EMBL Programme
2007–2011

EMBL

TABLE OF CONTENTS

List of boxes

List of figures

List of annexes

Abbreviations

ALMF	Advanced Light Microscopy Facility
ArrayExpress	The transcription database
ATC	Advanced Training Centre
BAC	Bioinformatics Advisory Committee
BMBF	Bundesministerium für Bildung und Forschung (German Ministry for Research and Education)
CBB	Cell Biology and Biophysics Unit, EMBL Heidelberg
CBCF	Chemical Biology Core Facility
CEA	Commissariat à l'Energie Atomique
CERN	European Organisation for Nuclear Research
ChEBI	Chemical Entities of Biological Interest
ChIP	Chromatin Immunoprecipitation
CNRS	Centre National de la Recherche Scientifique
CSHL	Cold Spring Harbor Laboratory
DB	Developmental Biology Unit, EMBL Heidelberg
DESY	German Electron Synchrotron Research Centre
DKFZ	German Cancer Research Centre
EBI	European Bioinformatics Institute
EC	European Commission
EFDA	European Fusion Development Agreement
EICAT	EMBL International Centre for Advanced Training
EIPP	EMBL International PhD Programme
EIROforum	European Intergovernmental Research Organisation Forum
ELLS	European Learning Lab for the Life Sciences
ELMI	European Light Microscopy Initiative
EM	Electron Microscopy
EMBC	European Molecular Biology Conference
EMBL-Bank	The DNA sequence archive
EMBLEM	EMBL Enterprise Management
EMBO	European Molecular Biology Organisation
EMCF	Electron Microscopy Core Facility
Ensembl	Genome information
ESA	European Space Agency
ESO	European Southern Observatory
ESPRIT	Expression of Soluble Proteins by Random Incremental Truncation
ESRF	European Synchrotron Radiation Facility
ETF	EMBL Technology Fund
EU	European Union
EXAFS	X-ray Absorption Spectroscopy
FCCF	Flow Cytometry Core Facility
FCS	Fluorescence Correlation Spectroscopy
FLAP	Fluorescence Loss After Photobleaching
FLIM	Fluorescence Lifetime Imaging Microscopy
FRAP	Fluorescence Recovery After Photobleaching
FRET	Fluorescence Resonance Energy Transfer
GE	Gene Expression Unit, EMBL Heidelberg
GO	Gene Ontology
HASYLAB	Hamburg Synchrotron Radiation Laboratory (DESY)
IBS	Institut de Biologie Structurale
ILL	Institute Laue-Langevin
IntAct	Protein–protein interaction databases
Integr8	A database that provides integration of gene-related information from other component databases
IntEnz	International standard for enzyme nomenclature

InterPro	An aggregation of different databases on protein functional motifs
IP	Intellectual Property
IT	Information Technology
ITC	Isothermal Titration Calorimetry
IVMS	Institut de Virologie Moléculaire et Structurale
JSBG	Joint Structural Biology Group
LIMS	Laboratory Information Management System
MACF	Monoclonal Antibody Core Facility
MMPU	Molecular Medicine Partnership Unit
MRI	Magnetic Resonance Imaging
MS	Mass Spectrometry
MSD	Molecular Structures Database
MX	Macromolecular crystallography
NCBI	National Center for Biotechnology Information
NIH	National Institutes of Health
NMR	Nuclear Magnetic Resonance
OIPA	Office of Information and Public Affairs
PCR	Polymerase Chain Reaction
PDB	Protein Database
PEPCF	Protein Expression and Purification Core Facility
PIR	Protein Information Resource
PSB	Partnership for Structural Biology
Reactome	Molecular processes and pathways database
RNAi	RNA interference
SAC	Scientific Advisory Committee
SAD/MAD	Single- or Multi-Wavelength Anomalous Dispersion
SAGE	Serial Analysis of Gene Expression
SAXS	Small Angle X-ray Scattering
SCB	Structural and Computational Biology Unit, EMBL Heidelberg
SFL	Strategic Forward Look 2006–2015
SIB	Swiss Institute of Bioinfomatics
SLS	Swiss Light Source
SME	Small-to-Medium Enterprise
SPIM	Selected Plain Illumination Microscopy
TAP	Tandem Affinity Purification
TB	Tuberculosis
TCF	Transgenic Core Facilities
UniProt	The global protein sequence database

A. EXECUTIVE SUMMARY

1. Molecular biology and Europe

In terms of its impact on society, molecular biology is possibly the most important current field of scientific research. Molecular biology is in the process of providing a detailed explanation of how living organisms function or, to put it another way, how life works. New technologies and research findings have revolutionised biology, and the rapid current rate of advance is set to continue for the foreseeable future. On the one hand, molecular biology research is a tremendously exciting intellectual adventure, comparable with exploring the composition of matter or the origins of the universe. On the other, understanding living systems is providing unprecedented insight into human health and disease, and thus has immediate impact on human welfare. Aside from its medical applications, insights from molecular biology are directly relevant to agriculture and other industries, for example the pharmaceutical industry, and molecular biology research has given rise to one of the most innovative sectors of current industry, biotechnology. These factors have contributed to the world-wide growth of research activity in molecular biology over the past decade.

Although molecular biology began in Europe, and is still actively pursued here, there is no doubt that we are losing ground to our competitors overseas. A reasonable measure of success in basic research is provided by bibliometric analyses of the impact of publications. The latest long-term survey of performance in molecular biology and genetics, which covered the period 1992–2002, ranked EMBL as the best research institute in Europe (see Annex 3). Although this was gratifying, we were one of only two European institutes in the world's top twenty. The previous ten-year survey had seen EMBL as one of five European institutes in the top ten!

Some of the reasons for this decline are easily discerned. The period under study saw an enormous increase in funding for life science research both in the US and in Japan, Europe's main competitors. Investment in this area in fast-growing countries, particularly China, is also considerable. Coupled to this is the fact that European research funding is often spread too thinly, and with too little focus on excellence, due to the still highly diverse nature of funding mechanisms in individual European countries.

2. EMBL's role

EMBL is a flagship laboratory for European molecular biology. The missions it carries out on behalf of its member states and the added value it brings to them depend upon the excellence of its research performance, that was briefly documented above, and on the mechanisms in place for ensuring a wide dissemination of EMBL's expertise and benefits to the scientific communities of the member states. The most important of these mechanisms are:

- EMBL's training programmes. EMBL trains scientists at all levels, from PhD students to young group leaders. EMBL's turnover system means that it produces world-class researchers at a much higher rate than any other research institute, 88% of whom move on to positions in our member states (see Annex 6).
- EMBL organises courses and conferences for outside researchers. Because of the dedicated involvement of the EMBL scientists who organise them, these are of very high scientific quality. Over 100 such meetings were organised in the last four years at EMBL Heidelberg alone, most together with our sister organisation EMBO, and attracted more than 9,000 participants (see Annex 7).
- EMBL directly provides services to the scientific communities in the member states. The most important of these are in the areas of structural biology and bioinformatics. EMBL provides access for life scientists to X-ray and neutron beamlines dedicated to structural biology at our Hamburg and Grenoble

Outstations (see Annexes 13 and 14). These are utilised by more than 2,000 visitors each year. A large fraction of all new macromolecular structures submitted to the databases rely on data collected in Hamburg or Grenoble (see Annex 12). The European Bioinformatics Institute (EBI) designs, builds and provides access to data resources that cover all aspects of core biomolecular data. These databases are accessed by more than 200,000 regular users. They serve a very broad community that includes basic biology researchers, clinicians, and researchers from the pharmaceutical industry, agriculture and ecology. The EBI is the only European centre that provides this data and it is also the European node in international collaborations that ensure access for European scientists to data collected throughout the world.

