.3) and with EMBO.
- Visiting scientists and scholars are hosted to develop collaborative research projects or to learn new technologies at EMBL. Although the capacity for such training measures are limited, they nevertheless enable a considerable number of scientists to benefit from working at EMBL and to transfer knowledge and technologies back to their home laboratories.
- Organisation of state-of-the-art scientific conferences, workshops, and specialised courses at different levels, as well as complementary summer schools, allows EMBL to catalyse the exchange of information and expertise, not only across disciplines, but also throughout Europe and the member states, as well as between the scientific community and the general public.
- Distance learning methodology and 'new media' provide an additional mechanism to meet the ever-increasing demand for training. EMBL has begun to use these technologies to create communities of trainers and trainees, who are partners in a large training network that ensures better dissemination of training activities across Europe. The EMBL-EBI e-Learning initiative, the European Learning Laboratory for the Life Sciences (ELLS) teacher training e-networking projects and the new EMBL Online Seminars, which we have piloted with a small number of lectures from the Science and Society series, are examples of recent developments in this area. Many more of our training activities will be adapted for dissemination in this way in the next Programme period.

2. EMBL International PhD Programme (EIPP)

The EMBL International PhD Programme has been a cornerstone of EMBL's training mission for more than 20 years. Constantly adapting to new requirements and challenges, the EIPP underwent several structural changes during the present Indicative Scheme. The programme has now grown to almost 200 PhD students and is directed by a recently appointed full-time Dean of Graduate Studies, whose tasks include optimising recruitment and mentoring of our PhD students as well as anchoring the EIPP visibly within the member states.

The two-month PhD core course offered on the arrival of each new predoctoral class has been reorganized into modules that better reflect the increasing importance of interdisciplinary research at EMBL and to stimulate networking across Units and research groups. Together with the addition of a 'basic science module' taught by second- and third-year students, the curriculum fosters the acquisition of knowledge across physics, mathematics, chemistry and statistics for biologists and allows the PhD students who have not studied biology as undergraduates to familiarise themselves with the latest concepts in the molecular life sciences. A detailed description of the course curriculum can be found in Appendix E.1.

Against the general trend, the number of applicants to the EIPP has doubled during the past two years to more than 1200 per year. Importantly, applicant numbers from the EMBL member states have grown proportionally with the total. We aim to intensify outreach activities, particularly in under-represented member states, to further increase the high reputation and inclusiveness of the programme.

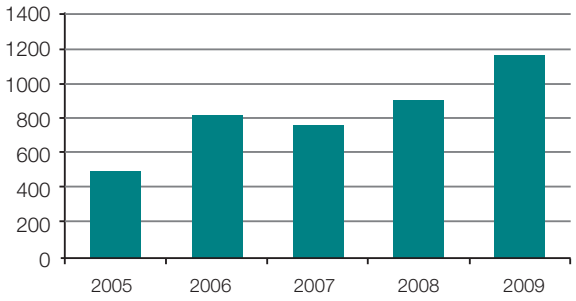


Figure E.1. Applications to the EIPP 2005-2009

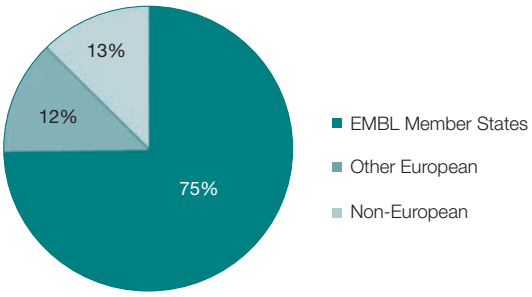


Figure E.2. Geographic distribution of admissions to the EIPP from 2008 to 2009 show that 75% of successful applicants come from EMBL member states.

The following new features will be implemented in the EIPP:

- The interdisciplinary nature of a PhD thesis at EMBL, reflected in the core course as well as in the composition of Thesis Advisory Committees, will be further advanced. Interdisciplinary projects involving co-advisors from different scientific backgrounds will become an integral theme of the EIPP, fostering the education of a new class of scientists prepared for the increasing complexity and cross-disciplinary nature of research projects in the future.
- Tailor-made mentorship and individual career development support for each predoctoral fellow.
- Forging tighter relationships to the undergraduate communities via student ambassadors and special Masters programme alliances together with refined advertisement strategies, workshops and recruitment events should allow EMBL to more specifically recruit students from under-represented member states.

Box E.1. The EMBL International PhD Programme at a glance

- 200 PhD students from over 40 nationalities are enrolled in the EIPP
- 50 new students join the EIPP on average every year
- 1 out of 20 applicants is admitted to the EIPP
- 19 partner universities in 29 countries
- 2:1 is the average ratio student to supervisor
- 3.9 years is the average time to thesis submission
- >50 annual graduations on average
- >95% of students successfully submit their thesis
- >90% of EIPP students achieve a first-author publication as a result of their PhD studies
- 2 first-author papers is the average number of publications of EIPP students
- 80% of EIPP graduates take up positions in the member states

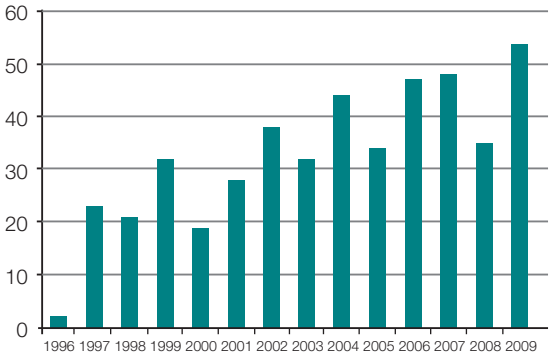


Figure E.3. Number of yearly PhD defenses since the inception of the EIPP in 1996

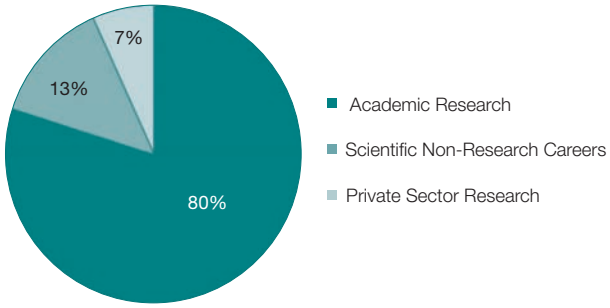


Figure E.4. Career progression of EMBL PhD students since 2004

- Developing the EIPP university partnership network by organising meetings of European Deans of Graduate Studies to exchange best practice information and explore collaborations.
- Sharing of resources with member state scientists, such as the "EIPP Shared Applicant Pool" that makes applications of excellent candidates that cannot be accommodated in EMBL's PhD Programme, and who agree to participate, available for screening by other academic institutions in the member states.
- The **EMBL Collaborative Training Programme** is EMBL's framework to provide expertise and models of best practice in the area of training to any interested institution in Europe. The programme has already led to the successful establishment of a structured pan-European PhD Programme within the EU Network BioMalPar, and also helps to establish new PhD programmes that follow the EMBL model in the three units of the Nordic Partnership for Molecular Medicine and Marine Biology.

3. EMBL Postdoctoral Programme

During the current Indicative Scheme, EICAT has acted vigorously to address the disparity between the effort expended by EMBL on its PhD students and the lack of a similar effort to attract, mentor and support outstanding postdocs. We are now well on our way to implementing postdoctoral mentoring schemes to prepare them for their future careers. As a first step, we have appointed a dedicated Academic Mentor for postdocs and a full-time Postdoctoral Programme Administrator.

Recognizing the rapidly increasing need for interdisciplinary expertise, particularly amongst postdocs, we utilised funding available for the establishment of cross-disciplinary centres to start the EIPOD Programme in 2007. Through this new initiative, EMBL underlined its commitment to promoting interdisciplinary research.

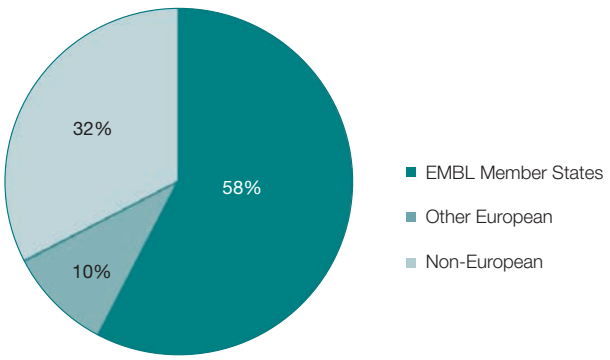


Figure E.5. Geographic distribution of the 286 postdocs working at EMBL 2008 and 2009.

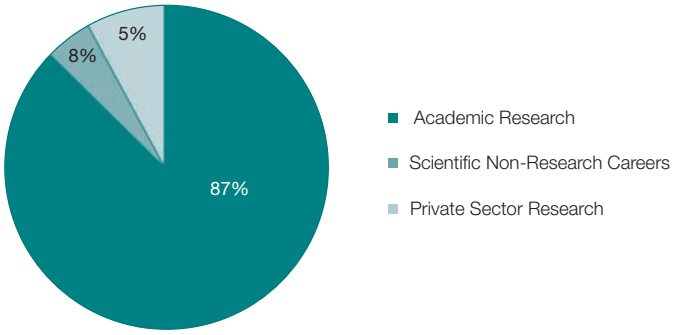


Figure E.6: Career progression of EMBL postdocs.

The EIPOD Programme encourages young scientists either to apply for one of several predefined interdisciplinary projects or to design an interdisciplinary project of their own that involves at least two EMBL groups. EIPOD selection is based solely on the scientific excellence of the candidates and research projects. The programme has quickly gained recognition across Europe. At present, EMBL can support 10-12 EIPODs per year, but we hope to obtain support from the member states to expand this valuable programme in the next Indicative Scheme. The success of the EIPOD Programme will strongly affect different facets of EMBL's postdoctoral community during the next decade. With the current growth of the scheme, interdisciplinarity will be established as a hallmark of the EMBL Postdoctoral Programme. Additional external funding sources, to complement dedicated core funding, will allow us to consolidate and expand this programme. The positive effect of the outstanding quality of the applicants and the collaborative research activities they engage in and stimulate can already be seen throughout EMBL and underlines the success of this initiative.

A second innovation for postdoctoral fellows is the new "Second Mentorship" scheme, introduced in September 2009. This scheme is devised to encourage all postdoctoral fellows to select an EMBL faculty member as a second mentor who can provide additional objective advice and guidance.

The third new aspect of both our postdoctoral and PhD programmes, already decided by EMBL Council, is to provide fellows with access to the EMBL social benefit programmes, at an initial cost of roughly €1.8 million.

4. Mentoring of young Group Leaders

Although not organized as a dedicated training programme, EMBL regards its provision of mentoring and complementary skills and management training to young group leaders as an important responsibility. New Group Leaders attend a mandatory course in leadership and lab management to help them face the challenges of running a group. The Heads of Units have a strong mentoring function for the Group Leaders in their Units and are generally the first point of contact if questions arise. To make sure the PIs have access to broad and interdisciplinary expertise, each Group Leader is also asked to choose a second mentor in addition to the Head of Unit. Additionally, the regular Unit reviews organized by the Scientific Advisory Committee provide guidance, advice and mentoring for Group Leaders, as do the informal interactions they engage in with their colleagues, in particular at regular retreats organized by the EMBL Units.

5. EMBL Visitor and Scholar Programmes

The EMBL Visitor and Scholar Programmes each year enable more than 400 students, postdocs and more experienced scientists from all over the world to benefit from the scientific environment, the new technologies and the individualised training on state-of-the-art equipment available at all five EMBL sites. Hosting visitors is an impor-

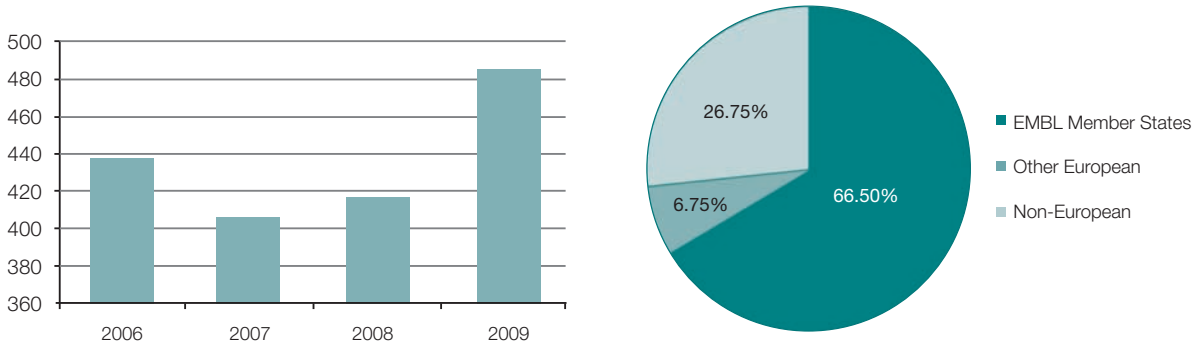


Figure E.7. Numbers and geographic distribution of visitors to EMBL from 2006 to 2009

tant EMBL mission, not only to foster collaborative research projects between EMBL scientists and their colleagues in member state research institutions, but also to allow the visitors to study or learn new technologies at EMBL and to take this knowledge back to their home institutes.

The Visitor Programme, in close collaboration with the EIPP, will strive to provide additional opportunities for short-term visits of undergraduate students, with the idea of introducing them to the EMBL scientific environment early in their careers. This effort is particularly focussed on students from disciplines such as physics, chemistry and engineering.

The EMBL Scholar Programme offers support to sabbatical visitors who wish to engage scientifically with EMBL. While necessarily limited in scope, the opening of the ATC means that EMBL Heidelberg will finally be able to offer dedicated office space to visiting scholars. Raising awareness in the EMBL member states about the benefits of the Visitor and Scholar Programmes is a challenge we aim to tackle during the next Programme.

6. Courses, conferences and workshops

Complementing the training opportunities within our research groups and Core Facilities, EMBL offers courses and conferences at all of its sites. The EMBL-EBI User Training Programme and the courses and conferences hosted in Heidelberg are numerically the most important, although courses and workshops are also offered in Grenoble and Hamburg for structural biologists and in Monterotondo for mouse biologists. With the opening of the Advanced Training Centre (ATC, see Box E.2.), we will enter a new phase in the provision of scientific courses and conferences at EMBL Heidelberg.

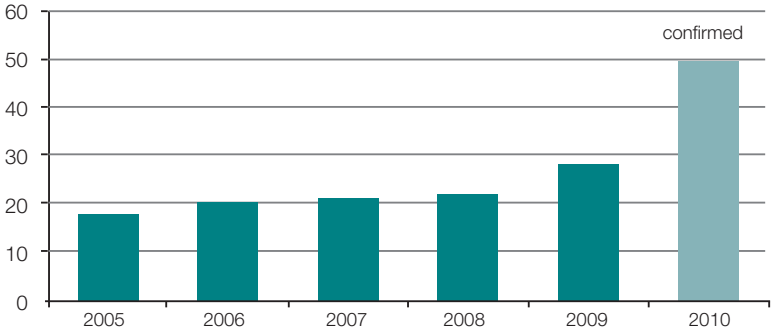


Figure E.8. Number of courses, conferences and workshop organised at EMBL Heidelberg from 2005 to 2009 and confirmed events for 2010

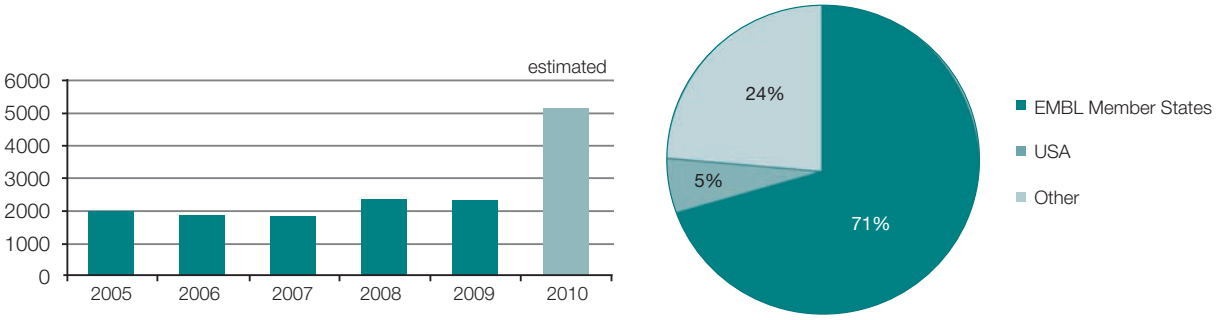


Figure E.9. Numbers and nationalities of past course and conference participants and estimated numbers for 2010

We expect to nearly triple the number of course participants. Although we are pleased to be able to accommodate a greater fraction of applicants than was previously possible, our explicit emphasis will remain on quality rather than on quantity in terms of both the scientific content and the organisation of all events. Gearing up for the increased activity, the EMBL Course and Conference Office (CCO) has been re-organised and expanded in order that scientific organisers can continue to dedicate their time to the content of the course. Importantly, the establishment of the EMBL Corporate Partnership Programme (see Section I.2) enables us to directly support young participating scientists who, for financial reasons, would otherwise be unable to attend.

Box E.2. The EMBL ATC

Although EMBL has been very active in the organisation and provision of courses and conferences in the past, the opening of the Advanced Training Centre (ATC) finally provides us with the opportunity to offer scientific training in a purpose-designed setting.

Constructed during the course of the current Indicative Scheme, the new training centre opened in early 2010. The ATC offers state-of-the-art teaching laboratory facilities for practical courses, a computer training lab, a poster exhibition area for up to 320 posters and a 470-seat auditorium. In addition, it will accommodate EICAT as well as other administrative and support staff. The ATC also includes a new canteen for employees and conference attendees.

With the ATC, we wish to take further steps towards the provision of convenient and affordable conference participation. At present, participants rely on accommodation provided by hotels in Heidelberg. However, we are aware that hotel costs in Heidelberg are high and that off-site accommodation means time-consuming commutes to and from the EMBL campus. To combat these problems and to offer conference attendees optimal networking opportunities, a guesthouse offering accommodation next to the ATC, for some 100-120 participants, is highly desirable. We have space for such an addition to the ATC and will engage in efforts to raise the funds required during the next Programme.

Our explicit ambition is to maintain EMBL's traditional role as a major hub for advanced life science training in Europe. Intensifying our collaboration with EMBO, we will implement several new activities:

- Jointly financed by EMBO and EMBL, the new EMBO | EMBL Symposia will provide a platform to discuss and exchange ideas on challenging, forward-looking topics and new developments in the life sciences. Five or six such symposia will be organised in the ATC in each of the years to come.
- Practical courses within the new training laboratories will be specially tailored to different levels of expertise in emerging techniques. In addition to the largely EMBO-supported Master Courses, we will implement two additional levels (Advanced and Introductory Courses).
- Reaching out by organising more events that appeal to other scientific fields – particularly the clinical research and medical communities – will be important to foster networking interactions between EMBL scientists and those from other research areas.
- Building sustainability into our training strategies is of vital importance. EMBL-EBI is piloting 'training support networks' for previous EBI roadshow training event hosts and undergraduate lecturers and similar schemes have already worked well for ELLS. There is considerable potential to increase Europe's scientific trainer base through dissemination schemes that use new communication technologies that have the potential to further increase the impact of our courses and conferences in the EMBL member states (see Box E.3.).

7. ELLS, EMBL's link to schools and school teachers

The main mission of the European Learning Lab for the Life Sciences (ELLS) is to offer professional development for European high-school teachers but it also provides a welcome opportunity for EMBL scientists to engage in these outreach activities by teaching on a voluntary basis. The core of ELLS activities are LearningLABs, which are three-day courses for small, international groups of high-school teachers that focus on hands-on experimentation and the development of innovative teaching materials. The long-term goal is to contribute to enhancing students' scientific literacy and to attract more young people into science careers. ELLS internal funding is limited but is supplemented by external grants.

Responding to increasing demand from teachers in Italy, ELLS activities were also established in Monterotondo in 2007. In collaboration with the EMBL alumni community, LearningLABs have been offered in several EMBL member states, and have been attended by more than 700 teachers from over 20 countries during the past five years.

ELLS also generates visibility for EMBL by exhibiting at outreach events in Europe and beyond. These include, among others, German Science Days, the Italian Brain Awareness Week and the European Science Open Forum (ESOF). ELLS also contributes to the highly successful EIROforum Journal 'Science in School'.

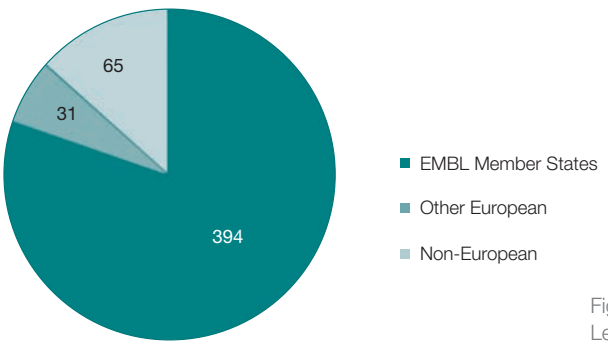


Figure E.10. Geographic distribution of Learning Lab participants (2007-2009)

Box E.3. The EMBL-EBI User Training Programme

Throughout the current Indicative Scheme, EMBL-EBI has built up an extensive User Training Programme. Its purpose is to enable users and potential users to make the most of the EBI's bioinformatics services and it comprises:

1) Training roadshows designed to provide bespoke bioinformatics user training to groups of users throughout Europe. This began with a pilot in June 2006, and since January 2007 we have delivered 42 roadshows in 18 countries, 31 in EMBL member states. This has enabled us to train more than 1400 people, mostly PhD students and early postdocs, and is a substantial contribution to EMBL-EBI's overall training effort. Funding from the European Commission for the SLING integrating action is now allowing us to take the roadshow to the new EU member states.

2) The hands-on User Training Programme was launched in September 2007 and provides regular training courses, aimed at bench biologists, in the EMBL-EBI IT training suite. Wellcome Trust and EMBO funding for some courses allowed us to invite external speakers, provide travel bursaries for a small number of attendees and provide accommodation on the Wellcome Trust Genome Campus at subsidised rates. We will try to obtain external funds for all of our hands-on courses in future but we will need to maintain support from core funds until this is achieved. Since its launch, the hands-on programme has trained over 450 researchers, mostly PhD students and postdocs.

3) In addition to the hands-on training programme, the new IT training suite allows us to host a large number of other courses and workshops. These include: Industry Programme workshops, internal training for staff and students, workshops organised through our many collaborative research and service projects and computer-based training offered by the Wellcome Trust Advanced Course Programme and the Wellcome Trust Sanger Institute. In 2009, the occupancy of the training suite reached 82% of working days.

4) We began building a pilot e-learning portal in December 2006. The original purpose of the portal was to enable us to train a much larger number of researchers than would be possible through face-to-face training. The pilot taught us that the portal was equally useful as a repository of training materials for trainers, and as an adjunct to face-to-face training. We are now testing platforms for a new portal, which will be populated with courses over the years to come.

Figure E.11. Locations of past EMBL-EBI roadshows and roadshows scheduled for 2010 in Europe

7.1 SET-Routes Programme

The EU-funded SET-Routes Programme is a pan-European network of women scientists at all career stages. It aims to inspire students, particularly female students, to study science and to pursue careers in science, engineering and technology (SET). During the current Indicative Scheme, the programme provided an excellent opportunity for ELLS to collaborate with the EMBL alumni community, EMBO and CERN to stimulate interest in science and to encourage female graduates to pursue careers in SET through a very active and visible School and University Ambassador Programme, an international conference – 'The Way Forward' – and a Multi-media Insight Lecture Series portraying 10 influential women scientists.

7.2 iNEXT Programme

Funded by the Robert Bosch Foundation in 2007, iNEXT, an interactive Network for Experimental Training, is a three-year pilot project, encompassing ELLS LearningLABs and students' workshops in a unique collaboration between EMBL and local schools to develop enquiry-based science education resources.