- More than 1,000 additional scientific visitors each year are attracted to the five EMBL sites, situated in four different member states. They come to use EMBL's facilities, engage in collaborative research or learn techniques (see Annex 17).
- EMBL has begun to establish cooperative partnerships with scientifically complementary institutes in the member states. These enable our partner institutes to benefit from EMBL's expertise in promoting research excellence and furthering internationality, in addition to engaging in collaborative research. This provides a unique opportunity to the member states to set up research institutes that function along the same lines as EMBL.
- EMBL scientists are very actively engaged in European-level scientific networks and in helping promote the European Research Area. Our scientists (84 group leaders in total) coordinate no fewer than 12 EU-funded network projects in the context of Framework Programme 6 and are partners in 46 additional such networks (see Annex 10).
- EMBL, through its technology transfer company EMBL Enterprise Management (EMBLEM), is actively engaged in ensuring that its basic research discoveries lead to innovation to the benefit of its member states' citizens. Eight EMBL spin-off companies have been founded to date, two of which are now publicly quoted.
- EMBL also engages in dialogue with the non-scientific public via extensive outreach and Science and Society activities.

3. Highlights of EMBL's strategy for 2007–2011

EMBL Council charged the previous administration to perform a detailed analysis of the prospects for molecular biology in the coming decade. This analysis, the "Strategic Forward Look: 2006–2015" (SFL) was published at the end of 2003 (http://www.embl.org/aboutus/news/publications/forwardlook.html). Its conclusions form the basis of EMBL's scientific strategy for the next five years, as briefly summarised here and discussed in later sections.

The major conclusion of the SFL was that developments in molecular biology over the past decade, including the sequencing of entire genomes, the broad advent of high-throughput technologies and a few excellent examples of the power of combining experimental analysis with computational modelling and simulation methods in biology, had paved the way for the advent of systems analysis in biology, or systems biology. Biological functions rely on the combinatorial use of multiple components, whose interactions modify each others' properties in non-intuitive ways. This dictates a requirement for interdisciplinary approaches to the understanding of biological phenomena, including the close integration of computational and experimental approaches. EMBL's response to this challenge will include the creation of EMBL Centres, horizontal activities that will bind the existing research Units and promote interdisciplinarity in three areas that we see as crucial for systems biology; computational biology, imaging and image analysis, and high-throughput

technology. We will expand our activity in chemical biology, mainly via recycling resources from biochemical instrumentation and structural and computational biology, because we see chemical perturbations as critically important for systems-level understanding. We will also need to further strengthen aspects of our scientific support structures, including EMBL's IT infrastructure and its Core Facilities, which provide both equipment access and technical support to our research groups.

The second research area highlighted in the SFL reflects the close relationship between molecular biology and disease. Also here, EMBL will respond by building up a Centre, the Centre for Disease Mechanisms, that will unite the numerous groups carrying out research on the basic mechanisms underlying disease states. Like other aspects of biology, research into disease is multi-faceted. Nevertheless, it is evident that the premier model organism for the study of human disease is the mouse. In reflection of this and in response to the recommendation of the Scientific Advisory Committee (SAC) of EMBL Council, we intend to increase our activity in this area by adding two research groups to our Monterotondo Mouse Biology Outstation.

The major investment needed at EMBL in the next five years will, however, not directly support our research activity. As noted above, molecular biology is an information-rich science. The advent of high-throughput methodologies has meant a further increase in the rate of data collection. This data is only useful if available to the whole scientific community. The EBI is responsible for the task of collecting, organising, annotating and providing it to European scientists. Both the rate of growth of this data and of its usage is increasing exponentially. Roughly 2 million pages of information are accessed on the EBI websites every day. This growth means that, in order to fulfil its mandate, the EBI must increase both its staff complement and its investment in infrastructure. In response to the urgent need, three UK research funders, the Wellcome Trust, the UK Medical Research Council and the Biotechnology and Biological Sciences Research Council, have generously agreed to fund an expansion of the existing EBI building. However, support for the required staff increase from 260 currently to roughly 400 by 2011, and the accompanying hardware and software infrastructure costs, will need to come in large part from the EMBL budget. Although we will work with the member states to increase the level of external funding coming into the EBI, particularly from the EU, the provision of the European data resources is a major responsibility of EMBL and its member states, and thus the largest item in our funding request.

In terms of service activity, the other large investment EMBL will be making over the next five years is in designing and building new beamlines for structural biology on the upgraded PETRA-III synchrotron radiation ring in Hamburg. We are very hopeful that the host country, Germany, will continue its tradition of very generous support for EMBL by providing the bulk of the required funding and that we can cover the remaining costs by redirecting funds available within the EMBL budget for the Hamburg Outstation to this project. A final decision on funding the beamlines should by available prior to the EMBL Council meeting in July 2006.

One of the biggest advantages that US science has over its international competitors is the availability of a large number of world-class centres that organise scientific meetings on various scales, including organisations such as Keystone or the Gordon Research Conferences and the Cold Spring Harbor Laboratory (CSHL). These conferences are open to the international community, but the costs of travel and accommodation mean that the vast majority of participants are nationals. This is particularly true of younger participants including PhD students, postdoctoral fellows and young faculty.

Within Europe, EMBL scientists have a strong tradition of organising small- to medium-sized courses, workshops and conferences, most of which are evaluated, selected and funded by our sister organisation EMBO. Our conference facilities however limit the scope and impact of these activities. We have inadequate teaching laboratories and no space suitable for poster presentation, a crucial aspect of the best scientific meetings.

Through generous support from Germany and the offer of an unprecedently large private donation, EMBL currently has the opportunity to establish an advanced training centre (ATC) that will host both state-of-the-art courses and conference infrastructure and many of our training activities. Our intention is to work together with EMBO to turn this centre into a meeting point for scientists from the member states that will help galvanise Europe's potential in the life sciences and provide world-class training for our young scientists.

The foundation whose donation makes building the ATC possible is also prepared to facilitate the additional financing required to complete the project. We are still actively engaged in discussions and in fund-raising, and will present alternative finance models to Council for decision in March 2006. An early decision on this project is essential as it must begin in 2006.

In making these proposals, we at EMBL are keenly aware of the difficulties faced by our member states in finding the means to support scientific research in the present economic climate. We have therefore made considerable effort to moderate our requests, and to focus these on areas that we feel will help EMBL provide a maximum in added value to the countries who support us. Given the commitment of many of our member states to the goal defined by the Barcelona agreement, of making Europe the world's leading knowledge-based economy by 2010, and our conviction that EMBL can continue to play a critical role in the integration and development of life science research in all of our member states, we feel it would be a serious error on our part not to ask you to support us in these efforts to the best of your ability.