During the past decade, initiatives to improve scientific literacy have mainly targeted students. Now, major Europe-wide efforts that address teachers are being implemented to foster evidence-based science teaching in European schools. With a decade of experience in this area, ELLS is in a prime position to actively contribute to these important educational initiatives.

The major focus of ELLS activities in the future will be:

- Employment of new media and innovative teaching methodologies to allow faster and more efficient dissemination of its resources to the member states.
- Developing a European Network of LearningLAB nodes across the EMBL member states involving EMBL alumni as local 'catalysts' to enable local development of LearningLAB modules adapted to specific national curricula.
- Expansion of the School Ambassador Programme, in collaboration with the EIPP and Postdoctoral Programme, to encourage young PhD students to pay visits to schools back home and to entice young future scientists to follow in their footsteps.

F. EMBL's impact on the economy and society

This section provides an overview of the many ways in which research in the life sciences in general and at EMBL in particular impacts on society and the economy.

Research at EMBL is aimed at elucidating basic principles of biology and contributing to the fundamental knowledge base. At the same time, science also lies at the heart of innovation. Basic research provides a deeper understanding of the world in which we live and highlights the fact that we need to interact with it in a responsible manner. The needs of researchers pursuing increasingly complex questions push the boundaries of technology and serve as a motor for progress. Worldwide, commercial hubs for innovative technologies primarily grow up around renowned basic research institutes and many of the technologies we routinely use in our everyday lives, such as computers, recombinant medicines or the internet, were initially developed for research purposes. In the biomedical sciences, EMBL's focus and strength – curiosity-driven research – aids in the development of new medical tools and devices, provides mechanistic explanations for diseases and identifies potential drug targets or small molecule leads that can be translated into new drugs and therapies.

In this section we present a few case studies of the applications of discoveries, inventions and technologies that were initially made at EMBL. The translation of a basic research finding into a product for end-users frequently takes many years. For drug development in particular, periods of 15-20 years from bench to application in patients are standard. This means that although the applications showcased here have been developed in the past five years, 2005-2009, the original finding(s) or invention(s) sometimes happened significantly earlier.

EMBL actively engages in technology transfer through its commercial subsidiary EMBL Enterprise Management Technology Transfer GmbH (EMBLEM) to facilitate the translation of basic research discoveries into practical applications. These activities will be described in the second part of this section. An invention arising from basic research often needs to be taken

through to a 'Proof of Concept' stage before it will attract the interest of a commercial partner and can be licensed. The same holds true for raising venture capital to fund a start-up company. The further development required is often beyond the core research remit of basic research institutes, yet without this step, an invention or discovery is at too early a stage to engage commercial partners. From the small technology transfer surpluses generated by EMBLEM, EMBL, in the past five years, has invested modest amounts into development funding through the Technology Development Fund. This funding has allowed four projects to be further developed and successfully commercialised (see below). Building on this success and in line with comparable research organisations in the member states (e.g. MRC (UK), INSERM (France) and MPG (Germany)) we envision the formal establishment of a more substantial 'Proof of Concept' Fund at EMBL for the next Programme.

Aside from its technology-transfer activities, there are many ways in which EMBL interacts with and supports European bio-industries, for example through the provision of access to crucial research infrastructures and scientific training, the EMBL-EBI industry programme, EMBL's active role in the European Commission's Innovative Medicine Initiative and Pistoia Alliance, collaborations with the patent office, the ATC Corporate Partnership Programme and through close interactions between beamline scientists, core facility staff and industry for beta-testing of new technologies and provision of expert user advice. These activities are described in more detail in the third part of this section.

Apart from their impact on the economy, science and technology are integral parts of our daily lives and form the basis of any knowledge-based society. The tremendous potential the life sciences hold for societal benefits endows scientists with the social responsibility to inform the public about advances in their research, as well as potential applications, inherent risks and benefits and any ethical implications. As a publicly funded research institute, EMBL strives to make the public aware of the value of life science research for society and human well-being. We see it as part of our mission to engage in a dialogue with the public about our science in order to allow citizens to make informed decisions on scientific matters and to make our scientists aware of public concerns. Our efforts in communicating science and EMBL's successful Science and Society Programme are described in the fourth and fifth parts of this section.

1. Case studies of research-driven impact

a) SCAN-R: an automated high content microscope

EMBL's Advanced Light Microscopy Facility (ALMF) is extremely active in developing technologies to push the boundaries of imaging as a tool for life science research. One example of this effort is SCAN-R, a fully automated

Figure F.1. Set-up of the SCAN-R automated high content microscope

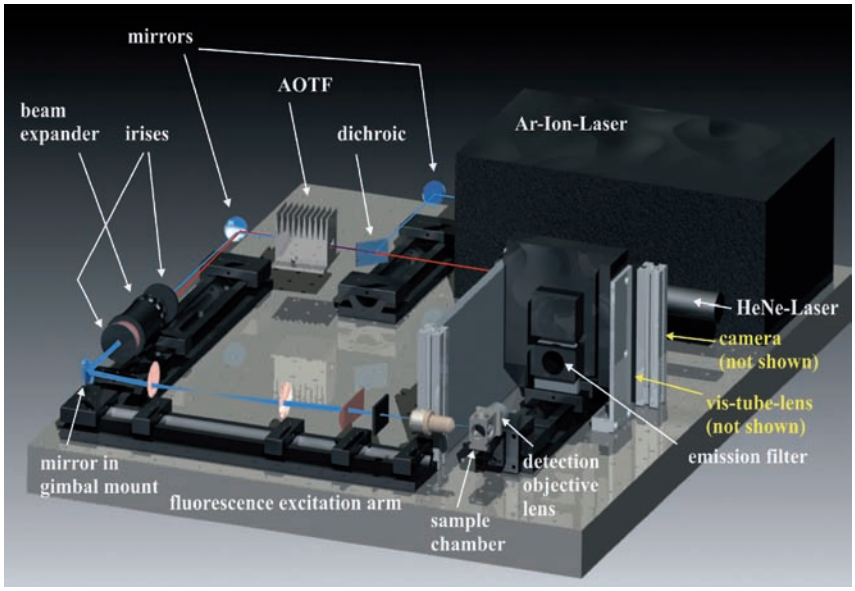


Figure F.2. EMBL's first SPIM system on an optical bench

high content screening system for light microscopy. SCAN-R includes automatic image recognition and data analysis. The system concept had been in development at EMBL since 2000 and was protected by a patent application in 2002. Following a commercial agreement in 2003, the system was jointly developed by EMBL scientists and Olympus, leading to the first commercial prototype of a high content screening microscope in 2005. In 2006 SCAN-R was successfully introduced onto the market and has since then evolved into one of EMBL's most successful commercial instrumentation applications, benefiting academic and industrial researchers worldwide.

b) ARP/wARP: A software suite for automatic structure determination from crystallographic data

A major bottleneck in crystallography is the interpretation of electron density maps produced in experiments that expose crystallized proteins to X-rays. Their transformation into three-dimensional structures requires complex mathematical modelling. In the 1990s, 'off the shelf' software solutions for protein-structure determination were rare and insufficient for research at the cutting edge of structural biology. This led scientists at EMBL Hamburg to start developing the software package ARP/wARP that was initially intended for academic purposes. Owing to its great success in the scientific community and a rapidly evolving demand in industry the software was marketed as a product for specific use in protein-structure research, targeted at a small commercial user community. It is licensed at no cost to academic users and multipliers and available for commercial use for a small fee. ARP/wARP now has several thousand users in academic and industrial research and has been continuously updated during the past ten years to meet changing user needs and hardware environments. The latest version, ARP/wARP 7.1, was launched in January 2010.

c) SPIM/DSLM: Single Plane Illumination Microscopy

Single Plane Illumination Microscopy, invented by EMBL scientists in 2002, revolutionised the field of light microscopy and live imaging. It allows scientists to study large, living specimens from many different angles, under physiological conditions and with minimal harm to the specimen. Images of the specimen obtained in different planes along different axes and at varying time points are assembled into three-dimensional images or movies, which provide insights into the dynamic cellular processes occurring in living organisms. The patent application for SPIM was filed in 2002. To allow the broader scientific community to benefit from SPIM, a license agreement was concluded with Zeiss in 2004 and the first prototype for commercial application was developed in 2005. It will be commercially available in the near future.

In 2008 SPIM underwent a major upgrade with the introduction of illumination of the specimen by a thin laser beam (Digital Scanned Laser Light Sheet Microscope, DSLM). DSLM makes it possible to scan an organism line

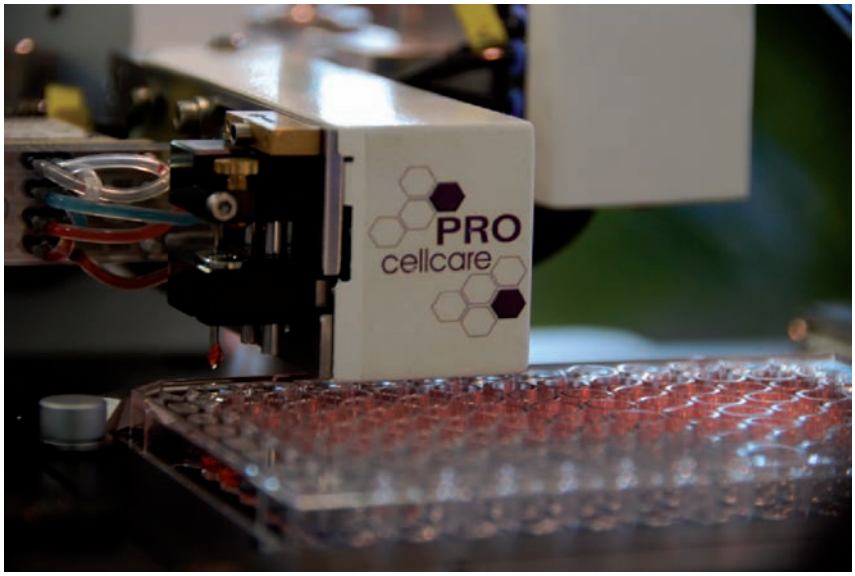


Figure F.3. The PROcellcare system

by line, horizontally and vertically, and thereby further minimizes the damage to the specimen. This upgrade allows living organisms to be studied over extended periods of time and was critical for obtaining the first complete developmental blueprint of a vertebrate, the digital zebrafish embryo (see Section C.1.1).

d) PROcellcare: An automatic dispenser system for microscopy

Based on an invention made at EMBL and supported by the EMBL Technology Development Fund, scientists from the Advanced Light Microscopy Facility and staff from EMBL's Workshops built a first demonstration model of an automated dispensing system in 2008. In collaboration with the external engineering company PROdesign, a commercial prototype of the system was developed. The system allows for fully automated dispensing and removal of liquids in an incubation chamber positioned in the optical axis of a microscope, assuring the stability of cell-culture conditions within the incubation chamber and thereby facilitating live-cell imaging under changing conditions. The dispenser was developed for commercial application within 6 months, patented and licensed to PROdesign through EMBLEM and has been available to academic and industrial researchers under the name 'PROcellcare' since October 2009. Future applications of the dispenser system include high-throughput compound testing in cell assays and *in vitro* online diagnostics.

e) SAVIRA: an EMBL spin-out company for the development of anti-viral drugs

In recent years, researchers at EMBL Grenoble have achieved multiple breakthroughs in determining how the influenza virus infects human cells. Influenza is a major concern for governments and health organisations around the world. Influenza pandemics, which arise when influenza viruses that infect animals develop the ability to infect and be transmitted between humans, pose a serious threat to global health. These potential pandemics call for new drugs that can be used to halt the spread of the virus.

Using a structural biology approach, EMBL scientists obtained high-resolution images of several crucial domains of the influenza virus polymerase, the enzyme that copies the virus' genetic material and allows it to multiply in human cells. The polymerase has long been considered a potential drug target for new therapies against influenza because experimentally inhibiting its function by genetic methods prevents the virus from replicating and spreading in the host. However, it has been impossible to produce the polymerase protein or to crystallize it and solve its structure in the past, making it very difficult to attempt to develop inhibitory drugs. The new high-resolution images produced at EMBL were key to identifying chemical compounds with properties that would allow them to bind to and potentially inhibit the influenza polymerase. Screens, financed by the EMBL Technology Development Fund and carried out in EMBL's Chemical Biology Core Facility, identified small molecules that bind to the protein *in vitro*.

In July 2009, the EMBL researchers and EMBLEM together with the biotechnology company Onepharm Research and Development GmbH founded Savira Pharmaceuticals, a start-up company focusing on the development of drugs for the treatment of influenza. Based on the first promising hits identified at EMBL, Savira, located in Vienna, Austria, runs screening and medicinal chemistry programmes aimed at the identification of further lead candidates and improving their properties as drugs. The EMBL spin-off company exploits the high-resolution structure information for structure-based drug design and will develop small molecule inhibitors that target the influenza virus polymerase to the stage at which larger pharmaceutical companies will step in to finance clinical trials. Savira and EMBLEM thereby help to bridge the gap between basic research and the pharmaceutical industry.

f) ELARA: an EMBL spin-out company developing oncology drugs for chemotherapy

ELARA Pharmaceuticals GmbH was founded in 2006 by EMBL scientists and EMBLEM, with the help of seed financing from EMBL Ventures, to exploit basic research findings made at EMBL in the field of oncology and translate them into novel anti-cancer drugs. ELARA is also supported by the German Federal Ministry of Education and Research (BMBF) under its GO-Bio initiative.

The spin-out company follows up on promising small molecule leads that have shown powerful anti-cancer actions in screening experiments in model systems conducted in EMBL's Chemical Biology Core Facility. ELARA develops promising drug candidate molecules further and evaluates their activity in animal models of various types of tumours, such as lung and breast cancer. In particular, ELARA focuses on molecular targets including the Hypoxia Inducible Factor (HIF) and oestrogen receptor signalling.

The direct flow of information that ELARA established from basic research to preclinical development and application studies helps to speed up the lengthy process of drug development.

2. Technology Transfer

2.1 Current Status

Groundbreaking basic research drives innovation. Pro-active technology transfer is a natural extension of EMBL´s activities and an integral part of the institute's mission to ensure that selected basic research-derived innovations are converted into marketable and consumable health products – such as medicines, diagnostic tools, machines and devices – to benefit the member states and society at large. The technology-transfer activities of EMBL are carried out exclusively by EMBL Enterprise Management Technology Transfer GmbH (EMBLEM), a wholly owned limited liability company, which was established in 1999.

EMBLEM´s pro-active technology-transfer approach ensures the rapid commercial development of promising innovations while concomitantly securing the free dissemination of knowledge for basic research purposes. This approach combines early sourcing and identification of promising innovations with professional technology assessment and development services, targeted intellectual property protection, active management of the intellectual property portfolio, and commercialisation. Through education and teaching initiatives coupled with site visits in member states and regular contributions to events of the Association of European Science & Technology Transfer Professionals, EMBLEM acts as a source of best practice in Europe for technology transfer from an academic setting.

During the past five years, EMBLEM has achieved major goals that were outlined in the EMBL Programme 2007-2011. Since 2005, 215 invention disclosures were received, 51 priority patent applications were filed and 24 software copyrights were secured. More than 800 license agreements were executed and annual turnover has more than doubled to over €4 million and totals €12.5 million for the period 2005-2009. The EMBLEM supervisory board, which includes the chairs of EMBL Council and the Finance Committee, has repeatedly expressed their surprise that EMBLEM has continued to increase its turnover through what has been a very difficult period for the biotechnology industry. They are of the opinion that the secrets of this success are the excellent performance of EMBLEM staff combined with the quality of the discoveries being made at EMBL. Significant royalty income has been gener-

Name	Phase
Sygnis Pharma AG (former Lion Bioscience AG) (1997)	Post IPO
Cenix Bioscience GmbH (1999)	2nd round
Cellzome AG (2000)	4th round
Anadys Pharmaceuticals, Inc. (2000)	Post IPO
Gene Bridges GmbH (2000)	-
ENVIVO Pharmaceuticals, Inc. (2001)	3rd round
Triskel Therapeutics Ltd. (2006)	-
Elara Pharmaceuticals GmbH (2006)	2nd round
BioByte Solutions GmbH (2008)	-
Savira Pharmaceuticals GmbH (2009)	Seed

Figure F.4. EMBL's spin-out companies

ated in the past five years, accounting for nearly 10% of the total turnover. Four EMBL spin-outs were created in the period, including Elara Pharmaceuticals and Savira Pharmaceuticals (see Section F.1e) and f))

Currently, EMBLEM manages an intellectual property rights (IPR) portfolio of more than 260 granted patents and patent applications and more than 90 copyrights and trademarks. These numbers mark an increase of approximately 40% compared with five years ago. The licensing portfolio has increased by 60% in the past five years and encompasses more than 400 active license and consultancy contracts. The licensing ratio of the IPR portfolio lies at 38%. There are 10 EMBL spin-out companies at present.

The success of technology-transfer activities is reflected in the broad engagement of scientific staff as well as in the large number of licensees of EMBL technologies. More than 400 EMBL scientists are on record as inventors. Taking staff turnover during the past 10-year period into account, statistically every third EMBL scientist has been actively engaged in technology transfer. Of the more than 250 commercial licensees of EMBL technologies, more than half are recurring customers interested in establishing a long-term relationship with EMBL and EMBLEM.

EMBLEM frequently collaborates with technology-transfer divisions of other national and international research institutes and universities and thereby is helping to establish a technology-transfer network across Europe. EMBLEM extends its services to EMBL alumni and, in a consortium with the technology-transfer division of the German Cancer Research Centre (DKFZ), has been responsible for the technology-transfer activities of the Heidelberg University Medical Faculty and associated clinics since 2007. These activities help strengthen the existing collaborations between EMBL and Heidelberg University and ensure that EMBLEM's IPR portfolio is enriched with additional high-quality inventions of direct medical relevance.

2.2 Future Outlook

2.2.1 EMBL "Proof of Concept" Fund

Many academic research institutes are faced with the problem that promising research inventions are just 'too early' to be embraced or licensed by commercial partners yet 'too advanced' to be further nurtured by the scientific inventors. Experts often refer to this gap as the valley of death. The technology transfer activities of EMBLEM broke even in 2004, a mere five years after inception, and since have generated a steady surplus year-by-year that has resulted in EMBL recouping its initial seed investments in their entirety and even a modest surplus. Part of this surplus has been used to fuel technology-transfer activities by further developing promising inventions from the laboratory. Several of these projects, carried out at the EMBL Chemical Biology Core Facility, entailed the identification and refinement of lead compounds against novel targets identified by EMBL researchers. All of these projects were successfully completed and licensed out. In another project, funds were used to develop the commercial prototype

of an automatic dispensing system invented at EMBL with an external engineering company. The system allows liquids to be dispensed into and removed from an incubation chamber positioned in the optical axis of a light microscope in a fully automated manner. It was developed for commercial application within 6 months and is now broadly available to academic researchers and industry (see Section F.1d))

Emboldened by the success of the projects described above and, in order to take more promising innovations through to enablement and commercial utility, the formal establishment of a 'Proof of Concept' Fund is envisioned. The intention is to increase the societal and economic benefits of EMBL's research and technology development. A minimum fund volume of €5 million for the next five-year period would be required. It is proposed that approximately half of the fund volume would derive from the EMBL budget or from surpluses generated by technology transfer and the other half would be matched by institutional investors. A thorough selection process involving the three external technology transfer professionals that sit on EMBLEM's supervisory board will decide which projects will be funded by the Proof of Concept Fund.

2.2.2 General Development

EMBLEM's commercialisation strategy balances short-term fixed income (upfront payments) against long-term variable commercial return in the form of milestone payments, royalties and equity stakes in start-up companies. For additional leverage, license agreements are often bundled with consultancy services and collaborations. This strategy will be continued between 2012-2016 to secure the long-term sustainability of technology-transfer activities.

Specifically in the period 2012-2016, EMBLEM expects to receive more than 200 invention disclosures and to grow its IPR portfolio by another 20-30%. The execution of more than 750 new licensing-, consultancy-, and collaboration agreements is planned. Cumulative turnover is expected to increase by approximately 50% to more than €18 million. Royalties, in absolute terms, are envisaged to increase by roughly three-fold. The goals of the activities in the forthcoming five years are: to maintain the high level of technology-transfer services to EMBL and the member states; to establish a Proof of Concept Fund and fund 10-15 promising inventions to commercial maturity; and finally, to have at least one EMBL-derived invention in clinical phase II studies.

Box F.1. EMBLEM technology transfer in numbers (2005-2009)

Turnover (k€)	12,500
Invention disclosures	215
Priority patent applications	51
Software copyrights	24
New license agreements	825
Material transfer agreements (MTAs)	>1,000
Confidentiality agreements (CDAs)	>150
Inter-institutional agreements (IIAs)	69
Start-ups created	4

3. EMBL's support for European bio-industries

Aside from its technology-transfer activities, there are many different and often less formal ways in which EMBL interacts with and supports European bio-industries. This chapter provides an overview of how EMBL benefits industry in the member states. It is our view that by working closely together with the life science industry, in partnership with other stakeholder groups, we can improve the competitive position of European researchers and enable better translation of fundamental research into new advances in medicine, health, and agriculture, for the benefit of society.

3.1 Collaborations and consultancies

EMBL does not carry out commissioned research for industry, but where research interests overlap and collaborations are mutually beneficial synergies can be exploited and EMBL scientists work together with industrial R&D groups. During the period 2000-2009, 42 EMBL scientists collaborated with small to medium-sized enterprises (SMEs) and 44 scientists worked jointly with pharmaceutical companies to solve biological questions. In addition to these project-based collaborations, EMBL scientists also make their expertise available through consultancies with companies. In the past ten years EMBL scientists have had 135 consultancy agreements with SMEs and 63 provided advice to the pharma industry through consultancies.

3.2 Research infrastructure

The operation of key research infrastructures is of great benefit to academia and industry alike. It is not by chance that centres of maximal innovation, such as Boston or the Bay Area in the US or Cambridge, Munich and Heidelberg in Europe, develop around concentrations of leading research institutions where this infrastructure is available. Research infrastructure forms the link between ideas and innovation: it attracts talent, generates knowledge, expertise and skills, drives technology development and transfer and is thus a pillar of any knowledge-based society. Industry represents at least 20% of the usage of EMBL-EBI's biomolecular databases and the structural biology services in Hamburg and Grenoble are also used by industry. Moreover, beamline engineers at the structural biology outstations and Core Facilities staff across EMBL engage in regular interaction with bio-industries, for example to engage in joint technology development, beta-test new technologies or provide input and advice as expert, high-end users.

3.2.1 The EBI Industry Programme

For the past 12 years the Industry Programme has been an important part of EMBL-EBI, providing regular contact with commercial users and generating feedback on our priorities. It provides extensive information and technology transfer, through transfer of open-access software and pipelines and access to expert staff. Our efforts address the needs of two different kinds of companies:

- Large, often multinational, companies that exploit the data and methods of bioinformatics and usually have their own bioinformatics departments. Most typically these are pharmaceutical companies, but they also include companies in the areas of agriculture and biomedical instrumentation.
- SMEs, which are typically service industries with diverse niches, and often have some of the larger companies as customers.

Among the larger companies there is more commonality of interest, and more scope for long-term strategic planning and coordination, making it easier for us to provide structured support. For the SMEs we arrange discussion forums of a general nature at different geographic locations as well as thematic meetings on an *ad hoc* basis.

The larger companies subscribe to the Industry Programme, the activities of which take several forms:

- Quarterly meetings, which allow us to present recent developments at EMBL-EBI and receive input from the partners about their needs and priorities. Knowledge exchange workshops with industry partners sometimes result in the identification and documentation of shared needs among industry companies that can be considered as 'pre-competitive'. Such projects may relate to the development of standards, support for data

resources in the public domain, public information integration activities and the development of new services.

- Hands-on training. From 2007 to 2009 we arranged 24 training workshops with a total of 968 places.
- Topic-specific meetings. For example, discussions of the European Innovative Medicines Initiative (IMI), automatic information extraction from literature, or gene-centric data summaries of drug targets.
- Specific development projects are sometimes undertaken with financial support from one or more of the companies.

The influence of the Programme partners has been crucial in several ways, including the establishment of cheminformatics and chemogenomics resources at EMBL-EBI, the function of which is to link biological data with the chemistry of metabolites and drugs. These new resources help to address the 'druggability' of potential targets and the integration of different types of data, including the literature, to obtain a systems-level perspective. Our industry partners have been convinced of the value of having these databases in the public domain to the extent that they have provided their own, previously confidential, data to populate sections of the data resources.

Box F.2.Members of the EBI Industry Programme (December 2009)

- AstraZeneca
- Bayer Schering Pharma
- Boehringer Ingelheim
- Galderma
- GlaxoSmithKline
- Eli Lilly and Company
- F. Hoffmann-La Roche
- Johnson & Johnson Pharmaceutical Research & Development
- Merck Serono S.A.
- Nestlé Research Centre
- Orion Pharma
- Philips Research
- Pfizer Ltd
- Syngenta
- Sanofi-Aventis
- Unilever

In future, we see our interactions with industrial partners growing even stronger, as the flood of data continues to rise and the need of industry to outsource and utilise public resources stimulates pre-competitive research collaborations, use of open-source software, and standards development. Through participation in efforts such as the IMI, an FP7 Joint Initiative fostering collaborative projects between the European pharmaceutical industry, academia, patient organisations and regulatory agencies, and the Pistoia Alliance, an initiative to provide an open foundation of data standards, ontologies and web-services to streamline the Pharmaceutical Drug Discovery workflow, EMBL aims to influence, support and encourage this transition.

In the coming years, if this trend towards increasingly outsourced data and tools is to succeed, new organisational entities must be created to provide services for industry. These services will be based mostly on public data provided by the EBI, but must be 'secure' (including aspects of patent disclosure and business risk), and must allow companies to combine private with public data. This mode of operation will require a carefully considered strategy of EMBL involvement, as such services could be provided on a commercial basis.

3.2.2 Interaction between beamline scientists and bio-industries

The X-ray-based structural biology services offered at EMBL Grenoble and Hamburg provide vital infrastructure not only for academic users but also for the European pharmaceutical industry, which forms one of the biggest user communities. At present, industrial use of beamlines for structural biology is largely part of their medical research aiming to understand the structure, dynamics, and function of biological macromolecules for drug discovery and design.

The Instrumentation groups at EMBL Grenoble and Hamburg develop state-of-the-art instruments for X-ray crystallography and biological small angle X-ray scattering (SAXS) experiments at synchrotron beamlines and col-

laborate with European industry to make them available to the scientific community. In Grenoble, for example, partnerships have been established with the companies Maatel, Bruker-ASC, Cedrat Industries, Hampton Research and Molecular Dimensions. During the past decade, more than 30 beamlines worldwide have been equipped with diffractometers and robotic sample changers developed at EMBL and a key contribution was made to define the SPINE European standard equipment to store and transport frozen crystals from laboratories to European synchrotrons. Our developments benefit European industry by generating commercial activity in high-end products and by helping maintain their technical know-how at a high standard.

Industry also directly benefits from the instrumentation development at EMBL through the use of our equipment during their experiments at DESY, the ESRF and indeed most other European synchrotron facilities. Major developments in automated crystal harvesting have recently started in collaboration with the companies Maatel and SwissCI. This development, if realised in combination with the advanced facilities proposed at the EMBL@PETRA3 and the ESRF sites will provide a fully integrated crystallography service with high-throughput screening capacity that should reduce sample processing costs and times.

3.2.3 Interaction between Core Facility staff and bio-industries

Early access to state-of-the-art equipment and technologies is key for the provision of advanced services in the context of EMBL's Core Facilities and interactions with industry partners are an essential aspect of horizon scanning.

Since its foundation, the Advanced Light Microscopy Facility (ALMF, Section D.3.1.6) has collaborated with leading microscopy manufacturers to further advance the technology and applications of commercially available equipment. This allows industry to improve their latest microscopy technologies and test it on concrete biological applications in the ALMF and receive expert user feedback from EMBL's scientists and visitors. This collaboration has lead to several new developments and products. In the past ten years the ALMF has partnered with 24 companies and currently has 15 ongoing collaborations with manufacturers. Amongst others, these collaborations have been very successful in the field of high-content screening microscopy together with Leica and Olympus, and in the development of liquid dispensing robotics for microscopy experiments in collaboration with ProCellcare.

The Genomics Core Facility (Section D.3.1.1) interacts regularly with manufacturers in the areas of new sequencing technologies and microarray-based tools. Moreover, due to its established reputation, companies frequently approach the Genomics Core Facility for beta-testing of and early access to new solutions.

3.3 The ATC Corporate Partnership Programme

The Advanced Training Centre Corporate Partnership Programme (Section I.2) is designed to create and enhance long-term, mutually beneficial relationships between EMBL and corporate partners. It aims to connect interested private sector companies with the latest developments in the molecular life sciences and to provide support for training young scientists. The financial contributions of the 15 corporate members, as of January 2010, are used to sponsor conferences and training programmes and to provide fellowships for young scientists. In return the CPP provides bio-industries with a platform to reach researchers and high-end users of their products, learn about the needs of the scientific communities with respect to technology and material and remain informed about new trends in the life sciences, both within EMBL and in the wider community.

4. Communication and Public Relations

EMBL's core communications unit, the Office of Information and Public Affairs (OIPA) in Heidelberg, addresses various target audiences including the wider scientific community, journalists, students, teachers, decision-makers and the public. The office manages the maintenance of the website, media work, production of publications and promotional material, offers graphics support for scientists and runs an outreach programme including visits to the Laboratory and exhibits at conferences and career fairs. Additionally, OIPA advises and oversees communications projects carried out in the outstations. Today, all but one of the outstations also have dedicated outreach representatives. This improves internal communication and helps the sites to respond to the growing need for local outreach activities. Because of its strong service mission, a large proportion of the communication efforts of EMBL-EBI are

focused on users. The EBI's outreach programme is therefore coordinated by a dedicated Outreach and Training Team, in close collaboration with OIPA.

One of the major communication challenges EMBL faces has arisen from rapid changes in the way information is received, processed and used in our society. New online, digital, interactive and social media are in the process of revolutionising communications. In an ongoing manner, EMBL will address these changes, explore ways to exploit new media for effective science communication and continue to fine-tune its outreach activities to diverse audiences and outlets.

Corporate design

To ensure a coherent, institutional identity across all sites, EMBL revised its corporate design including the logo, the colour scheme and its layout guidelines during the last Programme. All printed publications, promotional material and websites now share a uniform, recognizable "look" and standardised templates are available for official documents, letters and presentations.

EMBL websites

The websites are EMBL's most visible window to the outside world and are also used as a central information resource for EMBL staff. The internet sites need to continuously evolve to provide online access to key information, to be as interactive as possible and to remain up-to-date.

In 2009, EMBL launched completely redesigned, centrally managed websites for Heidelberg, Hamburg, Grenoble and Monterotondo. A content-management system was implemented, which allows editors distributed across the Laboratory to update and publish content continuously. To highlight the individual strengths of the different EMBL sites, each outstation provides specific local information on their homepage and defines both the menu and contents of its site. A new intranet portal has been introduced to facilitate the sharing of information among EMBL employees. On average, www.embl.de, the Heidelberg site, is visited about 6000 times a day by guests and the Heidelberg intranet about 800 times by EMBL staff. Because of its enormous user base and web presence, EMBL-EBI requires a dedicated web interface with the public and its users, as described in the context of the Bioinformatics Services (Section D.1.).

Future plans for the EMBL websites include:

- The development of a media gallery with images, videos and live streaming of events.
- Further development of the intranet into an internal communication and networking tool. This will include the implementation of forum pages to foster exchange between collaborating groups.
- Continued training to teach web editors how to use the content-management system and the provision of templates for the creation of websites about EMBL-related projects.

Publications and publicity material

OIPA regularly produces a range of documents, brochures and flyers for various audiences. The official EMBL Annual Report has continuously evolved over the past years and is now a modern publication that informs funders and decision-makers about the state of the Laboratory and showcases EMBL's scientists and research highlights. OIPA also annually produces the popular Research at a Glance booklet, which includes a brief summary of all EMBL research projects and is aimed predominantly at a scientific audience. Other brochures such as the EMBL overview brochure, the brochure for PhD applicants, site-specific brochures and leaflets on career opportunities, are updated regularly. The EMBL etcetera newsletter, a key publication for internal communication that reaches 5000 staff members and alumni in print and online, was redesigned in 2009.

In future:

- We will continue to improve EMBL's publications and select appropriate formats for different audiences.
- We will develop an electronic version of the EMBL etcetera newsletter.
- We will extend our graphics and publication capacities to comply with the increasing demand for graphics support and publicity material.

Figure F.5. A collection of press clippings from international media outlets 2007-2009

Media work

The EMBL press office coordinates all media relations between scientists and journalists and communicates important news in research, technology development, training and other EMBL activities to more than 1000 media contacts worldwide. The press office also supports the EBI Outreach and Training Team to promote EMBL-EBI activities with the media. During the past years, we have expanded our range of communications instruments to include video material and new formats for press releases, i.e. picture releases and press releases targeted at the specialist press, the latter being particularly important to inform EMBL-EBI's user community through specialist bioinformatics outlets. The new formats have proven very successful, with videos receiving over 100,000 hits on public portals such as YouTube and scientific images produced at EMBL being featured in newspapers, news websites and popular science magazines. In the period 2007-2009 we issued 90 press releases, which resulted in more than 3700 press clippings, half of which appeared in print media, including broadsheets, popular science magazines and technical journals, and half of which were covered by research-themed websites.

In future, to keep pace with the fast-changing media landscape and adapt to developments such as the reduction in print publishing and the shift towards online communication, EMBL will:

- Tailor our press work even more to specific target audiences.
- Focus more on working with regional media.
- Complement our press work by testing selected new media tools, such as social and interactive media.
- Develop support materials that the media can use with little or no modification.

Public relations activities

In the past years we have increased our presence at both major scientific conferences and careers fairs to raise awareness of EMBL's research and services, and to attract the best applicants to join the Laboratory. We also organize about 20 EMBL visits per year for school and university students, almost all from member states. In Heidelberg, we take part in the local Girls' Day initiative, when schoolgirls are invited to visit EMBL to learn about life in a research laboratory, and organise activities at the "Long Night of Science" festivals in Heidelberg and Hamburg, when research institutes and universities open their doors to the public till late in the night in what are remarkably popular events.

Summary of our future plans:

- We will ensure even better harmonisation of different EMBL programmes and activities and a uniform public image.
- We will target our activities to reach the scientific community, journalists, students, teachers, decision-makers, the public, the next generation of scientists, early-stage scientists, facility users and future employees collaborating with the outstations and respective in-house teams.
- We will continue to offer communication training to predoctoral fellows and EMBL staff and will provide support and advice to project administrators of EU-funded research projects.
- We will further develop EMBL's websites to include a media gallery with images, videos and live streaming of events and an intranet serving as an internal communication and networking tool.
- We will extend our graphics and publication capacities to comply with the increasing demand for graphics support and publicity material. In addition, we will continue to improve EMBL's publications and select appropriate formats for different audiences, including the creation of an electronic version of the EMBL etcetera newsletter.
- We will tailor our press work even more to specific target audiences, for example by developing support materials that the media can use with little or no modification, and we will complement our press work by testing selected new media tools.
- We will address the changes in the way information is received, processed and used in our society by exploring ways to exploit new media for effective science communication and continue to fine-tune our outreach activities to diverse audiences and outlets.

5. Science and Society

Towards the end of the 20th century the life sciences saw a number of critical developments, such as recombinant DNA and the birth of biotechnology, the market penetration of the first genetically modified crops, the exhilarating expectations associated with the Human Genome Project and the cloning of the first mammal, that had the potential to bring about radical change in many areas of life in society. These developments changed the role and the place of biomedical science in society and made the need for a new dialogue between scientists and the general public apparent. To stimulate this dialogue and provide a forum to disseminate information about molecular biology and its impact on society, in 1998 EMBL launched its Science and Society Programme. The Programme also provides a platform for EMBL scientists to be exposed to and address the evolving societal issues and ethical debates that accompany the harnessing of new knowledge and technologies growing out of basic research.

Since its inception, the EMBL Science and Society Programme has become an integral part of scientific life in the Laboratory. A variety of activities and events organised at EMBL bring together members of the life science community, scholars of other disciplines and members of the public, for discussion and communication extending beyond professional boundaries. While all of the EMBL Science and Society activities are open to the public, some are primarily targeted at the EMBL research community. An example of the latter is the EMBL Forum series of interdisciplinary seminars, with topics as diverse as building scientific capacity in developing countries, patenting of biological information, the relationship of science and art and the impact of meditation on our brain. EMBL Forum lectures take place approximately once every other month and attract an audience ranging between 50 and 300 people.

Every year the EMBL Science and Society Programme, together with EMBO, organises a Science and Society conference in Heidelberg. Over the past 11 years these meetings have been a great success attracting both local and international audiences to discuss a range of different topics including systems and synthetic biology and their social implications, food sustainability and plant science and genes, brain and behaviour.

In 2001, EMBL, the German Cancer Research Centre (DKFZ) and the Medical Faculty of the University of Heidelberg launched a public lecture series, 'Heidelberg Forum – Biosciences and Society', specifically aimed at enlightening and engaging local audiences. These lectures take place in a public auditorium and reach audiences between 150 and 200 attendees. The Heidelberg Forum benefits from generous sponsorship by a local philanthropic organisation, The Lautenschläger Foundation.

Since 2007, EMBL's Science and Society Programme has expanded beyond the main Laboratory in Heidelberg and now regularly organises activities in the outstations, including an annual half-day thematic symposium organised in Cambridge with the help of EMBL-EBI PhD students and seminars in Monterotondo and in Hamburg. Past topics of these events include 'Promises and pitfalls of public-private partnerships', 'Scientific and social aspects of ageing' and 'The personal genome'.

In 2008, together with partners from the European School of Molecular Medicine (Milan) and the Harvard School of Government (Boston), EMBL organised an EU-sponsored pilot project entitled the European Science and Society Summer School (E4S) in Heidelberg. The Summer School brought together 30 PhD students and postdocs from the life sciences, the humanities, and the social sciences with senior researchers and scholars from a range of different fields for a week of intense interdisciplinary exchange. Depending on the availability of external funding we intend to repeat this highly successful event and plan to build on and further develop all described Science and Society activities in future.

G. Member state relations

EMBL is a European ambassador for the life sciences and fulfils an important function in integrating biomedical research activities throughout Europe and building up international networks of collaboration and exchange. We devote much time and effort to carrying out this task because we believe that our integration and coordination efforts are a critical contribution to the future of research in our member states. In the coming years, more resources will be required to strengthen EMBL's scientific administration in order to allow us to increase the member states' participation in EMBL through partnerships, collaborations, a growing network of EMBL alumni and to enable us to extend EMBL membership to include more of Europe.

Member states and associate member states. EMBL maintains close relationships and interaction with its 20 member states and associate member state Australia**.** To further integrate European research efforts in the life sciences, the participation of as many European countries as possible in EMBL is desirable. We intend to continue discussions with the newer EU member states to encourage them to join EMBL and participate in and benefit from our unique network. Since 2008, Australia has been taking active advantage of associate membership to strengthen its life science links to Europe and to plan the adoption of EMBL practices in some of its national activities. It thus provides both a valuable partner for EMBL, a recent example for other prospective members of how membership can benefit the national community and a good test case of the expansion of EMBL's activities beyond Europe.

EMBL partnerships. EMBL engages in formal partnerships with selected member state institutions of similar vision, international orientation, overlapping research interests and complementary expertise. These partnerships involve the transfer of relevant aspects of the EMBL model, such as international recruitment, continuous staff turnover or the best ways to organise core facilities, to the member state institutes. The aim is to create an interlinked system of excellent institutions and thus enhance the development of the molecular life sciences across Europe.

C**ollaborations with institutions in non-member states.** The increasing globalisation of research will offer new opportunities and challenges. Although EMBL's formalised overseas relations are still few in number, the de-

mand for interaction with EMBL as a means of building bridges to the European life science community is growing and provides us with an opportunity to help foster closer links between Europe and the rest of the world. The main focus of EMBL's international activities to date has been in Asia, mainly Japan and India, as well as in Australia.

EMBL alumni. Europe's knowledge society will increasingly depend on a skilled work force and the excellently trained successive generations of young EMBL scientists make a tangible contribution to fulfilling this goal. More than 80% of EMBL's 4600 alumni have returned to EMBL's member states. They enrich their national systems with their experience and training, representing a major return on member state investment. Alumni act as local ambassadors for EMBL and, building on the many contacts they are able to make during their time at EMBL and the maintenance of these links, establish a network across Europe.

1. EMBL member states and associate member states

1.1. 20th EMBL member state: Luxembourg

Since it was founded in 1974, the number of countries participating in EMBL has doubled. In 2007, EMBL Council welcomed the Grand-Duchy of Luxembourg as the 20th member state. Luxembourg's activities in the molecular life sciences range from basic research in genetics, molecular biology and biomedicine to more applied areas such as cancer research and the development of health technologies. With the foundation of its University in 2003, Luxembourg is actively contributing to the education and training of Europe's scientific elite. Being part of EMBL's international network will provide a boost for Luxembourg's research in the growing field of molecular biology and will integrate its researchers even better into the European scientific community. In return, Luxembourg will contribute to EMBL's various activities bringing in complementary strengths and technical expertise. A new collaboration between EMBL's Genome Biology Unit, the Institute for Systems Biology in Seattle, USA and the newly founded Centre of Systems Biology Luxembourg was established in 2009 to address fundamental challenges in personalized medicine.

1.2. New member states

In general all European countries are welcome to join EMBL. Discussions are currently underway with the Czech Republic, Poland, Slovak Republic and Malta. In the next programme period, we wish to initiate proactive measures to facilitate the joining of the new EU member states, by for example organising mutual visits to explore opportunities for scientific collaborations and participation in EMBL training programmes. Some countries that are currently not members, for example Poland and Turkey, already send substantial numbers of young researchers to participate in EMBL activities, demonstrating the need for training that young scientists from these countries have and the desire of the scientists to interact with EMBL. Their participation is made possible through external funding obtained from the EC and other funding organisations, but full participation in EMBL's programmes requires membership, which has to be a mutual future goal. If approved by Council, our intention will be to make a small contribution to enable visits of scientists from candidate countries to learn about the opportunities that EMBL offers its member states, meet EMBL scientists and visit the Core Facilities. Such visits are organized on a regular basis for member state visitors who often also ask for advice in setting up new institutions or facilities locally or in the context of the ESFRI research infrastructure projects.

1.3. First associate member state: Australia

The joining of Australia as the first associate member is an important step in expanding EMBL's network beyond Europe. The agreement establishing EMBL foresees formal links with non-European countries and in 2003 EMBL Council agreed in principle to the associate membership scheme. In July 2007 Council welcomed Australia's request to become the first associate member state of EMBL and in October 2007 an Instrument of Cooperation was signed by the Australian government and EMBL. The associate membership started officially in January 2008 and will initially last for seven years.

Australia has become a central player in the landscape of molecular biology. With its special expertise, for example in the fields of medical epidemiology and stem cell research, it forms an excellent complement to EMBL's focus on basic research in molecular biology. Collaborations and exchange have been established between the five European sites of EMBL and leading Australian research institutions. Four major Australian research universities (Monash University, the University of Western Australia, the University of Queensland, the University of Sydney) and the Commonwealth Scientific and Industrial Research Organisation (CSIRO) made financial contributions to, and are now actively engaged in implementing, Australia's associate membership. Australia contributes to the diverse activities at EMBL by sending early-career scientists to join EMBL through a faculty development programme, while EMBL shares with Australian institutions its world-renowned expertise in researcher training and research infrastructure development. Australia is utilising associate membership to facilitate interaction between the Australian and European molecular biology communities and to encourage the exchange of scientists between the two continents. Combining their complementary expertise will benefit not only Australian science but also research at EMBL and in a broader context in Europe.

For the implementation of the associate membership the Australian government appointed Monash University as lead organisation. In October 2008 an agreement was signed between the four Australian universities and CSIRO with the approval of EMBL and the Australian government on the implementation of the associate membership in Australia. In April 2009 Richard Larkins, the former Vice Chancellor of Research at Monash University, was appointed as first chair of EMBL Australia Council.

Several visits have led to the establishment of collaborations in bioinformatics, regenerative medicine and plant biology. The Director General met with the Australian Research Minister Senator Kim Carr on several occasions to discuss how the relationship can be developed to mutual benefit.

The associate membership made provisions for two faculty positions at EMBL that would be funded by Australia for a five-year period followed by four years at an Australian research organisation. In addition, Australia is providing funding for pre- and postdoctoral fellows at EMBL.

In May 2009 eight million AUD were awarded by the Australian government to finance the initiation of an EMBL partnership laboratory in Australia. The plan is to begin with a Unit located at Monash University and then to expand to a network of laboratories located at the other partner universities in Brisbane, Sydney and Perth in the future. The concept is to copy the successful EMBL organisational model into the Australian national context in order to help promote internationalisation and rejuvenation of the life science community.

Extension of Australia's associate membership will require a review by EMBL Council and the Australian government that will take place in 2012/13.

2. Partnerships with member state institutions

2.1. EMBL partnerships

EMBL enters into partnerships with selected member state institutions that exhibit international vision and orientation and a desire to implement the EMBL organisational model. This scheme was first introduced to Council and the member states in the EMBL Strategic Forward Look 2006-2015 and has turned out to be extremely successful. The partnerships can operate either locally at one of the five EMBL sites or further afield to foster research collaborations at the institutional level in scientific areas overlapping with and complementary to those being pursued within EMBL. Their aim is to create an interlinked system of excellent institutions and thus enhance the development of the molecular life sciences across Europe and beyond.

EMBL contributes its know-how and experience to setting up the partner laboratories but cannot transfer funds to the partnership institutions. This means that the establishment of each partnership requires substantial efforts at the regional, national and international level, depending on the nature of the partnership, to raise the necessary funds. The partnership scheme has nevertheless been very popular and is being followed with interest in the member states. It has already led to the establishment of excellent research institutions that have been able to form

close links to EMBL, engage in scientific collaborations and benefit from an increased visibility and interactions with EMBL in their countries. EMBL scientists benefit from the partnerships by becoming better acquainted with the work going on in these research organisations, by setting up new scientific collaborations and by being involved in scientific reviews or the recruitment of young scientists. EMBL Partnerships are reviewed after 5 years to ensure that they are a scientific success and have implemented key aspects of the 'EMBL model', including regular external review, international recruitment and staff turnover.

- The first partnership was established in 2003 and currently eight partnerships are in place:
- Partnership for Structural Biology (PSB), Partner: ESRF, ILL, Institut de Biologie Structurale, UVHCI, Grenoble, France.
- Unit for Virus Host Cell Interactions (UVHCI), Unité Mixte Internationale (UMI 3265), Partner : CNRS, Université Joseph Fourier, Grenoble, France.
- Molecular Medicine Partnership Unit (MMPU), Partner: University Heidelberg Medical Faculty, Heidelberg, Germany.
- Partnership in Marine Biology with the Sars International Centre for Molecular Marine Biology, Bergen, Norway.
- Agreement with DESY to use Synchrotron Radiation for Life Sciences, Hamburg, Germany.
- Centre for Genomic Regulation - EMBL Partnership Unit for Systems Biology, Barcelona, Spain.
- Nordic EMBL Partnership for Molecular Medicine, Partner: University Oslo, Norway, University Umeå, Sweden, University Helsinki, Finland.
- EMBL Australia Partnership Laboratory, Partner: Monash University, Melbourne, University of Sydney, Sydney, University of Western Australia, Perth, University of Queensland, Brisbane, Australia.