B. INTRODUCTION

EMBL is a flagship project for Europe. It was conceived as a means to strengthen basic research in the molecular life sciences in Europe and is equipped with a set of missions that allow it to both supplement and foster related activities in the national research communities. The Laboratory is currently supported by the governments of 19 member states (Annex 1). From its beginnings over 30 years ago, EMBL has promoted molecular biology across Europe and has launched the careers of a significant number of Europe's leading molecular biologists. Many of our scientists are recruited to EMBL from overseas, where they undertake postdoctoral work. After establishing their reputations at EMBL, they move on to re-integrate into the national systems. Through its unusual structure, the Laboratory has developed a philosophy of looking outward into the scientific community, sharing its knowledge and expertise widely to the benefit of molecular biology in all its member states.

The Laboratory has five sites, with a total of 1,345 staff, including postdoctoral fellows and PhD students. A total of 67 nationalities are represented, including all member states. Within the Laboratory, eight Units carry out a combination of research and service functions that target large sections of the scientific communities of the member states.

- The main laboratory in Heidelberg, Germany, houses four research Units: Structural and Computational Biology (SCB), Gene Expression (GE), Cell Biology and Biophysics (CBB), and Developmental Biology (DB). Most of EMBL's Core Facilities and virtually all of its administration are situated in Heidelberg.
- The two structural biology Outstations in Grenoble, France, and Hamburg, Germany, are located adjacent to large-scale sources of synchrotron and neutron radiation, and actively pursue research in structural biology. Importantly, they also provide the infrastructure and assistance required by the large number of life science users of these facilities.
- The European Bioinformatics Institute (EBI) in Hinxton, UK, designs, builds, maintains and provides data resources and bioinformatics tools to a global user community, and has research programmes aimed at organising and extracting information from biological data.
- The newest Unit is the Outstation in Monterotondo, Italy, which is devoted to the study of mouse biology and provides expertise in mammalian physiology and the production of mouse models of human disease.

EMBL's scientific excellence is ensured by the stringent selection of excellent young candidates for positions as independent research group leaders and by regular expert review of all its research and service activities, organised by the Scientific Advisory Committee (SAC) of EMBL Council. A unique feature of EMBL is that none of the staff receive permanent contracts. The norm is that EMBL's employees stay for a maximum of nine years, and then return to a member state to continue their career. These alumni enrich the scientific communities they join and help to spread the international collaborative scientific culture for which the Laboratory is known throughout the member states. A minimal proportion of EMBL staff (currently 12–13%) has longer-term rolling contracts. These crucial staff are required to ensure the continuity and stability of the Laboratory, to provide mentoring to the constant influx of new, younger staff members and to help meet the many external demands on the Laboratory. The staff turnover system and EMBL's scientific environment, which is both stimulating and demanding, ensure that there is a constant evolution both of research personnel and of their goals.

With a mandate from EMBL Council, the Strategic Forward Look 2006–2015 (SFL) was prepared by the present EMBL Director General and his predecessor in 2003. It represented an assessment of the history of EMBL, its current status, and the way its scientific and organisational future should be planned. The SFL involved intensive discussion throughout EMBL and was overseen by a steering committee that was appointed by EMBL Council and included representatives of Council, the SAC and EMBL's sister organisations, the European Molecular Biology Organisation (EMBO) and the European Molecular Biology Conference (EMBC). Through the SFL, EMBL Council reconfirmed that EMBL has been entrusted by its member states with five major missions:

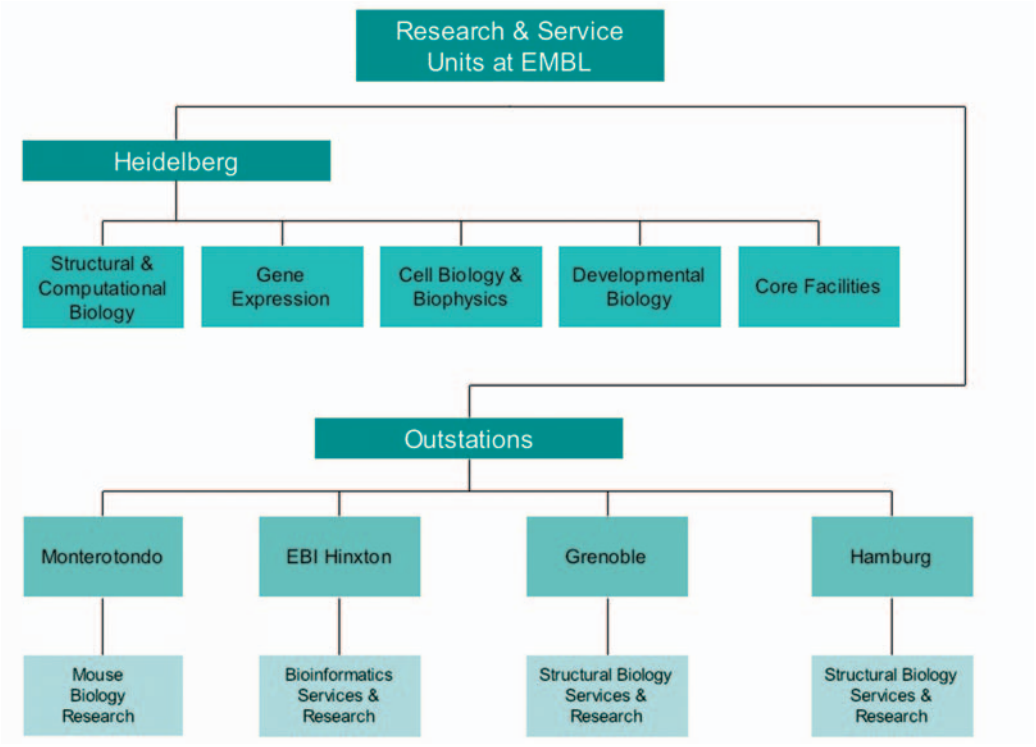
- to be a flagship laboratory for basic, investigator-driven research in molecular biology;
- to develop and help disseminate cutting-edge technologies and instrumentation for molecular biology;
- to provide facilities and services for the scientific community;
- to provide advanced training to individual scientists at all levels, from PhD students to independent investigators;
- to engage actively in developing its discoveries to the point where they can be used by society.

EMBL's Programme 2007–2011 provides a blueprint for the implementation of these principles. We will now summarise the major components of our plans before presenting the Programme in detail.

1. Research

EMBL's primary mission is basic research and, according to external evaluation, its research performance over the past two decades has been outstanding (see Annexes 2 and 3). The initial EMBL mandate was to develop both critical mass and excellence in many facets of molecular biology. This was achieved with a structure that consisted of relatively small independent research groups who enjoyed scientific freedom. The fact that EMBL was able to offer an attractive environment, including scientific independence, excellent working conditions, a collaborative international environment and sufficient resources to allow the establishment of a modestly sized research team, enabled it to attract outstanding young scientists. They themselves quickly realised that working together in a collaborative way greatly extended the scope and impact of their work. Although collaboration is a bottom-up activity, successive administrations also actively fostered interdisciplinarity and communication between EMBL's scientific Units (see Figure B.1. for EMBL Research and Service Units).

Figure B.1. Organisation of EMBL research and services. Please see Annex 4 for a list of EMBL faculty members



Over the 30 years of its existence, EMBL has assembled critical mass in several sub-disciplines of the molecular life sciences, including structural biology, biochemistry, genetics, cell biology, developmental biology and bioinformatics. As discussed in detail in the SFL document, these fields by no means represent a complete coverage of molecular biology, but they were chosen because together they have enabled EMBL to cover the range of biological scale from the atom to the organism, and have provided a framework that has allowed a broad interdisciplinary approach to a number of important biological problems. EMBL excelled in this era and some of the highlights of its research are listed in Annex 5.

The tremendous power of reductionism, enabled by the technological approaches of molecular biology, has allowed detailed understanding of the functions of individual biological macromolecules. This has led to unprecedented knowledge of specific facets of living systems. These insights have brought basic biological research and medicine much closer together, as molecular approaches have explained the defects underlying certain disease states. They have also spawned the biotechnology sector. Together, these changes have made the life sciences strategically even more important for all EMBL's member states. There is a growing international realisation that investment in basic research in the life sciences drives innovation. Unfortunately, in terms of increased research funding, the USA, Japan and some developing countries such as China have reacted much more forcefully to this realisation than Europe.

Together with this change in the social significance of life science research, a revolution in the capacity of researchers to analyse and understand living systems is taking place. This began with, and is partly the result of, the advent of genome sequencing. Because they provide the context for EMBL's decision to focus on systems biology in the future, some aspects of this revolution and EMBL's response to them over the past five years requires more explanation.

To recapitulate, EMBL has a history and considerable expertise in reductionist molecular biology. The usefulness of these powerful approaches for understanding living systems is far from exhausted. However, they have now been supplemented by cross-disciplinary approaches that provide new types of information: functional genomics, quantitative imaging and computational biology. The synthesis of all these diverse approaches with the iterative use of computational modelling and simulation methods in order to develop a detailed quantitative understanding of aspects of biological function is systems biology. This approach is the natural next step for EMBL to take and will form the heart of the next five-year Programme (see Box 1). The Programme builds on the strengths of EMBL's existing Units but will require new horizontal activities to further emphasise interdisciplinarity and to ensure the wide dissemination throughout the Laboratory of the essential expertise.

To achieve this, we intend to build up EMBL Centres. Three Centres will focus on functional genomics, imaging, and computational biology. The Centres will not supplant our current Units, rather they will recruit members from across EMBL in order to spread the use of these technologies, share expertise and help organise ambitious interdisciplinary and inter-Unit research projects. These projects will help the Laboratory face the challenge represented by biological complexity. A further innovation will be to assimilate chemical biology and biochemical instrumentation development, which have formed a separate Unit in Heidelberg, into the four remaining research Units of the headquarters Laboratory. This will promote the seamless integration of technology development with basic research activities.

1 Systems biology

A system, in the original terminology of Ludwig von Bertalanffy, is an entity that maintains its existence through the mutual interaction of its parts. It is these interactions that make the difference between a system and a collection of individual parts. This definition of systems allows the derivation of principles that remain valid at various scales even if the local mechanisms operating at these scales are entirely different. Therefore, "systems biology" concerns the study of the connectivity between the components of living matter and how the connectivity between these components can lead to emergent properties i.e. the properties of a system that are not derivable by summing the properties of its component parts. The approach focuses on collective behaviour and is to be contrasted with classical reductionist approaches that describe living matter by the identification of the parts and their individual properties.

Below we give just a few examples of questions at different scales that require a systems biology approach.

- Protein folding: the attempt to understand the structure of a protein by analysing the collective behaviour of a chain of amino acids.
- Protein and gene networks: the attempt to understand how complex functions (oscillations, switches, etc.) emerge from interactions between genes (through the proteins they encode) or between proteins.
- Cell morphogenesis: the attempt to understand how the collective behaviour of the structural components of cells interact with enzymatic networks to produce diverse cell shapes and forms.
- Tissue morphogenesis and embryogenesis: the attempt to understand how the collective behaviour of groups of cells with specific individual properties interact to produce an organ or tissue. Organs have specific functions related to their organisation that cannot be predicted from the properties of the cells from which they originate.

The essence of systems biology is therefore to understand the dynamic interactions between the components of the system. To achieve this, it is necessary to know the nature and properties of the components involved. Hence, it is often important for systems biology to use "omics" approaches (e.g. genomics, proteomics), data resources and computer simulations in combination with detailed experimental testing. EMBL has gathered much of the expertise required to take a systems approach over the past ten years and wishes to combine and extend these efforts to contribute to the deeper understanding of complex biological phenomena during the next five years.

1.1. Functional genomics

The publication of the complete DNA sequences of living organisms, beginning in 1995, represented a watershed for biologists. First, it marked the realisation that projects whose value to the community was large, but whose realisation required financial and human resources that were considerably above those available to an individual laboratory, could and should be undertaken. Second, it forced researchers who had been used to thinking in terms of single gene products and their functions to consider biological phenomena from a more global, integrated perspective. Third, it introduced robotics and automation to a life science community that had been unfamiliar with the use of these techniques and the possibilities they offer.

DNA sequencing remains enormously important not only for obtaining primary information on genome composition and organisation but also in order to understand genetic diversity and variability and the basis of genetic disease. In addition, a variety of other genomic technologies have been developed, three of which will be described here for illustrative purposes.

- Microarray techniques allow global analysis of the abundance of all the RNAs present in a given cell or tissue sample. This allows unprecedented insight into the differences between cell types in an organism, or the changes in gene expression that occur as a result of alterations in the environment or during the course of the development and differentiation of an organism. Because these methods are so powerful, they are now in widespread use in many EMBL Units.

- Proteomics methods reveal essentially equivalent information for the protein composition of biological samples, although the dynamic range of protein detection methods is not yet as great as for RNA or DNA. The combination of advances in mass spectrometry of protein samples with non-disruptive purification methods, both of which are fields to which EMBL has contributed critical technology developments, allows a picture to be obtained of the interactions that occur between the individual proteins in a cell. Such information is crucial for understanding cellular activity and function.

- High-throughput structural genomic approaches are also greatly accelerating the determination of macromolecular structure. Approaches taken in this area vary from projects that concentrate on a single type of target, such as kinases, to those that attempt to obtain as much structural information as possible on a single organism. EMBL, through its Hamburg and Grenoble Outstations, is actively involved both in the development of technologies that allow automation of structure determination and, more generally, in several targeted structural genomics initiatives.

Such examples illustrate and stress the tremendous change in thinking that these approaches permit. Biologists have been used to thinking in detail about one specific aspect of the process that interests them. This approach has been, and will continue to be, enormously powerful. Now, however, these reductionist methods can be supplemented by information on the global effects on biological systems of, for example, the loss of a specific gene product or the activity of a regulator of development.

1.2. Imaging

Light microscopic imaging has also been undergoing its own revolution. From being a technique that could be used to visualise the subcellular localisation of individual molecules in fixed samples, light microscopy has taken the huge step of becoming the method of choice for visualising *in vivo* dynamics. The use of fluorescent reporters and highly sensitive light detection systems now allows the dynamic movements of individual proteins to be monitored within living cells and organisms and, in a more limited way, the recording of interactions made between proteins. Technologies developed at EMBL over the past few years, including the so-called Selected Plain Illumination Microscopy (SPIM) and Fluorescence Lifetime Imaging Microscopy

(FLIM) microscopes (see Section C), allow researchers to follow the movement, and even the activity, of individual cellular components in time and space in living cells and tissues. Quantitative 4D information gained in this way will be crucial for developing a real understanding of the function of living systems.