Four partnerships were established before 2007: PSB, MMPU, Sars, DESY. Since 2007 the PSB and MMPU agreements were extended following external review and four new partnerships were established: the UVHCI, the CRG-EMBL Partnership Unit for Systems Biology, the Nordic EMBL Partnership for Molecular Medicine, and the Australian EMBL Partnership Laboratory.

2.2. Status of individual partnerships: past achievements and future development

2.2.1 Partnership for Structural Biology, Grenoble, France

The PSB was established in 2002 to provide a unique environment for state-of-the-art integrated structural biology. The partners were EMBL, ESRF, ILL and the Institut de Biologie Structurale (IBS), which is supported by the French CEA and CNRS. The Carl-Ivar-Brändén-Building opened in 2006 and **today houses 100 people** affiliated to the PSB partners and includes laboratory space and several of the PSB technical platforms. In 2008 a fifth partner, the newly established Unit of Virus Host Cell Interactions (UVHCI, Section 2.2.2) joined the PSB. 14 technical platforms are operated and jointly used by the partners; EMBL contributes the ESPRIT platform for the expression of difficult proteins, eukaryotic cell expression, high throughput crystallization and a share of the electron microscopy platform. In 2009 the PSB had 300 active scientists, 70 PhD students, and 60 postdoctoral scientists that published more than 150 publications. In the immediate future, the ESRF-ILL site will be partially redeveloped and the plans include a new on-site building to accommodate the IBS.

2.2.2 Unit of Virus Host Cell Interactions

The Unit of Virus Host Cell Interactions (UVHCI, UMI 3265 UJF-EMBL-CNRS) was created in January 2007 to develop the collaboration between EMBL Grenoble, the Université Joseph Fourier, and the CNRS. The UVHCI is located on the ESRF-ILL research campus in Grenoble and is directed for an initial five-year period by Stephen Cusack, Head of EMBL Grenoble. It is the second of Unité Mixte Internationale to be created in France out of 14 worldwide.

The objective of the UVHCI is to pursue internationally leading research in structural and molecular biology focused on virus-host cell interactions. Interdisciplinary research will cover virus structure, virus assembly and maturation, virus-host cell interactions, host and virus gene-expression mechanisms, cell biology of infected cells, innate immunity, as well as anti-pathogen drug design. New methods and technical platforms for structural biology will be developed, for example in the areas of high-throughput expression and crystallisation, synchrotron X-ray and neutron diffraction methods and instrumentation, Laboratory Information Management Systems (LIMS), and electron microscopy.

2.2.3 Molecular Medicine Partnership Unit, Heidelberg, Germany

In January 2002, the Medical Faculty of the University of Heidelberg and the European Molecular Biology Laboratory (EMBL) established the Molecular Medicine Partnership Unit that is staffed by both institutions and co-directed by Andreas Kulozik from Heidelberg University and Matthias Hentze, EMBL's Associate Director.

Research at the MMPU aims to contribute to the molecular understanding of the basis of common human diseases, its applications in diagnosis and therapy, and feedback from 'bedside medicine' into the research laboratory. The MMPU consists currently of five collaborative University/EMBL teams and the research themes include the diseases of childhood, particularly blood disorders and childhood cancers, diseases of adolesence such as chronic lung diseases, in particular cystic fibrosis, and others such as defective cholesterol regulation, hemochromatosis and colorectal cancer.

An external scientific review was carried out in 2009 and strongly recommended that a common facility should be provided to house the collaborating groups. This will be followed up in 2010/2011 when the Medical Faculty plans to provide space for the MMPU activities in a dedicated research facility.

2.2.4 Partnership with the Sars International Centre for Molecular Marine Biology, Bergen, Norway

The partnership with the Sars Centre has enabled collaborations in marine molecular biology involving the use of marine model organisms by scientists at EMBL. In return EMBL has taken an active role in advising and evaluating the development and progress of the Sars Centre, has participated in the recruitment of new faculty members and provided advice on setting up a PhD programme.

2.2.5 Agreement with DESY to use synchrotron radiation for life science applications, Hamburg, Germany

The agreement confirmed the mutual commitment for building beamlines for life sciences applications on the new PETRA-III synchrotron, which went into operation in 2009. The first EMBL-operated beamline, for small angle X-ray scattering, will be available towards the end of 2010 and further beamlines for protein crystallography will follow in 2011. In pursuit of this challenging project the collaboration with DESY has developed very well and has recently benefitted from the strong support of the new DESY director.

A new Centre for Structural Systems Biology will be established on the DESY campus in Hamburg, Germany, bringing together more than 10 research organisations, universities and EMBL. The German federal and regional governments are providing €73 Mio to build a state-of-the-art facility on the campus of the German synchrotron in Hamburg; start of operations is planned for 2012/2013. The centre will use integrated structural biology for research related to infectious diseases and will form a platform through which the potential use in biology of the X-ray Free Electron Laser, also under construction at DESY and scheduled for 2014, can be investigated.

2.2.6 CRG-EMBL Partnership for Systems Biology, Barcelona, Spain

The partnership between EMBL and the Centre for Genomic Regulation (CRG) was formed in 2007 to advance the understanding of complex biological systems. By combining EMBL's expertise in computational biology with the CRG's know-how in specific areas of genomics and proteomics the unit hopes to better understand some of the key aspects of biology relevant to human health.

Luis Serrano at the CRG, who was previously joint coordinator of the EMBL Structural and Computational Biology Unit in Heidelberg, heads the Unit. The research draws on the expertise of various scientific disciplines and spans

systems analysis across the entire range from molecules to cells. Five multidisciplinary groups, that are funded by the Spanish ministry and have fixed–term contracts following the EMBL model, form the unit.

In late 2009, three papers were published back-to-back in Science by EMBL and CRG scientists, who had collaborated over several years to produce the first comprehensive picture of a minimal cell. Their work was based on an extensive quantitative study of the biology of the bacterium *Mycoplasma pneumoniae*, that has a very simplified genome. A possible extension of the partnership to include formal links between the CRG bioinformatics groups and EMBL-EBI is currently under discussion with the Spanish research ministry.

2.2.7 Nordic EMBL Partnership for Molecular Medicine, Norway, Sweden, Finland

In October 2007, an agreement was signed between the University of Oslo, Umeå University, the University of Helsinki and EMBL to establish three Nodes of a Nordic EMBL Partnership for Molecular MedicineThe aim of the partnership is to combine expertise in basic and clinical research to promote translational research. The partnership facilitates scientific collaboration, access to scientific infrastructure, including databases, facilities and instrumentation, as well as to services and training activities provided by the partners.

In addition to their partnership with EMBL, the individual nordic research centres engage in collaborations with other national partners, including research and public health institutes, hospitals and research councils, with the aim of establishing an extensive nordic network for molecular medicine.

The objectives of the partnership are first, to facilitate and institutionalise scientific exchange and support in areas of common interest, or where one partner has a recognised expertise, which can be shared for the benefit of the other partners. Second, the partnership nodes will implement aspects of EMBL's administrative model, such as international recruitment, staff turnover, external scientific review and joint PhD and postdoctoral programmes involving all three nodes.

The partnership will focus on research in molecular medicine which will build on complementary strengths in all four partner institutes: EMBL's recognised research strength in areas such as molecular, cellular and developmental biology, bioinformatics and structural biology; Norway's strength in molecular mechanisms of disease; Sweden's strength in microbial pathogenicity and molecular infection medicine; and Finland's strength in integrated clinical and basic biomedical research coupled with population-wide genetic epidemiology and well characterised medical databases.

The Norwegian Node of the Nordic EMBL Partnership for Molecular Medicine is the Centre for Molecular Medicine Norway (NCMM). It is hosted by the University of Oslo in partnership with the Regional Health Authority SouthEast, and with links to the other Universities and Regional Health Authorities in Norway. The Swedish Node of the Nordic EMBL Partnership for Molecular Medicine is the Laboratory for Molecular Infection Medicine Sweden (MIMS). MIMS is established within the Umeå Centre for Microbial Research (UCMR) and is affiliated with both the Faculty of Medicine and the Faculty of Science and Technology and closely connected to the university hospital (Norrlands University Hospital). Support for MIMS was granted by the Swedish Research Council with the aim of strengthening Swedish research and enhance the dynamics in the field of molecular medicine, partly by promoting the career opportunities for young scientists.

The Finnish Node of the Nordic EMBL Partnership for Molecular Medicine is the Institute for Molecular Medicine Finland (FIMM). FIMM is a joint venture of the University of Helsinki, the Hospital District of Helsinki and Uusimaa (HUS), and the National Public Health Institute (KTL).

EMBL provided support in the international recruitment of young group leaders to each node and 15 faculty positions had been filled by the end of 2009. The number is expected to increase to over 20 by the end of 2010. A first conference was organized in early 2010 to bring together scientists from all three nodes and EMBL. Regular exchanges are planned for the future including the development of a joint Nordic partnership PhD Programme involving the three nodes and other activities.

2.2.8 EMBL Australia Partnership Laboratory

As part of the implementation of the associate membership, Australia has agreed to set up a partnership laboratory network. Its hub is to be located at Monash University in Melbourne and nodes are planned at the Universities of

Sydney, Queensland, and Western Australia. A mirror of some of EBI's data resources is being constructed at the University of Queensland. Initial funding has been obtained from the Australian government.

2.2.9 Proposed future partnerships

Informal discussions are ongoing in several EMBL member states and we expect a small number of additional partnerships to be established in the next five-year period. Since each partnership requires significant effort from the senior staff of EMBL, the Laboratory's Scientific Advisory Committee has recommended that EMBL avoid developing this programme beyond the level that can be managed.

2.3. Other formal collaborations with institutions in member states

- EMBL-EBI and the Wellcome Trust Sanger Institute are collaboratively responsible for Ensembl, which delivers high quality annotated genome databases for vertebrate species. The two institutes also collaborate on other DNA sequence-based data resources.
- The GKSS Research Centre, Geesthacht and EMBL Hamburg have joined their efforts to construct and operate the new small angle X-ray scattering beamline (BioSAXS) at the PETRA-III synchrotron.
- The German Cancer Research Center (DKFZ), the University of Heidelberg and EMBL jointly operate the Chemical Biology Core Facility at EMBL Heidelberg.
- The Imperial College-Magdi Yacoub Institute, Harwell, UK and EMBL Monterotondo are collaborating to use their complementary experience and expertise in cardiac morphogenesis, congenital heart disease, cardiac transplantation and transgenic models.

3. Collaborations with institutions in non-member states

EMBL has established many collaborations with other research institutions world-wide, mostly in the context of externally funded projects, but also some that are longer term collaborations for specific purposes. The prototypes of such activities are the various international collaborations and exchange agreements between EMBL-EBI and many other data resource providers throughout the world. More recently, EMBL entered into an agreement with the Indian National Institute of Immunology and ESRF to operate Beamline 14 at the ESRF synchrotron in Grenoble and agreed with the Japanese National Institute of Basic Biology to foster exchange of scientists and collaborations between Europe and Japan.

3.1. Indian National Institute of Immunology

EMBL has agreed to take over the operation of Beamline 14 (BM14) at the ESRF in Grenoble, France, in the framework of the Joint Structural Biology Group established with the ESRF in 1992 and in collaboration with the Indian National Institute of Immunology, from January 2010. The agreement is for a five-year period and its aim is to provide access to BM14 to users from the EMBL member states and India. BM14 is a dedicated macromolecular crystallography bending magnet beamline with a particularly impressive record in structure determination using single or multiple wavelength anomalous dispersion (SAD/MAD) method. From 2001 to 2009 it was owned by the UK Medical Research Council (MRC) and operated in partnership with EMBL. In 2009 the NII joined the collaboration and at the end of 2009 the MRC ended its operation of the beamline and handed it over to the EMBL-ESRF-NII consortium.

NII is an autonomous institution under the government Department for Biotechnology (DBT) and is representing a consortium of macromolecular crystallography communities in India who have come together, with funding from DBT, in order to enter the BM14 beamline partnership. Their objective is to enrich biomedical research by providing access to the synchrotron beamline for the Indian macromolecular crystallography community.

3.2. National Institute for Basic Biology (NIBB), Okazaki, Japan

The NIBB is part of the Japanese National Institutes for Natural Sciences and is one of Japan's leading research institutes. EMBL and NIBB started a very fruitful collaboration in 2005 with the aim of promoting joint research activities, inviting faculty members and researchers for exchange visits between NIBB and EMBL for lectures, workshops, conferences and other academic activities, to exchange graduate students for collaborative projects and visits and to exchange information in fields of interest to both organisation. A series of nine joint symposia have since been organised in the areas of developmental biology, light microscopy imaging, mouse biology, structural biology, epigenetics, systems biology and functional genomics. Many scientists have visited NIBB from EMBL and vice versa, most of them generously supported by NIBB funding. The initial five-year agreement was extended until 2015 in January 2010.

4. EMBL Alumni

One of EMBL's biggest assets – as well as one of the major benefits for the member states – is its distributed network of alumni. EMBL was founded as a centre of excellence to attract leading junior scientists from across the world to Europe and provide them with advanced training and ideal conditions to pursue research in molecular biology. Part of EMBL's mission is to make the skills and expertise that scientists acquire during their stay available to its member states. Helping to train top scientists and to create lasting networks and collaborations is a unique service that EMBL offers to its member states and European science. The turnover system, based on time-limited appointments, ensures that researchers stay a maximum of nine years at EMBL and more than 80% of alumni subsequently take up positions in academia or industry in member states or associate member states. These alumni act as EMBL ambassadors; they share their experiences of working in a unique international and interdisciplinary environment and often go on to successfully implement aspects of the EMBL model in member state institutions.

Since its foundation in 1974, EMBL has produced more than 4600 alumni, comprising approximately 3100 scientific, 1100 technical and 400 administrative staff at all stages of their careers. More than half of the former scientific staff now occupy senior positions. For example, 21 alumni have become Directors of institutes of the Max Planck Society in Germany.

In 1999, the EMBL Alumni Association was founded to build and maintain lasting connections with former staff. The Association's goal is to support EMBL in disseminating its expertise, scientific culture and organisational model to the member states for the benefit of the entire community. It provides a platform for EMBL staff and alumni to interact and benefit from one another's knowledge, expertise and resources, and through which they can forge life-long links with the organisation and with each other. The Alumni Association includes scientists, support staff and members of the administration from all five EMBL sites as well as from EMBO. It currently has 1600 members and during the past three years there has been a 60% increase in registration.

The EMBL Alumni Association pursues four broad aims:

1. Organising EMBL's global network of alumni
2. Engaging alumni and staff (=future alumni)
3. Fostering and intensifying relationships between alumni and staff
4. Developing an income stream devoted to supporting alumni activities.

The main focus of the past five years has been on the first two aims. Much groundwork has been done to develop a database that allows EMBL alumni to update their records online. The database is an ongoing task and will help to increase and maintain the usefulness of alumni records over time. To engage EMBL alumni and staff, the Alumni Association has concentrated on the development of an attractive profile, regular communication with alumni, as well as the provision of useful services.

The EMBL Alumni Association supports meetings organised by its local chapters to reach out to local alumni communities. Local chapters form a support structure for former EMBL employees who sometimes face problems when relocating into a new community. Who better to provide advice than alumni who have already gone through

the same process? Between 2007 and 2009, local chapter meetings were organised in Spain, Portugal, the UK, the USA, Germany and Greece. At these meetings alumni are updated on recent EMBL developments in the areas of research, services and training and have a chance to engage in informal networking. Regular local chapter meetings have proven a useful tool to reach alumni and will be continued as an initiative in future, with meetings in Ireland, Finland and France already planned for 2010-2011. For additional networking opportunities, the Alumni Association has created alumni groups on popular online communities such as LinkedIn and Facebook, and both the Association's website and the EMBL newsletter feature news about alumni. Moreover, regular staff-alumni events across all EMBL sites encourage interaction between alumni and current EMBL staff.

Intensifying the interactions between former and current staff has helped to raise awareness and generate support among alumni for different EMBL initiatives, such as the European Learning Laboratory for the Life Sciences (ELLS), the Science & Society Programme and conferences and courses. Moreover, services the Alumni Association provides in the area of career development allow fellows and junior members of staff to tap the extensive knowledge of the alumni network to plan and advance their own careers. In 2008, a careers wiki – accessible to staff and alumni – was created to provide general and country-specific information that facilitates integration into a new scientific or national community. A second career development initiative is the John Kendrew Young Scientist Award that, initially thanks to the financial support of the EMBL Pensioner's Association, was also launched in 2008. It aims to foster and support young scientists through early recognition of their achievements in science and/or science communication, and serves as an inspiration to aspiring EMBL pre- and postdocs.

For the period 2012-2016, the Alumni Association will focus on its third aim, to foster and intensify relationships with staff and alumni at two levels: firstly through the provision of an EMBL alumni portal through which staff and alumni can interact and share information; and secondly through increased local chapter meetings and regular staff-alumni reunions. Another major focus will be resource development to support alumni activities. Its first fundraising campaign has already been launched to ensure the continuation of the John Kendrew Award.

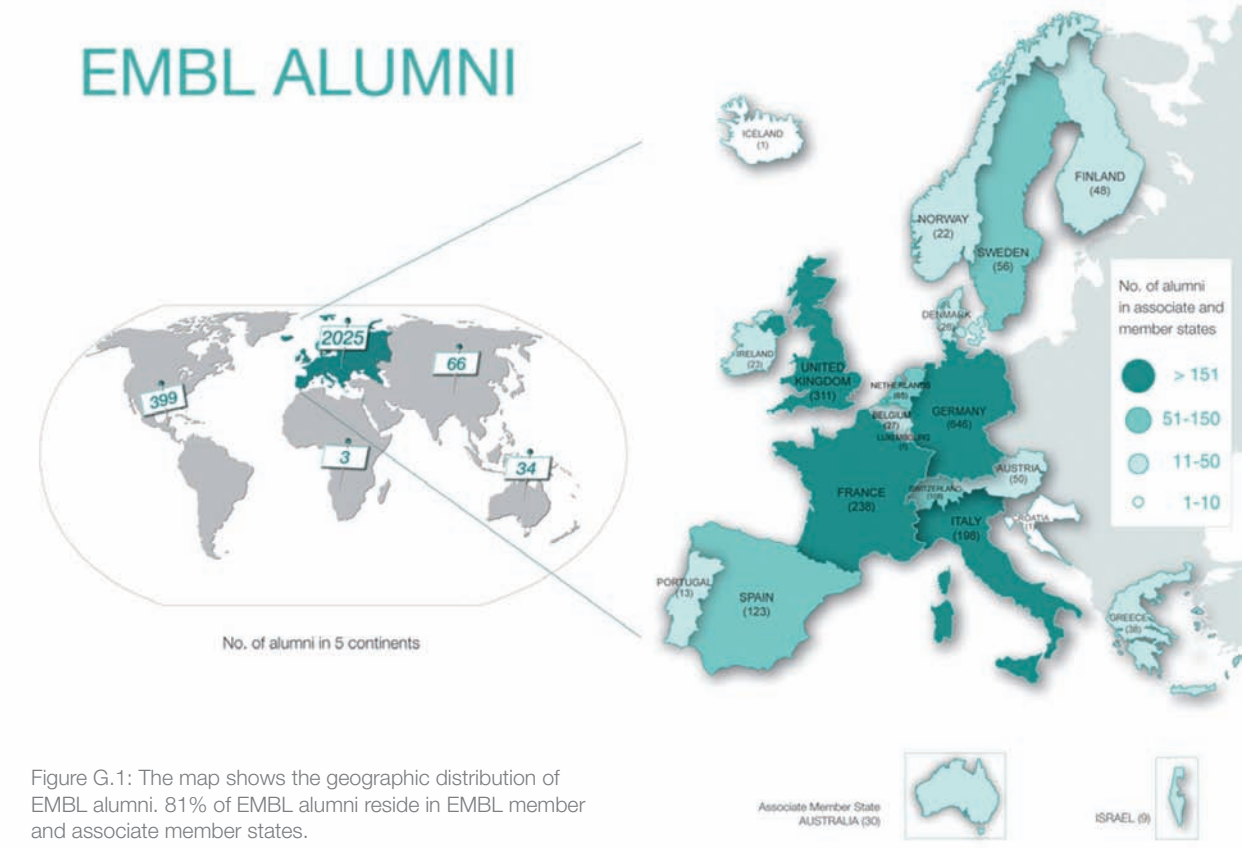


Figure G.1: The map shows the geographic distribution of EMBL alumni. 81% of EMBL alumni reside in EMBL member and associate member states.

H. European integration

As Europe's member states increasingly unite and integrate and as life science research diversifies into an interdisciplinary endeavour driven by international collaboration, a new level of coordination and integration is required that goes beyond the bottom-up networking activities and partnerships described above. A more integrated approach at the institutional level will allow Europe to address the major challenges in the life sciences and make the best use of the limited resources available. EMBL, as Europe's intergovernmental laboratory for molecular biology, has a central and strategic position in the European research landscape. It plays a unique integrative role in the life sciences and contributes in many ways to the development and integration of research in Europe as well as in forging links between Europe and the rest of the world

EMBL maintains close links with policy makers in the institutions of the European Union and other political or scientific interest groups established by the EU member states and associated countries such as the European Strategic Forum on Research Infrastructures (ESFRI). We either interact directly with policy makers or act in the framework of other organisations to which we belong. EMBL is a founding member of EIROforum, an organisation that was established by the seven major inter-governmental research infrastructure organisations in Europe to promote research excellence and provide input into science policy discussions. The European Life Sciences Forum (ELSF) and the Initiative for Science in Europe (ISE) were founded by EMBL together with EMBO, FEBS and, particularly in the case of ISE, many other European-scale scientific organisations, to provide a voice to the scientific community in European science policy and a means for wide consultation. With its central position in the European research landscape and several decades of experience in running an international organisation that integrates a considerable variety of research-related activities, EMBL has become a valuable motor for this level of integration.

1. European research infrastructures

A fairly recent aspect of European integration, but one that EMBL dedicates considerable effort to, is the proposed renewal and expansion of European research infrastructures (RIs) to support the needs of the biomedical science community. Infrastructure forms a core aspect of centres of excellence for research. RIs contribute to scientists' education and training and feed back on technology development and transfer, both important driving forces of economic and social progress. Thus, in order for European life science research to be competitive on an international scale it must have access to world-class research infrastructure. Because of their scope and the international nature of their user communities, RIs are a prime example of the benefits of integrated European action through which duplication can be avoided and coordination achieved.

EMBL plays an extremely active role in the ESFRI process, coordinating two and participating in another five of the ten ESFRI Biomedical Science projects. Our input to these projects is in part scientific, but our unique experience in the organisation and management of European-scale life science research infrastructure means that we can also provide invaluable input on the governance and organisational models for new research infrastructures in the biomedical science field. EMBL is the only intergovernmental research organisation in the life sciences in Europe. It has over 35 years of experience in running international research facilities that provide services and training. This not only bestows on EMBL the responsibility to organise integrative, infrastructure-based activities in certain fields of the life sciences but also puts it in the position of being able to help others handle this challenge. EMBL has the requisite expertise and experience to take a leading role in the areas of biomolecular data resources, biomedical imaging and aspects of structural biology, in which it has successfully provided services to European scientists for many decades. EMBL is therefore committed to play a central role in future RI development in these areas. EMBL may not only provide advice but in some other cases may also, dependent upon the decision of EMBL Council, serve as the organisational umbrella for a new research infrastructure. Over the past three years we have been actively involved in discussions with the member states and the European Commission to review the usefulness of existing organisations and the possible structure of new and upgraded major European RIs. Funding these infrastructures remains the biggest challenge and using an existing organisation such as EMBL, rather than constructing a new one, may be a cost-effective solution for some of the biomedical research infrastructures.