Similar to genomics methods, advanced light microscopy relies on a level of equipment and specialist staff that lies outside the capacity of a standard EMBL research group. In order to provide access to these technologies in the Laboratory, EMBL Council approved the plan of the previous Director General to found EMBL's Core Facilities. These came into existence during the last EMBL Programme and have been a tremendous success. They make it possible to maintain our independent research group structure without making it difficult for scientists to use expensive modern technology. For these reasons, the Core Facilities will be further strengthened during the next EMBL Programme, as described in detail in Section C.

1.3. Bioinformatics and computational biology

The flood of biological data and the burgeoning growth of the databases that make these data available to the community is changing the way biologists work. Today, virtually every project carried out in the life sciences is informed and directed to a greater or lesser extent by use of the available data resources. EMBL, through the EBI, is a major global provider of this information. Furthermore, EMBL was a pioneer in the support of Bioinformatics – the science of organising and extracting information from biological data – and is still very active in this field. Biology has entered an information-rich age, but much of the information we have is poorly integrated and almost all is qualitative rather than quantitative in nature, and therefore not always useful in a predictive sense. Bioinformatics research helps turn biological data into information by using evolutionary relationships and the available information on biological systems to integrate new data with existing knowledge.

The use of other computational methods, such as modelling and simulation, has a shorter history at EMBL. However, given the existence of much detailed information on specific biological systems that EMBL scientists study, and the new possibilities to supplement this with global information or quantitative data, we are convinced of the urgency to adopt computational approaches widely as a tool in our research.

Computational modelling can in fact help at every stage of a project, from experimental design through data interpretation to hypothesis generation and testing. Simulations can guide the researcher by defining which are the critical parameters whose quantitative values must be experimentally determined in order to allow predictions to be made about the behaviour of a system. Concordance between experimental and modelling data from the system over a wide variety of conditions provides confidence that it is well understood. Given the complexity and unpredictability of biology, everything points to the need for much more widespread use of computational methods in the future; this is a major goal in the next EMBL Programme with its focus on systems biology.

2. Services to the scientific community

EMBL has four main service activities: the EBI, the Structural Biology Outstations, the Visitors Programme and EMBLEM.

EBI databases are currently consulted roughly 2 million times per day, and provide over 200,000 regular users with access both to the world's core biomolecular data resources and to tools that enable their exploitation. The growth in the volume and diversity of biological data production, and its importance for modern research and innovation in many fields, cannot be overestimated. On the basis of independent external evaluation and advice, we have reached the conclusion that the EBI needs to grow to meet these demands, and we have to work with EMBL's member states to achieve a satisfactory, sustainable funding model for this Outstation.

The provision of access to state-of-the-art facilities for structural biology at the Grenoble and Hamburg Outstations is an important task for EMBL. Together, the two Outstations serve more than 2000 visiting scientists each year by devoting much of their time and effort to user support. A second major focus is in the area of beamline technology development. EMBL has produced leading equipment for the optimal manipulation of synchrotron-derived X-ray beamlines and for the automation of the beamline environment in order to increase the efficiency of use. A goal in the next period is automation of the entire pipeline for X-ray structure determination from sample preparation to the production of a structure model. In Hamburg, the next five years will see an ambitious investment in new beamlines to accompany the transfer to a high-performance synchrotron ring, PETRA-III. Current indications are that most of this upgrade will be made possible by the generous financial support of the host country, Germany. In Grenoble, the Partnership for Structural Biology (PSB) hat was formed during the last EMBL Programme is allowing an integrated approach to structural biology problems through the use of complementary techniques. We will continue to reshape the research programme in Grenoble to take maximal advantage of the new opportunities that this offers.

EMBL's Visitors Programme enables a large number of scientists, more than 1,000 per year, to come to EMBL to learn about and use equipment in its Core Facilities, as well as to carry out experimental work, often in collaboration with EMBL research or service staff. These visitors carry out projects that they would not be able to do in their home laboratory. We devote considerable time and effort to making EMBL an open and welcoming environment for scientists who can benefit from such visits.

3. Technology transfer

EMBL became active in technology transfer in the mid-1990s and founded EMBL Enterprise Management GmbH (EMBLEM) in 1999 as a wholly-owned subsidiary. The goal of EMBLEM is to aid the Laboratory in ensuring that the fruits of its basic research, where appropriate, are used to help innovation in Europe and thus to benefit the member states.

EMBLEM has established a relationship of trust with the research staff and has been very active in increasing EMBL's Intellectual Property (IP) portfolio, in licensing EMBL's IP, and in helping found and develop spin-off companies. It is our view that excellent basic research is the best motor for innovation and we are proud of the rapid progress made by EMBLEM in this regard in its first five years.

4. Advanced training

The most valuable resource EMBL provides to its member states is highly trained scientists. Our turnover system has been in place from the inception of the Laboratory, and has since served as a model for numerous academic institutes around the world. It ensures that young group and team leaders come to EMBL, often from the USA, establish their research programme and reputation at EMBL for a maximum stay of nine years, and then, in 88% of cases, go on to take positions in academia or industry in one of EMBL's member states (see Annex 6). These alumni, trained in excellence, also become part of the large European network of scientists who pass through EMBL. They are schooled in collaborative research and enrich the scientific communities that they join. They are also instrumental in forging links between scientists in the member states and EMBL.

Advanced training has always been a central mission of EMBL. Over the years, several very successful training programmes have been initiated at EMBL, most prominently the EMBL International PhD Programme (EIPP). We are now planning to organise all training activities in the EMBL International Centre for Advanced Training (EICAT) to use synergies and to increase the visibility of EMBL as a major training centre. Building on the enormous success of the EIPP, which currently encompasses 180 students, EICAT will coordinate all existing advanced training activities at EMBL plus several new ones. These include a recently initiated Postdoctoral Programme that will give future independent researchers the leadership and other management skills they need for careers in science. Vocational training will also be offered to all EMBL staff and will be organised in an internal training programme. This will improve the skills of staff members while they are at EMBL and will help them find employment once their time-limited contracts end. Scientists from our member states will continue to have the opportunity to spend time at EMBL as visitors, to collaborate, learn new techniques and use specialised equipment, or to spend a sabbatical. The European Learning Laboratory for the Life Sciences (ELLS) was established three years ago and will continue to organise training courses for teachers in Europe to help bring cutting-edge science to the classroom.

Aside from integrating, and thereby increasing the efficiency of, all our training activities, a major benefit that EICAT will contribute to the advanced training programme at EMBL is the organisation of a comprehensive programme of conferences, training courses and workshops. In this area, we work closely together with our sister organisation EMBO to develop a programme that is leading in Europe. EMBO selects, coordinates and funds a large programme of conferences, courses and workshops every year for a variety of scientists in diverse fields. Although EMBO is not the only source of funding for EMBL's external training activities, it is by far the main one. EMBL scientists, either alone or with colleagues from other scientific institutions, serve as scientific organisers, speakers or teachers at these events. More than 100 scientific meetings have been organised at EMBL in the past four years with more than 9,000 participants. For illustration, Annex 7 provides a detailed list. The current conference facilities are, however, inadequate to support these activities properly and in 2005 EMBL Council agreed to the construction of a Multipurpose Building that, among other things, will enable us to increase the number of participants significantly in our internal and external training courses. Germany agreed generously to fund this building.

Still missing is a suitable auditorium and space for poster display, which would hugely increase the impact of both EMBL and EMBO conference activities. We are currently in discussion with a donor who wishes to support the provision of such facilities on the Heidelberg site. The plan is an ambitious one, involving the replacement of the proposed Multipurpose Building with an integrated Advanced Training Centre (ATC). The model for the ATC is a similar centre at Cold Spring Harbor Laboratory (CSHL) in the USA, which for decades has served as a location for scientific exchange and training on that continent. Aside from CSHL, there are several other major conference series organised in the US each year, for example Keystone and the Gordon Research Conferences. We believe that it is critical to create parallel opportunities in Europe, to

allow European PhD students and postdoctoral fellows, who can rarely afford to attend US meetings, access to cutting-edge research and researchers. If the ATC can be developed as a focal point for regular European-based and -organised conferences, it will make a major contribution to our scientific competitiveness in the life sciences. This topic has already been introduced to EMBL Council to positive resonance, and will form part of our discussions with the member states during the preparation of the EMBL Programme and Indicative Scheme.

5. EMBL's interactions with the member states

EMBL has multiple levels of interaction with the member states. On an official level, representatives of the member states constitute EMBL's ruling body – EMBL Council. Through EMBL Council, and the SAC it appoints, the member states oversee all of EMBL's activities. On a day-to-day level, interactions mainly occur between member state scientists and EMBL. The scientists make use of our service and training facilities, visit our different sites, have joint funding applications with EMBL scientists, and so on. EMBL's scientists also heavily participate in scientific review processes, sit on advisory boards, and take part in strategic discussions on the development of scientific research at both the national and international level. All of EMBL's PhD students are registered at a member state university. This helps strengthen links between EMBL and the academic communities in the member states.

The newest and deepest type of interaction with scientific institutions in the member states – the formation of EMBL partnerships – was instituted during the last EMBL Programme. These are carefully selected formal agreements between EMBL and partner institutions that, although locally funded, enable EMBL's partners to benefit from our expertise in research and training, in promoting internationalism and in institutional organisation.

To be successful, these partnerships must be mutually beneficial. Although EMBL seeks no financial benefit, the partner institution should provide scientific expertise that is complementary to EMBL's in areas that the Laboratory, because of its limited size, cannot pursue. These include research on human disease, including regenerative medicine and human genetics, as well as marine biology, plant biology and a series of other areas.

EMBL has currently established four partnerships (Section F) and is in the process of discussing five more. The popularity of this programme has exceeded our expectations. We intend to pursue the ongoing discussions vigorously and to establish several additional partnerships in the course of the next five years. A critical factor that will need to be considered before expanding this programme further is EMBL's very limited leadership structure, and more administrative support will certainly be necessary if this programme continues to grow.

Another important aspect of EMBL's interactions with its member states is outreach activities. Life scientists today can no longer expect the public to be indifferent to what they do. Following the advent of genetic modification, reproductive technologies, regenerative medicine and the growing ubiquity of biotechnology as a pillar of industry, life science is very much in the public eye. In our view, it is absolutely essential that major research organisations engage in a dialogue with the public. The content of this dialogue should highlight the advances being made in the life sciences, the possibilities that result from these advances and the concerns that might be raised by their implementation. The form of the dialogue is all-important. Outreach should not be viewed as educating the public so that they will agree with scientists, but rather as a two-way discussion in which each party is listening to, and learning from, the other.

We will focus here on two of EMBL's activities in this area. The Science and Society Programme organises, in close collaboration with EMBO, annual symposia on such topics as ageing, diseases of poverty, genetic modification and the genetics of behaviour. These meetings cross many disciplines and attract a broad spectrum of speakers and attendees. In addition, in-house discussions on topics raised by, or of general interest to, EMBL staff will continue to be organised regularly.

The second activity is the organisation of courses for science teachers. There is a very worrying current trend throughout Europe for school children not to choose a science curriculum. In order to help reverse this tendency and to secure the educated manpower required for a knowledge-based economy, children need to be exposed to the most interesting aspects of science at an early age. EMBL, as part of a larger programme coordinated by EMBO, has organised a series of "Life Science Learning Laboratories". These courses help teachers find out what is new in our field and help them design practicals that can work in a school environment in order to illuminate these advances in an interesting way for children. We also intend to remain active, with our other EIROforum partners, both in organising the biannual "Science on Stage" science fair for secondary school teachers and in publishing a new journal for science teachers in Europe, *Science in School*, as part of the largest EU-funded Science and Society network (NUCLEUS).

6. EMBL in the context of a united Europe

Many of our member states are members of the EU. The EU has set itself the target of becoming the world's largest knowledge-based economy by 2010, and its member states have pledged to increase research and development spending considerably to help achieve this ambitious goal.

Although EMBL is not an EU institution, it is actively engaged in promoting European excellence and the European Research Area by providing expertise and access to cutting-edge research and innovation in its field throughout all its member states. EMBL therefore intends to do all it can to help improve Europe's science performance. The ways in which EMBL has worked together with the EU over the past five years and our intentions for doing so in the future are described in detail in Sections C and D.

Taking further opportunity to promote the scientific agenda in Europe, EMBL has joined forces with other intergovernmental organisations that provide infrastructural support to European scientists to form EIROforum, the European Intergovernmental Research Organisation Forum; these other organisations are the European Centre for Nuclear Research (CERN), European Southern Observatory (ESO), European Space Agency (ESA), European Fusion Development Agreement – Joint European Torus (EFDA-JET), European Synchrotron Radiation Facility (ESRF) and Institut Laue-Langevin (ILL). EIROforum wishes to provide its user communities with a voice in the planning of further European infrastructures and in improving strategic planning for the funding of those that currently exist.

The opening of Central and Eastern Europe brings with it a huge challenge to organisations whose mission is European in scale. Several Central and Eastern European countries that are already member states of EMBO and EMBC have engaged EMBL in membership discussions and one, Croatia, has just been accepted as EMBL's 19th member state at the time of writing. These countries share our view that membership of EMBL will be of enormous benefit to their scientific communities and for their development into knowledge-based economies. Just as EMBL has helped establish bridgeheads of molecular biology excellence in those of its present member states whose own communities were least developed when they became members, we see it as our duty to help integrate the "new" European countries as quickly as possible. We will make it a priority to enable as many additional members as possible from this region to join during the next EMBL Programme. Given the traditional strengths of these countries in areas that will be crucial for systems

biology, such as mathematics, physics and computational biology, we see this as a tremendous opportunity for mutual enrichment.