1.1 EMBL's participation in ESFRI Projects

The European Strategy Forum on Research Infrastructures (ESFRI) was set up by European Member States (the EU plus associated countries) and the European Commission as a platform to explore new initiatives for the development of European research infrastructures. All the ESFRI participants are convinced that if Europe is to become the world's leading knowledge economy, then its scientists must be provided with access to a new generation of world-class pan-European research infrastructures. After extensive discussion, ESFRI selected specific research infrastructures for inclusion in its European Roadmap for Research Infrastructures, first published in October 2006 and updated in 2008. 44 projects that broadly cover diverse fields of research were chosen for support from European Member States and the EC and their implementation is planned for the next decade. 10 out of the 44 projects are in the biomedical sciences. EMBL, as an intergovernmental research organisation, has the right to submit proposals directly to ESFRI after approval by EMBL Council. Reflecting its central role in European life science research, EMBL is currently involved in seven of the ten biomedical science projects. It coordinates two, ELIXIR and Euro-BioImaging, and participates in INSTRUCT, BBMRI, Infrafrontier, EU-OPENSCREEN and EMBRC. The ESFRI projects are pan-European and are intended to complement and coexist with national infrastructures. ESFRI provides an unprecedented opportunity for Europe to take an integrated approach to developing the best research infrastructures in the world, if the challenges of reconciling the interests of all the participating member states can be met and sufficient funding can be provided not only to build but also to run and maintain these research infrastructures in the future.

The projects on the first ESFRI roadmap are now completing their EU FP7-funded preparatory phase, whose aim is to decide on the scope of the infrastructure and establish a consortium of member states that will provide financial support for its construction and operation. This requires a detailed strategic plan including the scientific justification as well as decisions on the legal, management and governance structures of the infrastructure and the location of the site or sites at which it will be constructed. Securing stable funding is the greatest challenge and it is likely

that solutions will involve contributions from various sources, including not only the member states but also the EC, national research funding organisations and charities.

All biomedical research infrastructures will be distributed across several sites, reflecting both the distributed nature of biomedical research in general and the very large size of the biomedical user community. Almost by definition, distributed RIs will have to become international organisations if they are to work well because this is the only model that will allow free transfer of personnel and funds between the different sites at which the infrastructure operates. In the context of ELIXIR, a legal and governance structure was developed for an EMBL "special project" that would allow ELIXIR to become an integral part of EMBL, using EMBL's legal personality. This model was approved in principle by EMBL Council in summer 2009 and can now be offered as an umbrella structure to help in the initial phase of setting up new biomedical science RI projects. A 'special project' enables a new research infrastructure to make use of EMBL's legal and governance structures, while maintaining an independent budget. This can provide decisive help to the distributed biomedical sciences infrastructures on their way to becoming independent intergovernmental organisations, a process that generally takes years. There is no reason why this organisational model should not be maintained in the longer term and indeed this currently seems a likely outcome for ELIXIR and possibly a small number of other projects.

1.2. ESFRI Projects coordinated by EMBL

The two ESFRI projects that are coordinated by EMBL (i.e. where EMBL has taken on the scientific and organisational leadership on behalf of the consortium) represent areas in which EMBL is a leader in the field, in bioinformatics (ELIXIR) and in advanced light microscopy (Euro-BioImaging). In INSTRUCT EMBL has been one of the six core partners from the beginning, building on EMBL's strength in structural biology research and service provision at the Outstations in Hamburg and Grenoble and in the Structural and Computational Biology Unit in Heidelberg. EMBL's participation in the other four ESFRI BMS projects is mainly through EMBL-EBI and is driven by the need to link the data resources developed for biobanks (BBMRI), mouse biology (Infrafrontier), chemical biology (EU-Openscreen) and marine molecular biology (EMBRC) to the core biomolecular data resources provided by EMBL-EBI. All ESFRI BMS projects have requested and benefitted from EMBL's advice on organizational, legal and governance issues.

1.2.1. European research infrastructure for biological information: ELIXIR

The EMBL Programme 2007-2011 identified the rapid growth of, and growing demand for, biological data resources as the biggest challenge to EMBL and particularly EMBL-EBI. The EMBL member states agreed to a considerable increase in funding for that period, which has helped to alleviate some of the problems. However the development of new technologies, for example in the area of DNA sequencing, means that the EBI, Europe's major provider of biomolecular data resources and with 4 million database hits every day also Europe's most widely used infrastructure for the molecular life sciences, still faces significant problems. The ESFRI process has provided a new impetus to the search for solutions in this area, widely regarded as being of the highest importance for a healthy European biomedical research community. A planned European research infrastructure for biological information, built around EMBL-EBI as a core, was included in the first ESFRI roadmap in 2006. Financial support from EU FP7 was provided to enable the preparation of a detailed strategic plan for this research infrastructure, now called ELIXIR. The project is coordinated by Janet Thornton, the Director of EMBL-EBI. The aim is to plan within the community and with the member states how a pan-European sustainable research infrastructure for biological information can be established. EMBL and its member states play a critical role in this process and an EMBL Council Working Group was established to examine and evaluate possible outcomes; any solution that involves EMBL will require approval by EMBL Council.

ELIXIR's mission is to construct and operate a sustainable infrastructure for biological information in Europe to support life science research and its translation to medicine, the environment, the bio-industries and society and to provide the link to global life science data resources. It will permit the integration and interoperability of the diverse, heterogeneous information that is essential to generate and utilise biological knowledge. Various data resources located throughout Europe will be organised as a hub and several nodes that will be connected and integrated (see Figure H.1). ELIXIR will encompass the necessary computational infrastructure, both hardware and software, to store and organise biological data in a manner that enables rapid access and search through a sophisticated, user-

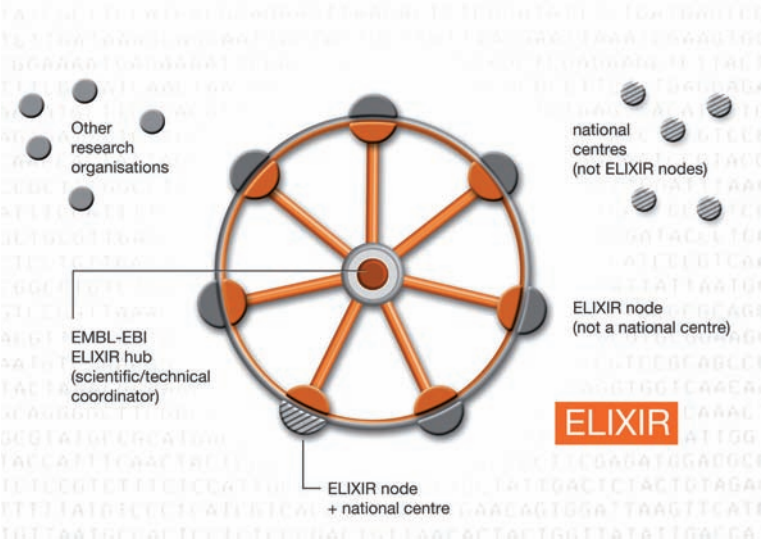


Figure H.1: The proposed ELIXIR structure

friendly portal. ELIXIR will not only manage data generated in Europe but also serve as the European component of the major international collaborations that collect, curate and annotate biological information world-wide. It will manage processes for (i) the integration of novel data-types (ii) supporting interoperability of analytical tools and (iii) developing standards and ontologies for biological information. ELIXIR is also needed to support the other ESFRI biomedical research infrastructure projects, not only by helping them with the organisation and distribution of the data they produce but also by providing them with access to the core biomolecular information and computational tools that they will need to operate their infrastructures and conduct their research.

The ELIXIR consortium currently encompasses 13 European countries and consists of 32 organisations, including 2 ministries, 14 research funding organisations and 16 associated scientific organisations. Although the preparatory phase referred to above is scheduled to close at the end of 2010, the implementation phase of ELIXIR will have already begun by then because several countries have already provided funding for the implementation of aspects of ELIXIR and we expect others to follow.

At the scientific and technical level the infrastructure will be distributed and will consist of a Hub, hosted by EMBL-EBI, and a number of nodes (Figure H.1).

The Hub will be responsible for holding the core data collections and enabling the development, interconnection and integration of nodes into a European-wide distributed infrastructure. The ELIXIR Database Providers Survey of 208 biological databases at 97 different institutions has shown that core biomolecular resources in Europe include those for nucleotide sequences and genomes, protein sequences, protein structures, protein-protein interactions and expression data. These data resources are mainly based at EMBL-EBI, though several involve major collaborations with partners elsewhere in Europe (e.g. the Uniprot database has grown up over many years as a close collaboration with the Swiss Institute of Bioinformatics), which makes EMBL-EBI currently the only viable option for an ELIXIR hub. In contrast, specialist data resources are widely distributed and are complementary to the core databases.

1.2.1.1 Funding of the ELIXIR hub

The following explains why we think that is it critical that the EMBL member states take responsibility for the running costs of the European bioinformatics hub at EBI. EMBL-EBI is already the European hub for biological information, being one of only two such centres worldwide and because of its strong links to the user communities in the member states is in the best position to maintain the leading position of Europe in the coming age of Information Biology. Continuing to develop and provide many of the largest and most frequently used data resources at EBI will facilitate the management and integration of Europe's other data resources and will certainly be cheaper for the

member states. The distributed structure of the nodes should leverage more funding at national level for national activities but will be unlikely to solve the problem of funding the central, integrative activities that are essential to the success of ELIXIR.

If EMBL-EBI provides the ELIXIR hub, it will enable the coordinated development of nodes in the member state countries, which can actively participate in a well-organised integrated infrastructure. Otherwise the infrastructure in Europe will either remain centralised (being almost entirely at EBI), or will continue to grow without coordination and end up being fragmented and therefore non-competitive. The role of coordination (both technical and leadership) is logically best taken on by an international organisation, representing all the member states, rather than a single member state.

An alternative would be to set up ELIXIR as a new international consortium, which would have the advantage that EMBL member states could opt in or out. The ELIXIR member state consortium could be larger than just the EMBL member states. An independent structure would result in more visibility for ELIXIR, but would be more challenging to manage and much more expensive because it would duplicate the EMBL structures already in place. Overall integration would be more difficult or impossible to achieve without a strong, centrally-funded hub even if it is possible that the political acceptance of ELIXIR may be higher if it were to create a new structure rather than to build on an existing successful one.

The development of this distributed research infrastructure for biological information in Europe will continue for many years. Although there is considerable will to establish and fund nodes in institutes around Europe there is no mechanism to build international agreement to fund the essential hub. The only practical location for this hub is at EMBL-EBI and without it the ELIXIR project will fail. It is our view that the future of European biomedicine hangs upon the efficient curatorship and exploitation of the scientific discoveries produced in ever-increasing volume by molecular biologists today. If we fail to provide the structures required to collate, interpret and disseminate these discoveries in a maximally informative way, key issues in future life sciences, in particular a systems level understanding of biology, will be impossible to achieve. The stumbling block for the creation of this critical multinational resource is the creation and funding of the hub element. EMBL's mission, set out in the Laboratory Agreement, states that:

> *"The Laboratory shall promote co-operation among European States in fundamental research, in the development of advanced instrumentation and in advanced teaching in molecular biology as well as in other areas of research essentially related thereto, and to this end shall concentrate its activities on work not normally or easily carried out in national institutions."*

It is our belief that the provision and funding of the hub is precisely the kind of activity that EMBL was created to carry out and we are therefore asking the EMBL member states to support the additional activities at EMBL-EBI in the next five year programme that will allow us to deal with the increasing quantity and diversity of data, take on some new activities and engage with the nodes that will be established in the member states to facilitate their operation and integration. Funding the hub by the EMBL member states would give ELIXIR a head start and allow the project, and the individual countries, to focus on creating or in some cases maintaining excellent national nodes.

Many of the activities and data resources that would be concentrated in the hub are already provided and funded by EMBL. The additional functionalities that would be required are; maintaining links and standards with nodes, providing coordination and technical support for the ELIXIR nodes, and increased computing and storage facilities would need to be funded in addition to EMBL's current budget. This investment, if agreed on, would provide a firm and lasting foundation for the ELIXIR project and ensure that European life science has the necessary tools and structures to build upon the discoveries of its scientists.

1.2.2 European research infrastructure for medical and biological imaging: Euro-BioImaging

A second ESFRI project, Euro-BioImaging, is coordinated by Jan Ellenberg, Head of the Cell Biology and Biophysics Unit, EMBL Heidelberg, together with the European Institute for Biomedical Imaging Research (EIBIR), Austria. EMBL's unique expertise, cutting-edge infrastructure and long-standing tradition in biological imaging, together with EIBIR's complementary expertise in medical imaging applications will address the imaging requirements of both the basic and medical imaging communities by deploying imaging infrastructure in a coordinated and harmonized manner and thus address the fragmentation of such efforts currently present in Europe.

As has been described throughout this document (see Section C.2.1) biological imaging at various levels and scales has become a crucial tool for life science research and will continue to gain importance in future as new technologies mature and come into widespread use. This makes world-class infrastructure for imaging studies an indispensible requirement for the European scientific community. EMBL, particularly through its Advanced Light Microscopy facility, has been very active in the development and use of light microscopy technology and in the generation of networks of users and providers of this technology throughout Europe. It is therefore a natural step for EMBL to coordinate the biological imaging aspect of this Euro-Bioimaging.

Euro-BioImaging is at an early stage and will enter its three-year preparatory phase in 2011 if funding is awarded by the EC. The aim of Euro-BioImaging is to provide access and training to imaging technologies across the full scale of biological and medical applications, from molecule to patient. The objectives of the preparatory phase of Euro-BioImaging will be the definition of the needs of the biomedical imaging user communities, the preparation of plans for construction and operation for the future research infrastructure, the definition of the legal and governmental framework, the commitment of member states to secure sustainable funding for Euro-BioImaging, and the integration of Euro-BioImaging in the European and global research infrastructure landscape. At the moment, the plans for the future infrastructure foresee the organization of strongly interlinked nodes, each focused on complementary imaging technologies addressing different aspects of biology, physiology and pathophysiology. Nodes will be newly constructed or undergo major upgrades in order to devote a large part of their capacity to external users. In planning, Euro-BioImaging will profit from the wide network of European imaging facilities, reprenting almost every ESFRI member state. Many of these are already participants in the European Light Microscopy Initiative, in whose formation EMBL played a major role.

Advanced light microscopy nodes will provide access to a broad range of imaging methods to European scientists and will become reference centres for the imaging communities in many ESFRI member states. *In vivo* molecular imaging nodes will address the newest developments in optical tomography, multi-modal molecular imaging in animal models and the development and testing of new imaging probes. For medical imaging, platforms for European collaboration in clinical imaging trials, in population imaging and in innovation and training of image guided therapies will be developed. Innovative technology nodes will provide access to imaging technologies which are not easily accessible to the broader imaging community, because they are new and still under development, require knowledge and training from experts, are not yet commercially available or are simply too expensive for individual institutes.

A planned Infrastructure for Biomedical Imaging Data Storage and Analysis will provide methodologies and protocols for image data management and quantitative data processing in close collaboration with ELIXIR. Furthermore, Euro-BioImaging plans to create a comprehensive and coordinated training platform for biomedical imaging at graduate and post-graduate level. Finally, user access to infrastructures, services and resources of Euro-BioImaging will be harmonized. Societal impact will be increased by fostering collaboration between industrial, regional, national and European authorities, and multidisciplinary scientists involved in the field of imaging. Relations of the Euro-BioImaging infrastructure to other European biomedical science infrastructures will be extensive.

1.3 Other ESFRI Projects with EMBL participation

1.3.1 Integrated Structural Biology Research Infrastructure for Europe (INSTRUCT)

Structural biology is the second main area of service provision at EMBL and thousands of users and visitors are served every year in Hamburg, Grenoble and Heidelberg (see Section D.2). Recent developments have shown that the integration of technology and service platforms in structural biology that produce data at scales ranging from Ångstroms to 10s of nanometres is an essential tool for the analysis of large, functional biomolecular complexes. This concept was introduced by EMBL through joint efforts across EMBL-Hamburg, Grenoble and Heidelberg and was then further enhanced in the Grenoble Partnership for Structural Biology. Building on the availability of the PETRA-III beamlines and the X-ray Free Electron Laser will also provide the intellectual framework for the new Centre for Structural Systems Biology in Hamburg and other INSTRUCT nodes such as the planned research complex at the DIAMOND synchrotron in the UK.

INSTRUCT will link information obtained by the major structural biology methods to state-of-the-art cell biology data to provide a detailed dynamic picture of key cellular processes. Major technology advances, from high-

throughput methods in protein production, through NMR and X-ray crystallography to electron microscopy mean that major investment in infrastructure will be required to maintain European competitiveness. EMBL is participating in the 3 year INSTRUCT preparatory phase project, that involves creating a pan-European structure comprising a number of centres across Europe. The project is coordinated by Prof. David Stuart, Oxford University, UK. A major aim of the preparatory phase is to establish the financial and legal mechanisms by which major funding bodies can work together to provide a coherent European infrastructure with broad user access. EMBL's future involvement and exact role will depend on the general strategy that INSTRUCT adopts after the preparatory phase project.

1.3.2 Biobanking and Biomolecular Resources Research Infrastructure (BBMRI)

BBMRI will sustainably secure access to biological resources required for health-related research and development intended to improve the prevention, diagnosis and treatment of disease and to promote the health of the citizens of Europe. BBMRI will collaborate with EMBL-EBI and in future with ELIXIR on the organisation and integration of its data resources.

1.3.3 Infrafrontier

EMBL Monterotondo is located on the campus that hosts the European headquarters of the European Mutant Mouse Archive (EMMA) and has been involved in a number of large EU-funded consortia such as EUMODIC, whose goal is to begin to build a European research infrastructure for mouse biology, specifically for establishing standards and protocols as well as accessible infrastructure for phenotyping mutant mice. In the context of ESFRI, Infrafrontier is now planning to organise two complementary and linked infrastructure networks for large-scale and comprehensive phenotyping and archiving of mouse models. Infrafrontier will serve the European genetics and biomedical research community to the ultimate benefit of human health. Infrafrontier is developing a strategy, in collaboration with EMBL-EBI, for managing and archiving data and linking it to the core biomolecular data resources at the EBI.

1.3.4 European Marine Biology Resource Centre (EMBRC)

This project was included in the second ESFRI roadmap. The consortium initially involves 10 leading marine biology stations and EMBL, which will provide expertise in genomics research, imaging and bioinformatics. The rationale for establishing a European Marine Biological Resource Centre (EMBRC) is to respond to the ever increasing demand for high quality provision of marine organisms to serve as models for research and development in diverse domains ranging from global change to bio-fuel production. By connecting existing marine biological research institutes located across the continent into a united organisation with common goals and strategies, Europe can take a leading role in structuring and promoting marine biological research and technical development. EMBRC and EMBL-EBI are together developing a strategy for managing the data EMBRC will produce and linking it to existing data resources.

1.3.5 European Infrastructure of Open Screening (EU–OPENSCREEN)

One of the scientific areas that has recently been expanded at EMBL is chemical biology. In 2004, a Core Facility with the capacity to screen for the effects of small molecules in biological assays was established in Heidelberg (see Section D.3.1.5) and in 2009 the EBI expanded its computational chemical biology effort by establishing ChEMBL as a chemogenomics resource. The second-generation ESFRI project EU-OPENSCREEN aims to build a pan-European infrastructure for Chemical Biology. The preparatory phase project will cover all aspects of chemical biology from high-throughput screening, with a dedicated compound library, to assay development, synthetic chemistry, bioprofiling and *in vivo* studies, as well as a central database that will be developed in close collaboration with EMBL-EBI and education programmes.

2. Relations with the institutions of the European Union

EMBL maintains tight links to the European Commission (EC) on various levels and has received significant funding from Framework Programmes 5, 6 and 7. The EC is in fact the most important external funder of EMBL's research. In 2007 and in 2008 42% of the external funding - €17 million - came from the EU. This funding is obtained competitively and allows EMBL to engage in new research projects, to provide user access to its infrastructure and facilities and to extend its training programmes to scientists from non-member states. To date EMBL researchers have coordinated 29 and participated in over 130 projects funded by Framework Programmes 6 and 7. The funding has fostered a large number of collaborations, which bring together scientists from all over Europe.

In 1995 the relationship with the European Commission was put on a formal basis by the signing of an Administrative Arrangement that not only outlines common goals within the European Research Area but also provides a framework for cooperation, including mutual consultation and exchange of information. Based on this arrangement the EC has observer status at EMBL Council meetings, which provides an effective mechanism for regular exchanges. Numerous interactions, consultations, and collaborations take place regularly and have led to a fruitful exchange between the EC and EMBL.

In large part, the success of the partnership between EMBL and the EC during the last five years is to the credit of the former EU Commissioner for Science and Research Janez Potočnik, who created an environment of mutual trust, building on initial contacts made by his predecessor, Commissioner Philippe Busquin. The renewal of the Administrative Arrangement is planned in 2010, with the signing of a Memorandum of Understanding by EMBL and the new Commissioner for Research and Innovation, Máire Geogheghan-Quinn.

Interactions with the European Parliament and the Council of Europe have so far mainly taken place in the context of EIROforum or ISE (see below) but in the process of gathering political support for the new biomedical research infrastructures we envisage the intensification of our interactions with members of the European Parliament and the Council of Europe.

3. EIROforum: past achievements, future plans

EIROforum is a partnership of the seven largest inter-governmental research organisations that operate research infrastructure in Europe: CERN, EFDA-JET, EMBL, ESA, ESO, ESRF and ILL. Each organisation is funded by a number of member states and their combined annual budgets for science are around €1.5 billion. EIROforum was established in 2002 and the organisations share their expertise in the areas of basic research and the management of large, international infrastructures, facilities and research programmes. The mission of EIROforum is to support European science in reaching its full potential through both their individual efforts and by launching common initiatives. Within its collaboration, EIROforum concentrates on certain fields that are of common interest and importance to all EIROforum Organisations and covers topics such as international relations, human resources, instrumentation, outreach and education and information technology. Joint interactions with the European Union and its institutions, national governments and industry secure a consolidated approach at European level. EMBL held the chair of EIROforum for the second time from July 2009 to June 2010 EMBL.

The field of research infrastructure is of enormous importance for European competitiveness and is one of the top priorities both of the member states, the EC and EIROforum. The research infrastructures operated by the EIROforum organisations are unique in Europe and in some cases in the world. Having managed large international infrastructures for several decades, the EIROs have enormous collective experience and regularly act as advisors in the field of research infrastructure in Europe and beyond. To give one recent example, during the preparation of the new regulation on a European Research Infrastructure Consortium (ERIC) by the EC as part of the ESFRI process an EIROforum representative acted as a member of the advisory committee to the Commission, the Sounding Board.

EIROforum is very willing to share its experience with those who will be responsible for constructing and operating new infrastructures. In 2010, EIROforum has therefore published a position paper on European research

infrastructures that describes its experience and provides input to other European stakeholders in research infrastructures.

EIROforum has a long tradition of outreach and education activities and organises or participates in several initiatives to support science teachers, students and journalists. EIROforum has, for example, established a series of teacher training courses. It will continue the publication of Science in School, a European journal for science teachers founded in 2005 that reports on new findings in the seven EIROforum organisations and successfully inspires science teaching. EIROforum is also a strong and active supporter of the complementary EU Contest for Young Scientists (EUCYS) organised by the European Commission as a competition between European school pupils.

Knowledge sharing is part of EIROforum's philosophy and is practiced in areas of European added value. EIROforum is regularly represented at conferences and science fora such as the Euroscience Open Forum (ESOF) or the American Association for the Advancement of Science (AAAS) to present and discuss its knowledge and philosophy with the public. In the field of science and technology EIROforum organised a conference on technology transfer in November 2009 at EMBL-Heidelberg that allowed exchange of knowledge and best practices across disciplines and between scientists, science funders, technology transfer professionals and industry.

An intensive collaboration and regular exchange exists between EIROforum and the European Commission. In 2003 this collaboration was formalised by the signing of a Statement of Intent. This was in 2010 to update the original agreement, incorporate new areas of cooperation and to reinforce the mutual commitment to a fruitful collaboration. The new agreement was signed for the EC by the Commissioner for Research and Innovation Máire Geogheghan-Quinn.

There are numerous examples of this fruitful exchange of information and expertise. In 2007, EIROforum published a formal response to the Commission's Green Paper 'ERA: Towards the future' describing six pillars of the European Research Area. A consultation with the stakeholders in Europe followed and EIROforum provided feedback emphasizing the importance of continuous efforts to build the European Research Area. EIROforum endorsed the Commission's proposal and offered support to constructing and upgrading major European research infrastructures to make European science competitive on a global scale.

4. The Initiative for Science in Europe

EMBL is also a founder-member of the Initiative for Science in Europe (ISE), an organisation of 'grassroots' European-scale scientific societies and organisations, that contributes to debates on, for example, the European Research Council, European Research Infrastructures or scientific career structure by organising meetings and debates involving scientists and science policy-makers. Examples of such activities in 2010 are the co-organisation of ECRI (European Conference on Research Infrastructures) with the EC and the Spanish EU presidency and the organisation of a meeting on the future of the European Research Council (ERC) at which representatives of all stakeholders spoke and discussed.

I. Administration

1. EMBL Administration

1.1. Current Status

With over 1530 staff from more than 70 nations distributed over five sites in four host countries and with continuous staff turnover, EMBL is an increasingly complex organisation. To address the ensuing challenges and to allow the research staff to focus entirely on their scientific activities, efficient, up-to-date administrative systems and services are required. EMBL employs less than 200 full-time administrative staff in all support areas, ranging from caretakers and gardeners to senior management staff. They are mainly located at the headquarters in Heidelberg with a small number of administrators providing local support at the outstations. 83 of the staff work for the 'classical' administration e.g. Human Resources, Finance and Purchase, SAP Team, etc. and a further 112 provide in-house services such as gardeners, drivers, canteen staff, childcare, the Laboratory of Animal Resources (LAR) and security.

Towards the end of the current Indicative Scheme a new Administrative Director was appointed. During the eight-year tenure of the departing Administrative Director significant improvements in staff relations and in the efficiency of the administrative services were achieved. These positive developments were reflected in a survey that was conducted in 2009 to assess both the efficiency of EMBL's administrative processes across all EMBL sites and the staff's satisfaction with these processes. The survey also identified areas in which adjustments have become necessary. Further to the excellent feedback constructive recommendations were made and the first improvements have been put in place.

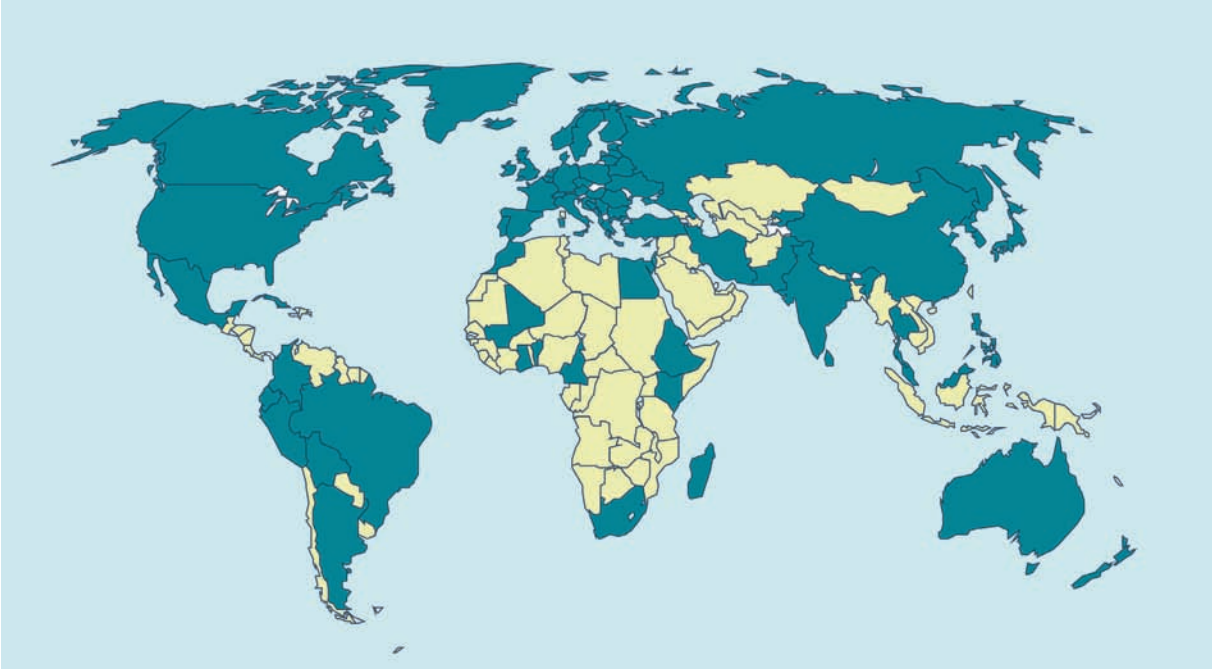


Figure I.1. Nationalities represented by EMBL staff (status 2009).

1.1.1 Human Resources

In 2010 the Personnel Department was renamed Human Resources and in an effort to improve the services provided, first steps toward the installation of an e-recruitment system were made. Recruitment staff were also organised into small, Unit-specific teams in order to better meet the specific needs of the different parts of EMBL.

In 2007 Human Resources and EICAT developed a General Training & Development Programme, which complements EMBL's activities in scientific training and specialist courses. It provides all staff with the opportunity to acquire important transferable skills relevant to their work and their career beyond EMBL, including management, communication, IT and language skills. 643 staff members participated in 133 courses in 2008.

The past years have seen drastic increases in healthcare costs across Europe, so that contributions to the Health Insurance Scheme had to be increased from 8.6% to 10.8% in 2010, with €1 million of the health insurance fund to be contributed towards the increased employer contributions in 2010 and 2011. Reductions in certain benefits were also made. Finally, amendments were made to ensure that the provisions for EMBL fellows are in line with the recommendations of the European Commission ('European Charter for researchers – The Code of Conduct for the Recruitment of Researchers'). Additionally, an update of the EMBL pension scheme is still under discussion in an EMBL Council working group and is to be decided before 2012.

1.1.2 Finance / SAP

As have many other international organisations, EMBL recently changed its accounting standards from cash-based accounting to accrual-based accounting. The major changes are the recognition of purchased goods and services at the time of receipt as well as the depreciation of fixed assets over their estimated life. This accounting practice provides for more meaningful and transparent financial reporting.

During the current Indicative Scheme EMBL will have completed the implementation and integration of the SAP-modules Finance, Controlling, Human Resources, Procurement, Asset Management and Travel. This presents a step change in the quality of information available to staff. Now real-time reports can be accessed by scientific staff anywhere in the world enabling them to make better-informed decisions about budgets, recruitment, staffing and purchasing.

1.1.3 Estate Management

The current Indicative Scheme has seen the construction of a number of buildings including the EMBL-EBI extension (East Wing) in 2007, a project supported by a significant investment from the Wellcome Trust as well as funding from the BBSRC, the MRC and the member states.

Generous contributions from the host country, Germany, the Klaus Tschira Foundation, the federal state of Baden-Württemberg and the member states made the construction of the EMBL Advanced Training Centre (ATC) with the new canteen possible. The ATC was inaugurated on 9 March 2010. Most of the Administration moved into the new building and the space vacated will be refurbished over time to serve scientific needs.

The main Laboratory underwent extensive refurbishment in all areas, including level 5 and 6 of the main building, level 2 annex building, level 3 main lab 1B, the steam boiler system, the 20 kV high voltage circuit system, the Kinderhaus and a complete roof renovation of all buildings on the campus including the cultivation of 'green roofs' as well as numerous individual lab renovations.

In Hamburg, the construction of building 48E related to the new beamlines at Petra-III was completed in 2010, incorporating three floors of laboratories, a robot room, offices and a seminar room, again with generous support from the host country and the member states.

Figure: I.2. Aerial view of EMBL Heidelberg in 2010

1.2. Strategic Objectives 2012-2016

Other sections of this document outline a variety of ambitious goals in the areas of life science research, services and training. On the basis of these plans EMBL's Administration has undertaken a strategic reassessment of its structure and services to identify administrative needs and ways to meet them in the future. As in its other activities, EMBL strives to be a model for Scientific Administration. In order to achieve this EMBL's Administration will pursue the following goal:

Excel in providing non-bureaucratic, timely and efficient administrative services through unobtrusive, flexible and effective processes and systems in support of an international, progressive and nurturing working environment with highest quality staff.

This ambition can only be achieved with a suitable administrative structure in place. The four pillars on which EMBL bases its future administrative strategy are organisation, processes, staff and communication.

Organisation

A complex and rapidly changing environment like EMBL requires an integrative organisational structure, which is both adaptable and proactive. In order to achieve this we will take a 'plug and play' approach that relies on deploying a limited workforce and infrastructure to address a broad range of administrative tasks with a large degree of flexibility.

Processes

Ideally, administrative processes at EMBL are results oriented. They address desired quality targets with a minimum of resources, their validity is frequently re-evaluated and they are flexibly readjusted towards changing objectives. An integral part of these processes and their coordination is the organisation's operational system (i.e. SAP for EMBL), which also addresses informational needs and actively supports all decision-making.

Staff

Being based on a system of continuous turnover, recruitment and retention of highly qualified staff is one of EMBL's priorities. Marketing and competitive conditions of employment are key factors in attracting suitable candidates. These instruments should be supplemented by comprehensive programmes aimed at raising the professional profile of EMBL staff during their employment.

Communication

Underpinning all of the above approaches are efficient operational networks ranging from internal communication and stakeholder consultation over adequate and smooth collaboration with local and national authorities to active engagement with other international organisations in matters of governance, infrastructure and know-how.

1.2.1 'Plug-and-Play'

Create an environment able to process any organisational requirement, manage all potential changes and actively strive for and pursue improvements.

In order to ensure continuity of operations and to safeguard the legitimacy, reputation and credibility of the organisation, EMBL needs to identify and assess key strategic, operational and financial risks for the organisation through a comprehensive Risk Assessment. This exercise will conclude in a Risk Management Concept integrating Disaster Recovery and Business Continuity plans.

Subsequently a gap analysis will identify inconsistencies and outdated organisational substructures as well as weaknesses in communication and interrelations between sites and implement an integrative structure generating synergies across all locations.

In this context, space preparation and allocation to organisational units play an essential role. With the completion of the ATC, office space has become available in the main Laboratory in Heidelberg and will be reorganised and refurbished over the coming years. The additional space will mainly be used to enhance the scientific infrastructure and to improve the organisation of the research and support Units in Heidelberg.

Large increases in commodity prices over the past few years, and the anticipation of further increases in future, mean that the development of a wide-ranging energy management strategy has become necessary.

Finally, EMBL-EBI's anticipated role as the central hub of ELIXIR will require an expansion of workspace providing offices, training rooms and meeting facilities for roughly 200 additional staff.

1.2.2 Processes

Establish non-bureaucratic, effective and transparent operational processes and standards generating the required quality and accountability.

The Administration will enter into an extensive re-engineering exercise, which will review processes to ensure outputs of the desired quality and maximum efficiency. This activity, which basically follows a zero base budgeting approach, will identify processes that can be consolidated, eliminated or streamlined in order to produce tangible results and generate budgetary savings. Typical processes that are subject to this exercise include recruitment, procurement and contract processing, payroll, inventory and invoice control, travel claims etc.

One focus of activity will be the development of a comprehensive and functional electronic document management system and archive with adequate browsing functionality that will encompass all the important documents at EMBL. This will be preceded by the creation of a set of guidelines on archiving, ensuring that the location of all final versions can be easily identified and enabling assembly of a consolidated 'documental history' for EMBL.

The operational system stands at the core of operational processes as their key facilitator. Added values of such an ERP system (SAP at EMBL) include information and reporting opportunities, automated workflows[1] and portals[2] in support of newly defined processes and efforts to become more efficient and client oriented. Even though progress has been made with internal reporting the further development of this facility will be a major focus of attention as information is key to all organisational development. EMBL is investigating the advantages of data warehousing in the context of implementing the SAP reporting suite, which will greatly facilitate providing scientists and other stakeholders with user-friendly and state-of-the-art reporting tools.

The existing grants database will also be revised to provide better information about both available and acquired grants. The new database will provide services that optimise both acquisition and management of grants, for example through monitoring tools that facilitate timely reactions and constant quality control of grant handling.

1.2.3 Staff

Attract, recruit, support and develop staff to maximise their contribution, enhance their work experience and raise their professional profile, promote an international spirit and respect diversities.

EMBL's strategic objective to recruit the very best will be pursued through a strategy based on using job advertising as a marketing tool for EMBL as a whole and achieving the highest possible congruence between job profile and the selected candidates. Targeted advertising, participation in relevant recruitment fairs and an internet area devoted to raising awareness of the benefits of EMBL employment will be some of the measures to generate value to the member states and achieve a balanced representation of nationalities among staff. These procedures will supplement EMBL's merit and quality-based recruitment policy.

Fast and effective staff integration will be supported through a comprehensive induction/onboarding package taking into consideration personal, professional and family needs and allowing a smooth start of employment at EMBL and acclimatisation to the new environment.

Throughout employment, a wide-ranging programme, including professional training, will be designed to improve the professional profiles of EMBL staff. The General Training and Development Programme takes career stages of individuals and the needs of different staff groups into account and provides pertinent monitoring. Particular focus will be put on management training, for which we are exploring the possibility of collaborating with a leading European business school.

[1] Automated workflows manage the different stages of administrative processes directly in the system including but not limited to electronic authorisation, thus reducing the need for manual input of information and related paperwork; examples of those are electronic handling of leave, employment contracting procedures etc.

[2] A portal is a website that serves as a single gateway to an organisation's information and knowledge base for employees, for example offering managers easy access to information and reports as well as entry points to automated workflows.

Figure I.3. Flyers of the 2010 General Training and Development Programme

The contribution of administrative staff to the organisation will be maximised through staff rotation. This offers the possibility to EMBL administrative staff to acquire work experience at a different EMBL site in order to learn new approaches and, ultimately, to provide for better communication, synergies and best practice implementation across sites. We will also improve the support offered to staff when they leave EMBL, for example by providing better career counselling.

All these measures are aimed at making EMBL an attractive employer and one whose past and present staff will be its best ambassadors.

1.2.4 Package

Define, implement and maintain internationally attractive and competitive conditions of employment.

EMBL needs to remain abreast of developments in terms and conditions of employment in similar scientific environments, competitive research centres and international administrations in order to maintain excellence and attractiveness. In this framework all aspects of the EMBL package will be regularly evaluated by the Administration to ensure that salaries, benefits and facilities available to staff are at desirably competitive levels and, most importantly, that they constitute the right kind of 'package' for relevant target groups both in science and in administration. In this respect, not only the visibility but also the international appeal of high-profile administrative positions will need to be carefully considered in order to permit EMBL to attract staff who will be able to perform to the advantage of both the organisation and the member states.

A central part of all competitive employment packages is social security and related systems. EMBL already provides a fairly attractive pension scheme that will be expanded to include new postdoctoral fellows during the next Indicative Scheme. This will provide them with access to the major social security provisions recommended by the European Charter and Code for Researchers, including sickness and parental benefits and pension rights. The continuity in provision of those benefits and rights will need to be secured. EMBL also provides its staff with a very attractive health insurance scheme, which is and will be regularly monitored in order to ensure its sustainability and continued competitive levels of coverage.

Any attractive employment package must also take into consideration other staff needs which are not part of the contractual agreement. Services related to short-, mid- or long-term housing, everyday catering as well as provisions for the families of staff, such as childcare or after-school care, are crucial factors that influence decisions regarding future employment. Business cases for the guesthouses/housing service, the canteen and the Kinderhaus in Heidelberg will be developed to evaluate cost efficiency and to illuminate areas in which services could be improved with minimal or no cost, for example through housing agreements with local authorities, advice on balanced and healthy nutrition and extension of the Kinderhaus activities, respectively. In this context some outstations face a particular challenge as they are less well-provided for and lack the critical mass to justify infrastructure similar to

the facilities at the main Laboratory. However, concerted efforts to find other means of addressing such needs will increase the attractiveness of those outstations for potential applicants.

Legal and tax advice have also become of increased importance to EMBL's dynamic, multicultural staff. The Administration will strive to supplement internal expertise through collaboration with other international organisations, which face similar questions from their staff.

The overarching regulatory framework for the terms and conditions of employment are EMBL's Staff Rules and Regulations. These have evolved over the past years alongside the growth and changing needs of EMBL and its staff. During the next Indicative Scheme EMBL will undertake a review of Staff Rules and Regulations, a major exercise that will aim to bring more clarity, incorporate best practice and modernise the existing document whilst consolidating the historical data and its evolution.

1.2.5 Collaboration & Outreach

Liaise and collaborate with relevant partners in order to generate added value on matters of governance, infrastructure and know-how

EMBL's increasing complexity also necessitates further improvement in communication and exchange of information between scientific staff and the Administration to achieve holistic approaches to tasks and implementation of solutions. Such approaches will generate synergies across sites and functions and lead to a more uniform corporate identity. To achieve this, regular information-exchange meetings with scientists will take place, virtual platforms (improving the availability of information about the Administration) will be implemented and additional assistance to help with the understanding of administrative procedures will be provided.

In parallel, links to local and national authorities and associations will be strengthened to ensure smooth collaboration and adequate support, and to improve the diffusion of EMBL activities into the local and national communities.

2. Resource Development

In May 2007, EMBL created an Office of Resource Development whose long-term objective is to establish a sustainable fundraising operation at EMBL by developing funding opportunities that complement EMBL's support by the member states. In the fast-changing world of molecular biology, unique opportunities can arise at short notice. Philanthropic donations allow EMBL to seize these opportunities and to pursue exceptional and ambitious projects without having to make additional ad hoc demands on the member states.

The first fundraising project of the Office of Resource Development was the EMBL Advanced Training Centre Corporate Partnership Programme (CPP). The programme was implemented in parallel with the construction of the ATC building and directly supports the activities to be carried out there. Through the CPP, EMBL aims to connect interested private sector companies with the latest developments in the molecular life sciences and to provide support for training young scientists. The programme is designed to create and enhance long-term, mutually beneficial relationships between EMBL and corporate partners. It offers three levels of engagement with three corresponding levels of annual contributions. As of the start of the active phase in January 2010, the programme has attracted 15 members[3] and currently receives €385 000 per annum. The contributions are used to sponsor conferences and training programmes and to provide fellowships intended for young scientists from less well-funded laboratories and countries. Most companies have committed to the programme for an initial three-year term starting in 2010. In the future, the CPP will be continued and moderately expanded.

On the basis of the success of the CPP, whose revenue handsomely exceeds EMBL's investments, the Office of Resource Development will embark on a long-term growth path. The main objective is to develop sustainable growth in terms of income raised from voluntary and philanthropic sources. To support these growth plans and to access the required know-how, EMBL has commissioned an internationally leading consultancy firm that specialises in

[3] Four companies at the Founder Partner Level (GE Healthcare, Leica Microsystems, Life Technologies, Olympus); five members at the Corporate Level (Becton Dickinson, Boehringer Ingelheim, Perkin Elmer, Qiagen, Sigma Aldrich); and six partners at the Associate Level (Eppendorf, Illumina, Merck Serono, Novartis, Sanofi Aventis, Thermo Fisher Scientific).

advising non-profit organisations in the fields of higher education, culture, science and research. This firm has many decades of experience and an in-depth knowledge of the specifics of fundraising and non-profit management.

During the 2012-2016 period, this advice will be implemented gradually and the fundraising efforts will be successively embedded within EMBL's scientific and training activities. Among other things, we will define a range of fundraising projects to strategically support EMBL's missions and establish a pro-active fundraising board that includes key EMBL staff and external supporters.

Appendix B.1

EMBL Missions

1. Excellent research in molecular biology
2. Service provision to member states
3. Advanced training
4. Technology and instrumentation development
5. Technology transfer

EMBL's goals and strategic objectives 2012-2016

Goal 1: Forefront life science research: setting trends and pushing the limits of technology

Strategic objectives

1. Promote EMBL's scientific excellence.
2. Pursue cutting-edge interdisciplinary research with a focus on 'information biology' and understanding the function of living systems in molecular detail.
3. Pursue a combination of experimental and computational approaches; integrate computational biology and bioinformatics seamlessly with wet-lab research.
4. Combine global, high-throughput approaches with hypothesis-driven, mechanistic analyses.
5. Integrate structural and imaging technologies operating at different resolutions to bridge scales from molecules to cells to organisms.
6. Push the limits of research by technology and instrumentation development in imaging, structural biology and computational approaches and exploit new technical developments in these areas by applying them to challenging biological problems.
7. Advance a network of formal and informal internal and external collaborations and joint activities that promote interdisciplinarity and excellence in European life science.

Goal 2: Providing world-class research infrastructure and services to the member states

1. Lead the provision of synchrotron radiation and integrated structural biology facilities in Europe.
2. Integrate synchrotron-based services with complementary imaging techniques to elucidate the structural biology of complex systems.
3. Provide Europe with the biological data that serves basic research and innovation in biology, health and agriculture.
4. Construct a new European research infrastructure for biological information with EMBL-EBI as the hub and nodes at national research organisations (ELIXIR).
5. Provide excellent services with state-of-the-art equipment, user support and appropriate computational infrastructure through the Core Facilities.
6. As Europe's intergovernmental research organisation for life science research EMBL will be a major player in the construction and operation of the next generation of biomedical research infrastructures in Europe.

Goal 3: Training and inspiring the next generation of scientific leaders

1. EMBL will build on its provision of excellent intra-mural training to develop the skills of all scientific staff.
2. Maintain EMBL's International PhD Programme as a model for other PhD programmes.
3. Implement and expand the EMBL Postdoc Programme fostering collaborative, interdisciplinary and outstanding science.
4. Provide appropriate training to all staff to boost their performance and enhance their career progression at and after EMBL.
5. EMBL will provide a world-class extramural training programme to enable the EMBL member states to benefit from EMBL's expertise, wide collaborative networks and facilities by providing high quality training to scientists.
6. Deliver a cutting edge course and conference programme that enhances interactions between disciplines and more than doubles the current participants to make optimal use of EMBL's facilities.
7. Promote the life sciences and their impact on society beyond the scientific community.

Goal 4: Driving research, innovation and progress through technology development, interaction with industry and technology transfer

1. Develop new technologies driven by scientific needs to further improve EMBL's research performance.
2. Actively engage in technology transfer activities through EMBLEM and make EMBL's discoveries and innovations available to society.
3. Establish a "Proof of Concept" fund to allow EMBL scientists to develop basic research findings further towards marketable products.
4. Engage in formal and informal interactions with bio-industries to make available the expertise of EMBL scientists and service providers.

Goal 5: Playing a leading role in the integration of life science research in Europe

1. Remain a centre of excellence and serve as a role model for life science research in Europe.
2. Develop EMBL partnerships as platforms for scientific collaborations and exchange of know how.
3. Make EMBL facilities and know-how available by hosting visitors at all sites.
4. Grow the network of EMBL alumni, developing activities of mutual benefit.
5. Encourage all European countries to join EMBL.
6. Play an active role in science policy through the interaction with decision-taking bodies and as a member of scientific interest groups such as EIROforum.

Goal 6: Excel in providing non-bureaucratic, timely and efficient administrative services through non-intrusive, flexible and effective processes and systems that support an international, progressive and nurturing work environment that attracts the highest quality staff and encourages them to fulfill their potential.

1. Create an environment able to process any organisational requirement, manage all potential changes and actively strive for and pursue improvements.
2. Establish non-bureaucratic, effective and transparent operational processes and standards generating the required quality and accountability.