Finally, it is important to stress that EMBL's influence extends beyond Europe. Our reputation for excellence has attracted interest from as far afield as Asia and Oceania, areas of the scientific world that had previously looked to the USA for collaboration and exchange. EMBL Council approved an Associate Membership scheme in 2003 and we expect significant progress towards the adoption of associate member states over the next five years. These new contacts will promote the global cooperation necessary to carry out large-scale research initiatives in the life sciences, and will bring benefits to Europe, in terms of communication and exchanges with distant scientific communities, that extend far beyond EMBL.

c. RESEARCH

The success of molecular biology is a result of its ability to provide mechanistic insight into the functioning of living organisms. Until recently, this insight has come almost exclusively from reductionist approaches, involving the detailed study of individual genes and gene products at a specific scale. A given gene product, say an enzyme, might be studied by structural biologists to uncover its atomic organisation and to probe how it achieves catalysis. Biochemists and cell biologists would be more interested to know how the expression of the gene is regulated, where the gene product is localised in the cell, how its enzymatic activity is affected by the proteins with which it interacts, and how it contributes to cell growth and organisation. Organismal biologists, by contrast, want to know what the gene product does in the context of a whole organism, in which tissues the gene is expressed, and what effect the removal or mutation of the gene would have on organ and tissue development or on complex phenotypes such as behaviour, learning, susceptibility to cancer etc. In most institutions, these sub-disciplines have been pursued separately.

EMBL's past leadership had the foresight to organise the Laboratory such as to enable biological problems to be understood across scales, from the atom to the organism. The first Director General, Sir John Kendrew, was awarded the Nobel Prize for his contributions to structural biology, and ensured that EMBL was active in this field from the beginning. He also recruited biochemists, mainly utilising and developing recombinant DNA technologies, and later, with significant contributions from Kai Simons, he initiated activity in the field of molecular cell biology. His successor, Lennart Philipson, consolidated these areas and, with considerable foresight, added a critical component – bioinformatics – at a time when the field was in its infancy. The databases that were part of this pioneering activity later led to the foundation of EMBL's UK Outstation EMBL-EBI. With some exceptions, the most notable being the Nobel-Prize-winning work of Christiane Nuesslein-Volhard and Eric Wieschaus on *Drosophila* early development, organismal biology was not a focus of EMBL in its first 20 years. The third Director General, Fotis C. Kafatos, encouraged the further evolution of the Differentiation Programme, which was then led by Thomas Graf, to concentrate on developmental biology and initiated a new Outstation in Italy that was devoted to research on the mouse. He quickly realised the value of genomic approaches, and was very active in adapting the Laboratory to enable the adoption of functional genomics as the focus of efforts in the 2001–2005 Programme.

Thus, EMBL was in a position to integrate its research across several levels of biological scale. The staff turnover system, the modestly sized individual research groups and the carefully nurtured collaborative culture all helped this integration and led to the current structure of the Laboratory. We provide two types of documentation of the success that this organisational model led to. The first is a collection of summaries of the reports of the committees that are appointed by SAC to review our research Units individually once every four years (see Annex 2). It is clear that these committees, composed of internationally recognised leading scientists, feel that EMBL's research is world-class and of considerable value to the member states. The second type of documentation comprises the results of recent bibliometric analyses of the impact of published research carried out by independent institutes (see Annex 3). These support the claim that EMBL is one of the world's top research institutions in the molecular life sciences. Although we caution against a narrow interpretation of the bibliometric data, which suffer from numerous imperfections as a measure of research excellence, the fact that EMBL has consistently been evaluated as one of Europe's top biomolecular research laboratories in multiple recent studies supports the conclusions of the SAC review panels that the Laboratory has been performing very well.

The reductionist approaches that garnered these successes will remain of great value to life science research for the foreseeable future, as they are essential to provide detailed insight into biological systems. However, significant developments have taken place over the past decade that mean that these approaches are no longer sufficient to support a cutting-edge research strategy. The first development is the accumulation of information from the many thousands of reductionist projects that have been carried out by the scientific community to date. The second stems from the availability of the sequences of entire genomes and the subsequent development of high-throughput functional genomics technologies. This information provides the context in which new projects should be planned and the new hypotheses formulated, but its use requires

considerable expertise. Biologists can now look at the expression of all the genes in a given cell, tissue or organ and globally monitor the changes that occur when conditions are altered. These developments have made bioinformatics analysis of published information and computational treatment of the data gathered with high-throughput approaches essential skills for molecular biology research groups. An additional factor has been the growing realisation that, in biology, virtually no gene product acts on its own. Rather, it exists in either stable or dynamic interaction with other macromolecules or small metabolites that influence its location, modification state and functional activity.

Interacting systems have what are called "emergent properties" that are not predictable from their individual components. Two often-cited examples illustrate what this means. The properties of hydrogen and oxygen as individual elements provide no indication of the properties of the water that is formed by their interaction. Another example is a thought experiment: if one were to determine the molecular composition of a mouse one second before and one second after its death, the minimal change in composition that would be detected would provide no indication of the fundamental change in the properties of the whole organism. A different approach is needed to understand the properties of complex systems, and the emerging discipline called "systems biology" that encapsulates this approach is the next logical step in EMBL's development (see Section B, Box 1).

Systems analysis presents biologists with an immense challenge. Although the number of human protein-coding genes, at roughly 23,000, at first appeared surprisingly modest, there are many ways by which organisms achieve additional diversity in their number of gene products. Alternative splicing allows many human genes to produce several functionally distinct protein products. These proteins can be differentially modified by the addition of chemical groups such as phosphates and thus have their activity further diversified. Most importantly, however, the fact that biological function depends on combinatorial use of gene products means that 23,000 genes can give rise to an incomprehensibly large array of different functions. Even if an infinitesimal fraction of the possible combinations of our gene products produces ensembles of distinct function, there is ample material from which to generate a complex organism. Beyond such combinatorial considerations, there are issues related to the difficulty of predicting how thousands of molecular species will interact dynamically to produce a shape or functional module. The examination of such behaviour is thus not likely to be approachable solely by experiments and intuition. The careful use of computer models and stochastic simulations will be unavoidable, and this combination of computation with experimentation is at the heart of what we mean by systems biology.

In light of the above, it will come as no surprise that EMBL's priorities in the future are designed to help our research groups tackle biological complexity in an interdisciplinary way. The recently implemented concept of EMBL Centres is aimed at allowing us to develop and promote the most important sets of expertise that we need to complement our existing disciplinary skills. The Centres will coordinate training and disseminate expertise in the use of functional genomics technologies, quantitative imaging at different scales, computational analysis and research related to disease. Core Facilities will provide technical support, ensuring that each research group has access to a range of experimentation that extends far beyond its individual expertise. Finally, and centrally, we will pursue the analysis of functional, structural and regulatory networks in biology at the systems level by combining modelling and simulation approaches with "wet-lab" experimentation in an iterative manner. We believe that it is currently premature to produce detailed models of entire eukaryotic cells or organisms. We currently lack information on many aspects of such complex entities, for example their complete parts list or the spatial organisation of their components. Gathering this information has begun, but will take time. However, the time is ripe to take a systems approach to functional modules, individual subcellular structures and even organelles. Detailed examples will be described in subsequent sections.

1. Major research themes

1.1. Structural biology at EMBL

Today, at the beginning of the 21st century, important scientific challenges exist in the field of structural biology. They include the resolution of the structures of large functional complexes using microcrystals, the automation of data collection and structure determination, the linking between computer modelling and structure determination, and the bridging between structural and cell biology through the combination of structural techniques such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, electron microscopy (EM) and light microscopy.