3. Attract, recruit, support and develop staff to maximise their contribution, enhance their work experience and raise their professional profile, promote an international spirit and respect diversities.
4. Define, implement and maintain internationally attractive and competitive conditions of employment.
5. Liaise and collaborate with relevant partners in order to generate added value on matters of governance, infrastructure and know-how.

Appendix C.1

Selected research highlights 2005-2009

Structural Biology

- Global analysis of the proteome and the composition of macromolecular complexes in *S. cerevisiae* using a combination of tandem affinity purification, mass spectroscopy and bioinformatics.

  Gavin A.C. et al. (2006) Proteome survey reveals modularity of the yeast cell machinery. Nature 440(7084):631-6.

- Groups from EMBL Heidelberg, Grenoble and Hamburg contributed their expertise in structural and computational methods to the international project 3D Repertoire, aimed at obtaining the 3D structures of the protein complexes found in yeast. Since 2006, the project has succeeded in obtaining structural information for a considerable fraction of the yeast proteome using X-ray crystallography, NMR and electron microscopy.

  Fernández-Tornero C. et al. (2007) Insights into transcription initiation and termination from the electron microscopy structure of yeast RNA polymerase III. Mol Cell 25:813–23.

- An integrated systems biology approach to the bacterium *M. pneumoniae* has provided the first-ever blueprint of a minimal cell. Combining structural, biochemical and computational methods EMBL scientists have characterised the proteome, transcriptome and metabolome of the bacterium.

  Yus E. et al. (2009) Impact of genome reduction on bacterial metabolism and its regulation. Science 326(5957):1263-8.

  Guell M. et al. (2009) Transcriptome complexity in a genome-reduced bacterium. Science 326(5957):1268-71.

  Kühner S. et al. (2009) Proteome organization in a genome-reduced bacterium. Science 326(5957):1235-1240.

- A 3D tomographic analysis of HIV shows the assembly of the virus in human cells at unprecedented detail.

  Briggs J.A. et al. (2009) Structure and assembly of immature HIV. PNAS 106(27):11090-5.

- The first 3D reconstruction of a human skin cell was produced by electron tomography and was used to reveal the organisation of cadherin molecules that interlink skin cells.

  Al-Amoudi A. et al. (2007) The molecular architecture of cadherins in native epidermal desmosomes. Nature 450:832-7.

- The identification and structural determination of several potential drug targets in the bacterium *M. tuberculosis* using X-ray crystallography.

  Ma Q. et al. (2006) The Mycobacterium tuberculosis *LipB enzyme functions as a cysteine/lysine dyad acyltransferase. PNAS 103(23):8662-7.*

- Structural determination of the rabies virus nucleoprotein-RNA complex.

  Albertini A.A. et al. (2006) Crystal structure of the rabies virus nucleoprotein-RNA complex. Science 313(5785):360-3.

- Crystal structures of four domains of the influenza virus polymerase including both active sites for cap-snatching.

Dias A. et al. (2009) The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. Nature 458(7240):914-8.

Guilligay D. et al. (2008) The structural basis for mRNA cap-binding by influenza virus polymerase subunit PB2. Nat Struct Mol Biol 15(5):500-6.

Tarendeau F. et al. (2007) Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. Nat Struct Mol Biol 14(3):229-33.

Tarendeau F. et al. (2008) Host determinant residue lysine 627 lies on the surface of adiscrete, folded domain of influenza virus polymerase PB2 subunit. PLoS Pathog 4(8).

- Deciphering the mechanism of action of a novel boron-based antibiotic.

  Rock F.L. et al. (2007) An Antifungal Agent Inhibits an Aminoacyl-tRNA Synthetase by Trapping tRNA in the Editing Site. Science 316:1759-1761.

- Demonstration of cooperative binding of two acetylation marks on a histone tail by a single bromodomain through X-ray crystallography.

  Moriniere J. et al. (2009) Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. Nature 461(7264):664-8.

- Determination of the structure of the Slit2-Robo1 complex, which is important for axon guidance.

  Morlot C. et al. (2007) Structural insights into the Slit-Robo complex. PNAS 104(38):14923-8.

- Determination of the structure of the t60/Dt91 subcomplex of yeast transcription factor IIIC.

  Mylona A. et al. (2006) Structure of the tau60/Delta tau91 subcomplex of yeast transcription factor IIIC: insights into preinitiation complex assembly. Mol Cell 24(2):221-32.

- Determination of the structure of the abscissic acid receptor PYR1 in complex with abscissic acid.

  Santiago J. et al. (2009) The abscisic acid receptor PYR1 in complex with abscisic acid. Nature 462(7273):665-8.

- Quaternary structures of the tumour suppressor p53 and its specific p53 DNA complex in solution have been solved by combining SAXS with EM, NMR and crystallography.

  Tidow H. et al. (2007) Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex. PNAS 104:12324-12329.

- Based on the time-resolved SAXS data during insulin amyloid fibrillation, an oligomeric precursor has been detected and a novel elongation pathway of the fibrils proposed, which defines a conceptually new basis for drug design against amyloid diseases.

  Groenning B. et al. (2007) A novel elongation pathway of insulin amyloid fibrils is proposed defining a conceptually new basis for drug design against amyloid diseases. PLoS Biol 5.

- An integrated structural biology approach on a protein/protein complex has led to the assembly of a model of the high-resolution structure of an important regulatory domain of the largest protein of the human genome (titin).

  Zou P. et al. (2006) Palindromic assembly of the giant muscle protein titin in the sarcomeric Z-disk. Nature 439(7073):229-33.

Cell Biology

- The demonstration that reaction-diffusion processes can pattern signalling pathways in the cell cytoplasm.

  Athale C.A. et al. (2008) Regulation of microtubule dynamics by reaction cascades around chromosomes. Science 322(5905):1243-1247.

- A new high-speed generalisable simulation tool for cell biology applications and its use to illuminate various microtubule self-organisation problems.

  Karsenti E. et al. (2006) Modelling microtubule patterns. Nat Cell Biol 11:1204-11.

  Kozlowski C. et al. (2007) Cortical microtubule contacts position the spindle in C. elegans *embryos. Cell 129(3):499-510.*

- A new high-throughput spindle assembly assay.

  Dinarina A. et al. (2009) Chromatin shapes the mitotic spindle. Cell 138(3):502-13.

- The *in vitro* reconstitution of microtubule Plus Tip Tracking from purified and recombinant proteins.

  Bieling P. et al. (2007) Reconstitution of a microtubule plus-end tracking system in vitro. Nature 450(7172):1100-5.

- Demonstration that the Golgi can reform *de novo* from the endoplasmic reticulum exit sites using laser cutter microsurgery and micro-fluidic methods.

- Demonstration that actin filaments move chromosomes and the spindle to the cortex in animal oocytes.

  Lenart P. et al. (2005) A contractile nuclear actin network drives chromosome congression in oocytes. Nature 436(7052):812-8.

- First experimental and computational demonstration of the mechanism of spore number control by environmental cues in *S. cerevisiae*.

- Demonstration that a reaction-diffusion system involving FGF signalling increases order within a migrating epithelial tissue.

  Lecaudey V. et al. (2008) Dynamic Fgf signaling couples morphogenesis and migration in the zebrafish lateral line primordium. Development 135(16):2695-705.

- Demonstration of a function for the chemokine receptor Cxcr7 in the control of cell migration in the zebrafish lateral line.

  Valentin G. et al. (2007) The chemokine SDF1a coordinates tissue migration through the spatially restricted activation of Cxcr7 and Cxcr4b. Curr Biol 17(12):1026-31.

- Full cell volume reconstruction by 3D electron tomography in fission yeast *S. pombe* (including the modelling of the whole interphase microtubule array and the main organelles).

  Höög J.L. et al. (2007) Organization of interphase microtubules in fission yeast analyzed by electron tomography. Dev Cell 12(3):349-61.

- As part of the international project MitoCheck, which aims to identify all the proteins involved in mitosis, EMBL scientists have developed a high-throughput phenotypic screening platform combining gene silencing by RNA interference, time-lapse microscopy and computational image processing. A genome-wide phenotypic profiling of each of the ~21,000 human protein-coding genes by live imaging and quantitative scoring by computational image processing identified hundreds of human genes involved in diverse biological functions including cell division, migration and survival.

  Neuman B. et al. (2010) Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. Nature 464(7289):721-7.

Genome Biology

- Advances in the understanding of the dosage compensation process and the regulation of sex chromosome expression in *Drosophila* including identification of a role for nuclear pore proteins in the process.

Kind J. et al. (2008) Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in Drosophila. Cell 133(5):813-28.

Mendjan S. et al. (2006) Nuclear pore components are involved in the transcriptional regulation of dosage compensation in Drosophila. Mol Cell 21(6):811-23.

- Generation of the first high-resolution map of meiotic recombination outcomes in yeast.

  Mancera E. et al. (2008) High-resolution mapping of meiotic crossovers and non-crossovers in yeast. Nature 454(7203):479-85.

- Discovery of pervasive transcription and bidirectional transcription initiation at promoters in the yeast genome.

  Xu Z. et al. (2009) Bidirectional promoters generate pervasive transcription in yeast.. Nature 457(7232):1033-7.

  Neil H. et al. (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature 457(7232):1038-42.

- Mapping a resistance factor that prevents the transmission of malaria parasites to mosquitoes with genomic methods.

  Blandin S.A. et al. (2009) Dissecting the genetic basis of resistance to malaria parasites in Anopheles gambiae. *Science 326(5949):147-50.*

- Construction of a high-resolution map of large structural copy number variants on human chromosome 21 that lead to phenotypes associated with Down's syndrome.

  Korbel J.O. et al. (2009) The genetic architecture of Down syndrome phenotypes revealed by high-resolution analysis of human segmental trisomies. PNAS 106(29):12031-6.

- The identification and structure determination of conserved Argonaute-interacting linear peptide motifs and their function in RNA interference.

  Till S. et al. (2007) A conserved motif in Argonaute-interacting proteins mediates functional interactions through the Argonaute PIWI domain. Nat Struct Mol Biol (10):897-903.

- Completion of a systematic genome-wide genetic screen in *Drosophila* for mutants with Polycomb phenotypes.

- Generation of a high-resolution atlas of *cis*-regulatory module (CRM) combinatorial and temporal occupancy, representing the most comprehensive CRM atlas for any system during metazoan development to date.

  Sandmann T. et al. (2006) A temporal map of transcription factor activity: mef2 directly regulates target genes at all stages of muscle development. Dev Cell 10(6):797-807.

  Sandmann T. et al. (2007) A core transcriptional network for early mesoderm development in Drosophila melanogaster. *Genes Dev 21(4):436-49.*

- Prediction of spatio-temporal CRM activity at a global scale.

  Wilczynski B. et al. (2010) Challenges for modeling global gene regulatory networks during development: Insights from Drosophila. Dev Biol 340(2):161-9.

- Elucidation of a translational control mechanism by microRNAs and RNA-binding proteins.

  Gehring N.H. et al. (2009) Disassembly of Exon Junction Complexes by PYM. Cell 137:536-548.

  Thermann R. et al. (2007) Drosophila miR2 induces pseudo-polysomes and inhibits translation initiation. Nature 447:875-879.

  Chekulaeva M. et al. (2006) Bruno acts as a dual repressor of oskar translation, promoting mRNA oligomerization and formation of silencing particles. Cell 124:521-533.

- Definition of the *in vivo* roles of iron regulatory proteins.

  Galy B. et al. (2006) Iron homeostasis in the brain: complete iron regulatory protein 2 deficiency without symptomatic neurodegeneration in the mouse. Nat Genet 38:967-969.

  Galy B. et al. (2008) Iron regulatory proteins are essential for intestinal function and control key iron absorption molecules in the duodenum. Cell Metab 7:79-85.

- Discovery that histone Macro domains interact with NAD metabolites, including PARP, and that this interaction is used to drive chromatin remodelling.

  Timinszky G. et al. (2009) A macrodomain-containing histone rearranges chromatin upon sensing PARP1 activation. Nat Struct Mol Biol 16(9):923-9.

Organismal Biology

- Analysis of microRNA functions *in vivo*: defining features of target sites that are important for function and development of computational tools for target site prediction. This has led to the finding that microRNAs are often in reciprocal relationships with their target – both spatial and temporal – and to the idea that some miRNAs act to confer robustness by keeping target levels low.

  Stark A. et al. (2005) Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. Cell 123(6):1133-46.

- Deciphering roles of specific small regulatory RNAs in neurodegeneration, growth control, metabolism, germ cell development and in regulatory feedback loops.

  Thompson B.J. et al. (2006) The Hippo pathway regulates the bantam microRNA to control cell proliferation and apoptosis in Drosophila. Cell 126(4):767-74.

  Karres J.S. (2007) The conserved microRNA miR-8 tunes atrophin levels to prevent neurodegeneration in Drosophila. Cell 131(1):136-45.

- Splicing is required for directed transport and localisation of an mRNA in the *Drosophila* oocyte.

  Hatchet O. et al. (2004) Splicing of oskar RNA in the nucleus is coupled to its cytoplasmic localization. Nature 428(6986):959-63.

- Visualising assembly and intracellular transport of an mRNP in the *Drosophila* oocyte.

  Trucco A. et al. (2009) Assembly of endogenous oskar mRNA particles for motor-dependent transport in the Drosophila oocyte. Cell 139(5):983-98.

- Vertebrate morphogenesis visualised at the single-cell level using single plane illumination microscopy: individual cell migration serves as the driving force for optic vesicle evagination.

  Rembold M. et al. (2006) Individual cell migration serves as the driving force for optic vesicle evagination. Science 313(5790):1130-4.

- Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy provided the first 3D blueprint of embryonic development.

  Keller P.J. et al. (2008) Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. Science 322(5904):1065-9.

- Transcription factors in development, tissue homeostasis and disease: Foxl2 function is essential to maintain sexual identity in females.

  Uhlenhaut N.H. et al. (2009) Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation. Cell 139(6):1130-42.

- Identification and analysis of the conserved transcription factor Bsx indicates that it may be the

long-sought "fidgeting" gene that couples energy metabolism and behaviour.

Sakkou M. et al. (2007) A role for brain-specific homeobox factor Bsx in the control of hyperphagia and locomotory behavior. Cell Metab 5(6):450-63.

- Discovery through a combination of developmental genetics, biochemistry and single motor analysis *in vitro* of the unexpected function of a MAP in promoting productive recruitment of a specific motor to microtubules and an additional level of kinesin regulation.

  Sung H.H. et al. (2008) Drosophila ensconsin promotes productive recruitment of Kinesin-1 to microtubules. Dev Cell 15(6):866-76.

- Using *Drosophila* border cells as a model system for invasive and guided migration of groups of cells.

  Bianco A. et al. (2007) Two distinct modes of guidance signalling during collective migration of border cells. Nature 448(7151):362-5.

- Insights gained from studies of platynereis into the evolution of the vertebrate eye/photoreceptors, the vertebrate central nervous system, the vertebrate hypothalamus and of miRNA expression in bilateria.

  Arendt D. et al. (2004) Ciliary photoreceptors with a vertebrate-type opsin in an invertebrate brain. Science 306(5697):869-71.

  Tessmar-Raible K. et al. (2007) Conserved sensory-neurosecretory cell types in annelid and fish forebrain: insights into hypothalamus evolution. Cell 129(7):1389-400.

- A systems biology approach combining neurobiology and biophysics to understand larval swimming behaviour/phototaxis (in a marine zooplankton).

  Jékely G. et al. (2008) Mechanism of phototaxis in marine zooplankton. Nature 456(7220):395-9.

- Retrotransposons versus the vertebrate genome – the identification of a genetic pathway that regulates the activity and impact of several families of mobile elements.

Molecular Medicine

- Demonstration that HFE-haemochromatosis is a liver disorder resulting from defective HFE function in hepatocytes.

  Vujić Spasić M. et al. (2008) Hfe acts in hepatocytes to prevent hemochromatosis. Cell Metab 7(2):173-8.

- Identification of the Ca^{2+} channel blocker nifedipine as a compound to reverse iron overload in a murine model of haemochromatosis.

  Ludwiczek S. et al. (2007) Ca2+ channel blockers reverse iron overload by a novel mechanism via divalent metal transporter (DMT)-1. Nat Med 13(4):448-54.

- Identification of branched pathways for mammalian nonsense-mediated decay.

  Ivanov P. et al. (2008) Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. EMBO J 27:736-747.

- Elucidation of exon junction complex assembly and disassembly.

  Gehring N.H. et al. (2009) Disassembly of Exon Junction Complexes by PYM. Cell 137:536-548.

  Gehring N.H. et al. (2009) Ordered Assembly of the Exon Junction Complex by the Spliceosome. PLoS Biol 7.

- Development of novel FRET reporters for quantifying matrix metalloprotease activity in macrophages and lung lavages as biomarkers for chronic inflammatory lung disease.

Cobos-Correa A. et al. (2009) Membrane-bound FRET probe visualizes MMP12 activity in pulmonary inflammation. Nat Chem Biol 5(9):628-30.

- Design of a clinical phase II trial for cancer vaccines based on target structures predicted by a bioinformatic approach.

Mouse Biology

- EMBL scientists made an important contribution to understanding cell motility by showing that an actin-processing molecule called n-cofilin is critical for regulating cell movement.

  Gurniak C.B. et al. (2005) The actin depolymerizing factor n-cofilin is essential for neural tube morphogenesis and neural crest cell migration. Dev Biol 278(1):231-41.

- Discovery that NF-kB signalling within brain cells influences whether they live or die after a stroke.

  Herrmann O. et al. (2005) IKK mediates ischemia-induced neuronal death. Nat Med 11(12):1322-9.

- Demonstration that blocking neural signal NF-kB alleviates the symptoms of a mouse model for multiple sclerosis.

  van Loo G. et al. (2006) Inhibition of transcription factor NF-kappaB in the central nervous system ameliorates autoimmune encephalomyelitis in mice. Nat Immunol 7(9):954-61.

- Decoding the molecular signals forming blood: an intracellular communication pathway involving beta-catenin plays a central role in determining whether blood cells form.

  Kirstetter P. et al. (2006) Activation of the canonical Wnt pathway leads to loss of hematopoietic stem cell repopulation and multilineage differentiation block. Nat Immunol 7(10):1048-56.

- Demonstration that enhancing IGF-1 signalling or blocking the NF-kB signalling pathway protects muscle from degenerating after injury and improves muscle healing in mice.

  Mourkioti F. et al. (2006) Targeted ablation of IKK2 improves skeletal muscle strength, maintains mass and promotes regeneration. J Clin Invest 116:2954-2945.

- A combination of highly defined genetic mouse models of TrkB mutagenesis with behavioural studies and *in vivo* recordings establishes the first genetic link between learning and LTP in living mice.

  Gruart A. et al. (2007) Mutation at the TrkB PLC{gamma}-docking site affects hippocampal LTP and associative learning in conscious mice. Learn Mem 14(1):54-62.

- A mouse model of anxiety-related disorders reveals that the receptor (1A) for the messenger serotonin and a neural circuit involving the hippocampus play crucial roles in mediating fear responses in ambiguous situations.

  Tsetsenis T. et al. (2007) Suppression of conditioning to ambiguous cues by pharmacogenetic inhibition of the dentate gyrus. Nat Neurosci 10(7):896-902.

- A mouse model of Sudden Infant Death Syndrome reveals an imbalance of the neuronal signal serotonin in the brainstem is sufficient to cause sudden death in newborn mice.

  Audero E. et al. (2008) Sporadic autonomic dysregulation and death associated with excessive serotonin autoinhibition. Science 321(5885):130-3.

- From stem cells to skin: EMBL scientists discover two proteins, C/EBPα and C/EBPβ, control when and how stem cells switch to being skin cells.

  Lopez R.G. et al. (2009) C/EBPalpha and beta couple interfollicular keratinocyte proliferation arrest to commitment and terminal differentiation. Nat Cell Biol 11(10):1181-90.

Bioinformatics and Computational Biology

- A new computational method for the analysis of genomes produced the most accurate tree of life to date and provided insights into the origins of bacteria and the last common universal ancestor of all life on earth.
  Ciccarelli F.D. et al. (2006) Toward automatic reconstruction of a highly resolved tree of life. Science 312(5774):697
- Development of a new algorithm for species choice for comparative genomics and biodiversity conservation.

  Pardi F. et al. (2005) Species choice for comparative genomics: being greedy works. PLoS Genetics 1.

  Pardi F. et al. (2007) Resource aware taxon selection for maximizing phylogenetic diversity. Syst Biol 56:431-444.
- Development of a new phylogenetically aware method of sequence alignment.

  Löytynoja A. et al. (2005) An algorithm for progressive multiple alignment of sequences with insertions. PNAS 102:10557-10562.

  Löytynoja A. et al. (2008) Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. Science 320:1632-1635.
- Investigation of the systems biology of neuronal signalling using modelling at different scales.

  Fernandez E. et al. (2006) DARPP-32 is a robust integrator of dopamine and glutamate signals. PLoS Comput Biol 2.

  Le Novère N. et al. (2008) DARPP-32: molecular integration of phosphorylation potential. Cell Mol Life Sci 65:2125-2127.
- Using protein structural data, combined with chemical information, to elucidate enzyme catalysis.

  Holliday G. et al. (2007) The Chemistry of Protein Catalysis. J Mol Biol 372:1261-1277.

  Holliday G.L. et al. (2009) Understanding the functional roles of amino acid residues in enzyme catalysis. J Mol Biol 390:560-577.
- Functional genomics of ageing – the identification of new longevity-associated genes in mouse.

  Selman C. et al. (2009) Ribosomal protein S6 kinase 1 signaling regulates mammalian life span. Science 326:140-144.
- Predicting protein function from structure with the help of a new pipeline performing a set of computational analyses on new structures, to help to assign their function.

  Watson J.D. et al. (2007) Towards fully automated structure-based function prediction in structural genomics: a case study. J Mol Biol 367:1511-1522.

  Laskowski R.A. et al. (2005) Protein function prediction using local 3D templates. J Mol Biol 351:614-626.

  Nobeli I. et al. (2009) Protein promiscuity and its implications for biotechnology. Nat Biotechnol 27(2):157-167.
- The computational analysis of the initial ENCODE data, comprising 1% of the human genome, was led by the EBI and discovered that the human genome has a far more complex transcriptional landscape, with many more sites of transcriptional initiation than previously thought, and that regulatory elements are dispersed across the genome.
- ENCODE Project Consortium. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447:799-816.

Margulies E.H. et al. (2007) Analyses of deep mammalian sequence alignments and constraint predictions for 1% of the human genome. Genome Res 17:760-774.

Tress M.L. et al. (2007) The implications of alternative splicing in the ENCODE protein complement. PNAS 104:5495-5500.

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

- A new comparative metagenomics method to analyse environmental DNA provides insights into the microbial composition of different habitats, from soil to water, and reveals that microbes evolve faster in some environments than in others.

  Von Mering C. et al. (2007) Quantitative phylogenetic assessment of microbial communities in diverse environments. Science 315(5815):1126-30.

- Identification of the human gut microbiome.

  Qin J. et al. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464(7285):59-65.

- EMBL scientists developed a computational method that compares how similar the side effects of different drugs are and predicts how likely the drugs are to act on the same target molecule with the potential to hint at new uses of marketed drugs.

  Campillos M. et al. (2008) Drug target identification using side-effect similarity. Science 321(5886):263-6.

Appendix D.1

Research highlights from the external scientific community in structural biology

A. EMBL-GR/ESRF

1) Ribosome

Last year two long-term users of ID14-4, Venki Ramakrishnan and Ada Yonath, were awarded the Nobel Prize in Chemistry along with Thomas Steitz for their studies on the ribosome. This is undoubtedly the most complex structural biology problem to be studied on ID14-4.

Neubauer, C., Yong-Gui, G., Andersen, K. R., Dunham, C. M., Hentschel, J., Gerdes, K., Ramaakrishnan, V., Brodersen, D. E. (2009) Cell **139**, *1084-1095.*

Schmeing, T. M., Voorhees, R. M., Kelly, A. C., Gao Y-G., Murphy, F. V., Weir, J. R., Ramakrishnan, V. (2009) Science **326**, *688-694.*

Weixlbaumer, A., Petry, S., Dunham, C. M., Selmer, M., Kelley, A. C., Ramakrishnan, V. (2007) Nat. Struct. Mol. Biol. **14**, *733-737.*

Selmer, M., Dunham, C. M., Murphy, F. V. 4th, Weixlbaumer, A., Petry, S., Kelley, A. C., Weir, J. R., Ramakrishnan, V. (2006) Science **313**, *1935-1942.*

Petry, S., Brodersen, D. E., Murphy, F. V. 4th, Dunham, C. M., Selmer, M., Tarry, M. J., Kelley, A. C., Ramakrishnan, V. (2005) Cell **123**, *1255-1266.*

2) Complement system

Another long-term user of ID14-4 is Piet Gros of Utrecht University. His group has used ID14-4 for many of their structural studies on the complement system and this research has led to a more detailed understanding of underlying molecular mechanisms involved in innate immunity.

Rooijakkers, S. H., Wu, J., Ruyken, M., van Domselaar, R., Planken, K. L., Tzekou, A., Ricklin, D., Lambris, J. D., Janssen, B. J., van Strijp, J. A., Gros, P. (2009) *Nat. Immunol. 10*, 721-727.

Wu, J., Wu, Y. Q., Ricklin, D., Janssen, B. J., Lambris, J. D., Gros, P. (2009) *Nat. Immunol. 10*, 728-733.

Milder, F. J., Gomes, L., Schouten, A., Janssen, B. J., Huizinga, E. G., Romijn, R. A., Hemrika, W., Roos, A., Daha, M. R., Gros, P. (2007) Nat. Struct. Mol. Biol. **14**, *224-228.*

Janssen, B. J., Christodoulidou, A., McCarthy, A., Lambris, J. D., Gros, P. (2006) Nature **444**, *213-216.*

Janssen, B. J., Huizinga, E. G., Raaijmakers, H. C., Roos, A., Daha, M. R., Nilsson-Ekdahl, K., Nilsson, B., Gros, P. (2005) Nature **437**, *505-511.*

3) Plant photosystem I

The structure of the plant Photosystem I, one of most complex membrane proteins to be solved, has provided fundamental insights into how sunlight is converted into chemical energy. The final datasets were collected on ID14-4 (2003) and the ID23-2 microfocus beamline (2007) but thousands of crystals were screened (many with the SC3) over the years on all the MX-beamlines.

Amunts, A., Drory, O., Nelson, N. (2007) Nature **447**, *58-63.*

Ben-Shem A., Frolow F., Nelson N. (2003) Nature 426, 630-635.

4) Structure of the human beta2 adrenergic G-protein-coupled receptor

The structure of the beta2 adrenergic G-protein-coupled receptor (GPCR) was solved by screening more than 800 crystals over two years on various synchrotron beamlines. Selecting the best crystal and collecting on multiple parts of this crystal at the microfocus ID23-2 allowed a rare insight into the molecular details of one of the most important GPCRs.

Rasmussen, S. G., Choi, H. J., Rosenbaum, D. M., Kobilka, T. S., Thian, F. S., Edwards, P. C., Burghammer, M., Ratnala, V. R., Sanishvili, R., Fischetti, R. F., Schertler, G. F., Weis, W. I., Kobilka, B. K. (2007) Nature 450, 383-387.

5) Structural insights into tubulin regulation

Microtubules are dynamic cytoskeletal polymers involved in important cellular process. The structure of tublin with various factors was mostly collected on ID14-4 and this work provides mechanistic insights into tubulin dynamics.

Dorléans, A., Gigant, B., Ravelli, R. B., Mailliet, P., Mikol, V., Knossow, M. (2009) Proc. Natl. Acad. Sci. USA 106, 13775-13779.

Cormier, A., Marchand, M., Ravelli, R. B., Knossow, M., Gigant, B. (2008) EMBO Rep. 9, 1101-1106.

Gigant, B., Wang, C., Ravelli, R. B., Roussi, F., Steinmetz, M. O., Curmi, P. A., Sobel, A., Knossow, M. (2005) Nature 435, 519-522.

Ravelli R.B., Gigant B., Curmi P.A., Jourdain I., Lachkar S., Sobel A., Knossow M. (2004) Nature 428, 198-202.

6) Structural studies on ESCRT

ESCRT complexes are essential for various activities, including membrane trafficking and viral budding. A number of structural studies over the years on BM14, ID14-4, ID23-2 and various other beamlines have led to a more detailed understanding of the ESCRT assembly mechanism.

Obita, T., Saksena, S., Ghazi-Tabatabai, S., Gill, D. J., Perisic, O., Emr, S. D., Williams, R. L. (2007) Nature 449, 735-739.

Gill, D. J., Teo, H., Sun, J., Perisic. O., Veprintsev, D. B., Emr, S. D., Williams, R. L. (2007) EMBO J. 26, 600-612.

Muzioł, T., Pineda-Molina, E., Ravelli, R. B., Zamborlini, A., Usami, Y., Göttlinger, H., Weissenhorn, W. (2006) Dev. Cell 313, 360-363.

Teo, H., Gill, D. J., Sun, J., Perisic, O., Veprintsev, D. B., Vallis, Y., Emr, S. D., Williams, R. L. (2006) Cell 125, 99-111.

Teo, H., Perisic, O., González, B., Williams, R. L. (2004) Dev. Cell 7, 559-569.

7) Structure of the rabies virus nucleoprotein-RNA complex.

The crystal structure of the rabies virus nucleoprotein–RNA complex was solved on ID14-4 and provides vital information on how negative-strand RNA viruses protect their genomes from the innate immune response.

Albertini, A. A., Wernimont, A. K., Muziol, T., Ravelli, R. B., Clapier, C. R., Schoehn, G., Weissenhorn, W., Ruigrok, R. W. (2006) Science 313, 360-363.

B. EMBL-HH

Support and scientific assistance by the EMBL-HH beamline scientists is acknowledged in all papers

1) Structure of the ribosome-dependent endonuclease, RelE

The crystal structure was solved with X-ray data from beamline X12. It reveals how mRNA is specifically cleaved in the presence of the ribosome.

Neubauer, C., Gao, Y. G., Andersen, K. R., Dunham, C. M., Kelley, A. C., Hentschel, J., Gerdes, K., Ramakrishnan, V., Brodersen, D. E. (2009) Cell 139, 1084-95.

2) Structure of the catalytic core of a membrane-associated eukaryotic polyphosphate polymerase

The crystal structure was solved from X-ray data from beamline BW7A. It shows how adenosine triphosphate winds through a tunnel-shaped pocket of the polymerase catalytic domain.

Hothorn, M., Neumann, H., Lenherr, E. D., Wehner, M., Rybin, V., Hassa, P. O., Uttenweiler, A., Reinhardt, M., Schmidt, A., Seiler, J., Ladurner, A. G., Herrmann, C., Scheffzek, K., Mayer, A. (2009) Science 324, 513-6.

3) Structure of the bifunctional catalytic fragment of the Rel/Spo homologue from *Streptococcus dysgalactiae*

The crystal structure was solved from X-ray data from beamline BW7B. The structure of Rel(Seq) reveals two conformations of the enzyme corresponding to known reciprocal activity states: (p) ppGpp-hydrolase-OFF/(p)ppGpp-synthetase-ON and hydrolase-ON/synthetase-OFF.

Hogg, T., Mechold, U., Malke, H., Cashel, M., Hilgenfeld, R. (2004). Cell 117, 57-68.

4) Structure of the sarco(endo)plasmic reticulum Ca2+-adenosine triphosphatase

The crystal structure was solved from X-ray data from beamline X11. The structure revealed conformational changes that accompany the reaction with ATP, pull the transmembrane helices and close a cytosolic entrance for calcium ions.

Sørensen, T. L., Møller, J. V., Nissen, P. (2004) Science 304, 1672-5.

5) Structure of the coronavirus main protease

The crystal structure was solved from X-ray data from beamline X13. The structure has been used as basis for design of anti-SARS drugs.

Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., Hilgenfeld, R. (2003). Science **300**, *1763-7.*

Appendix D.2

Research projects that have been enabled by Core Facilities in the period 2005-2009

1. Genomics Core Facility

- *Studying vascular progenitor cells in the epicardium*: The outer cell layer of the adult heart (epicardium) purportedly contains a vascular progenitor cell pool for cardiac regeneration, but there is a current scarcity of defining markers to enable further study. Laser microdissection and microarray analysis provided by the Genomics Core Facility have been used to identify several novel gene loci expressed specifically in the epicardium.
- *Systems biology of the minimal cell* Mycoplasma pneumoniae: The facility provided access to robots, protocols and trained scientists for the cloning of ~700 genes. GeneCore also contributed to the design of microarrays and provided guidance, reagents and protocols for the preparation of the RNA samples and the use of the microarrays required for the analysis of genome and transcriptome of the bacterium.
- Platynereis dumerilii *genome sequencing and miRNA analysis*
- *ChIP DNA sequencing in mouse embryonic stem cells*: Illumina GA sequencing of chromatin-immunoprecipitated DNA to resolve binding sites of several pluripotency-related transcription factors (Stat3, Rex1, Klf4) and histone modifications in mouse ES cells.
- *RNA sequencing in mouse embryonic stem cells*: Illumina GA sequencing of total RNA in mouse ES cells to enable global gene expression profiling in response to Stat3 transcriptional regulation.
- *RNA sequencing in human neural cancer stem cells*: Illumina GA sequencing for transcriptome analysis of glioblastoma cancer stem cell lines and their iPS cell counterparts, normal neural stem cells and human embryonic stem cell controls.
- *Genome-scale assessment of the regulation of dosage compensation*: Array hybridisations for transcriptomic and ChIP measurements
- *Genome-scale assessment of nucleoid-associated proteins in bacteria*: Array hybridisations for transcriptomic measurements. Solexa sequencing for ChIP-Seq and RNA-Seq.
- *Mapping the global distribution of meiotic recombination sites in yeast using tiling arrays*
- *Time resolved analysis of DNA methylation*; Participation in the research showing that DNA methylation is a dynamic process and adds another layer of complexity to the regulation of gene expression

2. Proteomics Core Facility

- *Characterisation of the proteome of the minimal bacterium* Mycoplasma pneumoniae*:* access to the MS instrumentation (MALDI-TOF), the gel imaging systems, but also expertise (protocols, procedure, and data storage, etc).
- *Tandem mass spectroscopy for histone modifications*: semi-quantitative analysis of histone H4 hyperacetylation in spermatids by tandem mass spectroscopy
- *The* Mycobacterium tuberculosis *LipB enzyme functions as a cysteine/lysine dyad acyltransferase*: Identification of covalent modification in crystals (decanoic acid)

3. Protein Expression and Purification Core Facility

- *Structural determination of Polycomb complexes and bromodomains*: ITC measurements to identify binding affinities of different "chromatin" domains for modified histone tail peptides.
- *Determination of protein-lipid interactions*: Large-scale expression and purification of proteins in *E. coli* and ITC measurements
- *Characterising the proteome of* Myocplosma pneumoniae: setting-up of gel filtration systems for the validation of protein complexes
- *Identification of a novel acetyltransferase with a conserved role in mechanotransduction*: expression and purification of the protein, providing advice that solved stability problems with the protein and choosing appropriate conditions for *in vitro* assays.
- *Purification of protein complexes regulating Notch signalling*: expression of recombinant proteins using the baculovirus expression system
- *Synthesis and expression of transcription factor proteins*: expression of tagged protein domain fusions for monoclonal antibody production.

4. Monoclonal Antibody Core Facility (MACF)

- *Identification of a novel acetyltransferase with a conserved role in mechanotransduction*: elucidation of the protein's function with the help of antibodies produced by the MACF
- *Monoclonal antibody production for ChIP and protein complex isolation*: raising antibodies to several human and mouse transcription factors to be used for chromatin IP and protein complex isolation
- *Detection of small regulators of Notch signaling pathways* with antibodies produced by the MACF
- *Generation of an 'Argonaut antibody tool box'*

5. Chemical Biology Core Facility (CBCF)

- *Identification of small molecule inhibitors of the influenza virus polymerase through chemical screens*
- *Screens for novel agonists of a Drosophila ion channel called Painless*: the channel is essential for nociception in flies and in its absence fly larvae no longer respond to noxious stimuli. The CBCF designed a medium-throughput assay that utilizes a novel luminescent calcium sensor to measure channel activity.
- *Dissection of the molecular basis of estrogen signaling*: screens run by the CBCF have revealed small molecules that disrupt the pathway and can be used as biotools
- *Developing inhibitors for Aurora kinase A:* Aurora kinase A is an enzyme with a central role in cell division. The CBCF found small molecules that block Aurora kinase acting through an allosteric site. The biotools are helping dissect the function of this important cancer target in the cell cycle and may provide the starting point for the development of novel anti-cancer drugs.
- *Developing inhibitors for motor proteins:* the CBCF found inhibitors for motor proteins of the parasite *Leishmania*, the causative agent of sleeping sickness. The inhibitors are being used to decipher the precise role of the motor proteins in cell division.

6. Advanced Light Microscopy Facility (ALMF)

- *Live imaging of trafficking pathways in the early Drosophila embryo*
- *Development of a new system to study the interaction between microglia, the phagocytes of brain, and damaged neurons in living zebrafish embryos*: a UV pulse laser is coupled with a confocal microscope to create small and large lesions in the brain to investigate the response of the microglial population to the dying neurons. The basic technical and instrumentation requirements for the project are available at the ALMF.
- *Large scale RNAi screening*: identifying all genes involved in mitosis in the context of the international MitoCheck project
- *DNA damage repair genome wide screen* with member state outside users

7. Electron Microscopy Core Facility (EMCF)

- *Full cell volume, 3D reconstruction of* S. pombe *by electron tomography*
- *Analysis of microtubule-tip-tracking-protein mutants* (mal3 and tip deletion mutants) in fission yeast by light microscopy and electron tomography
- *High-resolution analysis of RNP transport complex assembly*
- *First 3D reconstruction of a human skin cell revealing the organization of cadherin molecules that interlink skin cells*: applied EM tomography using plastic sections that has been pioneered in the core facility

8. Flow Cytometry Core Facility (FCCF)

- *Role of lipids in killing mycobacteria by macrophages*: evidence for NF-kB-dependent and -independent killing induced by different lipids
- *Cellular uptake of PNA-terpyridine conjugates and its enhancement by* $Zn^{(2+)}$ *ions*
- *A role for brain-specific homeobox factor Bsx in the control of hyperphagia and locomotory behavior*
- *Loss of GLIS2 causes nephronophthisis in humans and mice by increased apoptosis and fibrosis*
- *Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation*

9. Mouse Transgenic Core Facilities

- *Identification of molecular markers for different subsets of sensory neuron and to manipulate and record from these neurons in vivo*: several mouse lines are being generated that express fluorescent biosensor molecules for measurement of electrical activity and ion flux *in vivo*.
- *Identification of a novel acetyltransferase with a conserved role in mechanotransduction*: A conditional mouse knockout is being made to examine the function of the protein *in vivo*
- *Studying vascular progenitor cells in the epicardium*: The outer cell layer of the adult heart (epicardium) purportedly contains a vascular progenitor cell pool for cardiac regeneration. To manipulate gene expression in this cell layer, the facility is using state-of-the-art intracytoplasmic sperm injection techniques to generate a series of mouse lines carrying transgenic BAC loci into which inducible Cre recombinase cassettes have been integrated. These Cre driver strains will capture the expression patterns of the novel loci, and facilitate lineage tracing and targeted genetic manipulation specifically in the epicardial layer.
- *Mouse genetics in the investigation of sex determination and locomotory behavior*

Appendix E.1

Curriculum of the EMBL EIPP Core Course

The Predoc course is a two-month interdisciplinary course that is mandatory for all EIPP students. It introduces the new PhD students to EMBL, its research and other activities, and teaches interdisciplinary science to provide all the students, that often come from different disciplines, with sufficient general background knowledge before they embark on their individual PhD projects. The course consists of seven modules which span the major research activities of EMBL: Chromatin and Translation, Cell Biology, Disease Mechanisms and Pathogens, Genomics, Proteomics and Structural Biology, Developmental Biology and Behaviour and Evolution. Each module lasts for about a week and contains lectures and practicals. Lectures are taught in an interdisciplinary fashion by group leaders from several EMBL Units and outstations and collectively involve close to every research group at EMBL.

Between the scientific modules, the course features general training modules, such as lectures on good scientific practice, science communication, and patenting. Students are engaged in discussions with their colleagues, through practical presentations, a symposium organized by the previous year's Predocs, visits to the outstations and a weekly Journal Club. In addition, the core course provides training on complementary skills, and introductions to the scientific services, to the EMBL administration services as well as the EMBL outreach and science & society activities.

In the second year of their PhD the predocs all attend a mandatory 1-week course on bioinformatics and computational biology tools, databases and approaches that is organised at EMBL-EBI. By the second year of their PhD projects, most of the students can select modules on data resources and bioinformatics tools of direct relevance to their own work.

1. Students teaching students: Basic Teaching (3 days)

Biologists, chemists, physicists, informaticians and mathematicians all join the EMBL International PhD programme, contributing diverse backgrounds and complementary knowledge. To take full advantage of the interdisciplinary environment, they must learn the general principles and specific terminology of different scientific disciplines.

This introductory module was introduced in 2009 and is designed and taught by current EMBL Predocs. It is structured in two parallel streams, one for non-biologists and one for biologists. New Predocs can decide which course they want to join, according to their previous education. The second year students share their expertise with the new arrivals from distinct disciplines. Finally, both groups merge for joint discussions and presentations.

2. Cell Biology (4 days)

This module gives an introduction into the living cell as the basic functional unit of life. The focus is on the principles governing spatial organisation and temporal development of functional modules inside eukaryotic cells.

Within this module, students are given an overview of the methods that are currently used to address the most pressing questions about cell organisation and function. The course explains why these methods are required in order to achieve a quantitative understanding of the complex behaviour of cells that goes beyond what can be achieved by classical qualitative approaches.

3. Chromatin & Translation (5 days)

Life relies on translating genetic information into physically and biologically functional gene products. This flow of information from genes to effectors (RNA and proteins) is complex and tightly regulated. Gene expression allows the genetic blueprint to produce complex phenotypes, thus giving living organisms the ability to develop and respond to their ever-changing environments.

In this module, students learn about fundamental principles of gene expression control. Topics span from chromosome architecture to understanding how protein synthesis is governed by small interfering (si)RNA molecules. The module also includes the experimental study of gene regulation, first by joining EMBL research teams at the bench and also in dedicated literature-based seminars.

At the end of this module, students have an up-to-date understanding of many of the molecular mechanisms involved in gene expression regulation and have learned about useful inter-dependent methods, tools and approaches in this field.

4. Genomics (4 days)

Genomics has revolutionised the way that biology is done: research now aims to investigate the complete set of components of functional biological systems, rather than just selected items; evolution is studied more precisely in terms of DNA sequences than of morphology; diseases can be diagnosed by their molecular profiles; and researchers are beginning to unravel the molecular mechanisms behind inter-individual differences. It is becoming increasingly possible to spatially and temporally map out and model the interactions of cellular components such as proteins, metabolites, different types of RNA, and DNA.

Since genomics is a fast-moving field that is driven by technology, its development strongly relies on advances in computational biology and related disciplines such as statistics, computer science, and mathematics. The whole array of computational innovations is important to structure massive amounts of data, to create maps of the system and eventually predictive models.

The module begins by giving an overview over the recent developments in DNA sequencing, followed by lectures on sequence alignment technologies, basic statistical concepts, DNA sequence variations and phenotypes and the computational methods required for genomics studies. Discussions include the ethical, legal and social implications of personal genomics. An additional section covers progress in the area of functional genomics technologies.

5. Proteomics & Structural Biology (7 days)

Proteins are the main biological effectors. Logically therefore, Proteomics and Structural Biology has emerged as a key discipline, that aims at characterising and defining (functionally, structurally, etc.) the full sets of proteins expressed from the genome. This involves the development of new techniques as well as expansion of classic methods to formats where high-throughput technology can be employed.

In this course PhD students learn about methods and concepts in protein structure and function. Starting from experimental approaches such as electron microscopy, X-ray crystallography and nuclear magnetic resonance, they are guided through the principles, technical challenges and applications of those methods in lectures and practicals. They are exposed to classic concepts and major developments in mass spectrometry and to the study of protein-protein interactions in cells. The course is designed to expand the view from the individual molecular structure to the 'omics' scale and includes examples for the use of structural methods in high- (or at least medium-) throughput formats.

In addition to the Journal Club discussions group of classical papers, the module comprises a seminar in which the course participants present and discuss selected papers in the fields of Proteomics and Structural Biology.

6. Disease Mechanisms & Pathogens (4 days)

This module covers topics at the interface of biology and medicine. Subjects covered are cancer, genetic diseases and bacterial and viral pathogenesis.

On the molecular level, there is a particular focus on structural aspects of the proteins involved in disease and on drug design.

7. Developmental biology & behaviour (5 days)

Developmental biology is at the core of organismal biology, since it is concerned with how the body patterns of multicellular organisms are designed and implemented. It deals with the process by which the genes in the fertilised egg control cell behaviour in the embryo and so determine the nature of the emerging animal or plant. This process converts one-dimensional information present in the genome into three-dimensional information expressed in the structure of the organism.

Whereas classical developmental biology relied heavily on embryology and genetics, recent advances in the areas of cell biology, genomics, *in vivo* imaging, and computational biology have had a strong impact on the field, and are increasingly transforming developmental biology into a highly interdisciplinary field concerned with understanding the dynamic behaviour of entire networks of genes, cells, and organs in developing plants and animals.

Since the structure and function of mature organisms can only be understood in light of their development, developmental biology is also concerned with how genes important for development have major effects on the function of organs in the mature animal. This can be clearly seen in the case of genes important for development of the brain, which have profound effects on behaviour.

8. Evolution (5 days)

This module is designed to explore evolution both theoretically and experimentally, using as an example the evolution of gene networks. Students determine what aspects of the network architecture are prone to evolutionary change, and what aspects tend to be highly conserved. Is it the coding sequence or rather non-coding sequence elements? How about gene regulatory regions in comparison to transcription factor binding sites? How easily does differential gene expression change and what are the consequences? What is the level of conservation of protein-protein interaction and protein-gene interaction between species?

To answer these questions, the lectures and practical focus on a well-characterised pair of duplicate genes, *pax*6 and *pax*2, which play well-characterised roles in eye and brain development. The practical traces these genes and the gene networks they form part of through animal evolution, looking into fly, *Platynereis*, fish and mouse experimentally, and at many more vertebrate and invertebrate animals bioinformatically.

EMBL Programme 2012-2016

Planning, co-ordination, writing
Iain W Mattaj, Anna-Lynn Wegener

Graphics
Nicola Graf, Petra Riedinger

Layout and composition
Nicola Graf, Vienna Leigh

Print
ColorDruck Leimen, Germany

EMBL group and team leaders, all the Heads of Units and multiple other contributors throughout EMBL provided excellent draft input. We thank them for their cooperation.

A special thank you to the EMBL Scientific Advisory Committee and to the additional reviewers for their valuable input and feedback.

EMBL

EMBL Heidelberg
Meyerhofstraße 1
69117 Heidelberg
Germany

Tel. +49 (0)6221 387 0
Fax +49 (0)6221 387 8306

EMBL Grenoble
6, rue Jules Horowitz, BP 181
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 89 902 0
Fax +49 (0)40 89 902 104

EMBL-EBI
Wellcome Trust Genome
Campus, Hinxton
Cambridge CB10 1SD
United Kingdom

Tel. +44 (0)1223 494 444
Fax +44 (0)1223 494 468

EMBL Monterotondo
Adriano Buzzati-Traverso Campus
Via Ramarini, 32
00016 Monterotondo (Rome)
Italy

Tel. +39 06 900 912 85
Fax +39 06 900 912 72

www.embl.org