Why is EMBL in such a good position to address all of these challenges? First, it has the critical mass required for the integrated structural approaches that are necessary to pursue ambitious projects at the cutting edge of current life science research: three of EMBL's Units have structural biology as a major focus; the Structural and Computational Biology (SCB) Unit in Heidelberg and the Outstations in Grenoble and Hamburg. Each of these Units is unique, in part owing to differences in the focus of the research groups and teams that compose them and in part because they are embedded in different environments and are designed to take optimal advantage of the activities surrounding them. The features of each Unit will be described below. The two Outstations are located next to synchrotron facilities where much of today's atomic resolution structural biology is conducted (see Annex 12). They are engaged in the development of new techniques to study smaller and fewer crystals, robotics for automated data collection, and software for automated protein structure determination. The SCB Unit in Heidelberg is perfectly positioned to gather structural insights into the cellular machines that are investigated functionally by the surrounding Units. The physical proximity of the SCB Unit to research that employs lower-resolution imaging technologies and its strong bioinformatics component makes it possible to work towards the ambitious goal of integrating these disciplines towards an ambitious, long-term goal: cellular images with atomic resolution.

The different Units will continue to develop, implement and provide complementary technologies in the above key areas of structural biology. The fields of activity will include the development of hardware and equipment, as well as new software packages and automated experimental pipelines. These activities will allow EMBL to retain a leading position in a scientific area that is strongly associated with innovative technology development.

Crystallographic applications in molecular life sciences are enormously varied and recent developments have reached towards largely unexplored areas such as integral membrane proteins, large multi-component protein complexes and even partly unfolded proteins. It is therefore not surprising that X-ray crystallography has evolved into the strongest activity that is shared by the three structural biology Units. The Units in Hamburg and Grenoble will, over the course of the next five years, both be situated at world-leading synchrotron and laser facilities, and ideally complement each other. Whereas Hamburg has taken a major responsibility in the independent building and provision of beamline facilities, the Grenoble Unit has become a major provider of state-of-the-art technology, know-how and research experience for the ESRF beamlines. The Units have complementary activities in technology development. Hamburg has become a world leader in the development of novel software packages for automated data interpretation, whereas Grenoble is playing a leading role in the development of advanced beamline equipment such as automatic sample mounting and diffractometer systems. The Unit in Heidelberg combines X-ray crystallography with other technologies including NMR, EM tomography, chemical biology and biochemistry. The three structural biology Units are key partners in several large-scale European projects such as SPINE, BIOXHIT (coordinated by V. Lamzin, Hamburg), SAXIER (coordinated by D. Svergun, Hamburg) and 3D-Repertoire (coordinated by L. Serrano, Heidelberg). Although the Hamburg and Grenoble Units are mostly focused on synchrotron radiation-associated structural biology activities, the Heidelberg SCB Unit is in a unique

position to combine and integrate a broad range of available methods, frequently in collaboration with scientists from other programmes at the headquarters Laboratory.

Taken together, the capacities of these Units will enable EMBL to tackle the most challenging frontiers in structural biology in the next five years: structural genomics and automation, integration of X-ray crystallography with a variety of other methods including NMR, small-angle X-ray scattering (SAXS), cryo-EM, EM tomography, biochemistry and bioinformatics. This integration is being pursued in different ways at the three EMBL sites. Proper coordination of the activities will be maintained by a steering committee involving the Heads of these Units and the Director General, complemented by regular bilateral or trilateral project-oriented meetings of joint interest groups. For instance, there is an ongoing series of bilateral meetings of the beamline instrumentation groups in Hamburg and Grenoble. In addition, similar joint activities involve staff from Heidelberg, Hamburg and Grenoble that are involved in technology development to tackle one of the critical remaining challenges: the efficient preparation and characterisation of samples for structural analysis.

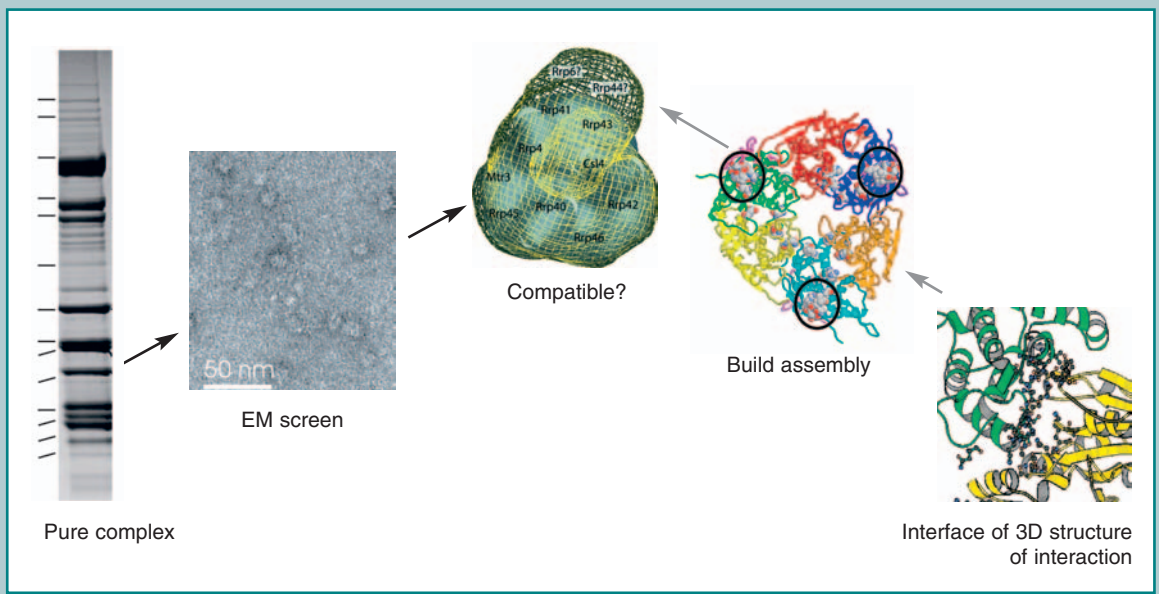
1.1.1. Structural analysis at different functional levels

EMBL's structural biology groups will focus on the following major research tasks:

- **Solution of biomolecular structures to help illuminate important biological problems.** These projects are often carried out in collaboration with research groups from other EMBL Units. Fields where we will be active include: chromatin and transcription; RNA metabolism (processing, degradation, translation and its regulation); intracellular transport, including transport of RNA and proteins between the nucleus and cytoplasm and selected aspects of membrane trafficking (endocytosis, vesicle transport); analysis of regulatory molecules involved in signal transduction and in cell organisation (e.g. mitotic kinases and their substrates, titin in muscle cells, glycosyltransferase enzymes that modulate signal transduction); and cytoskeletal proteins involved in cell morphogenesis.
- **Structural analysis of complex molecular machines.** It is abundantly clear that proteins usually do not function on their own, but instead are part of large macromolecular complexes. Structural work in all of the defined research areas listed above needs to take account of this fact (see Box 2). Each of the three structural biology Units has certain advantages in adapting from the analysis of single molecules to that of complex molecular machines. One of the most crucial is the ability of the structural biology Outstations to design and build X-ray beamlines capable of supporting the most exacting projects (Section D), but there are several others such as the integration of experimental data obtained at different levels of resolution to build structural models of large complexes.
- **Obtaining images of cells at atomic resolution.** This is a complex project, with many facets, but in essence relies on obtaining images of cell sections by EM tomography with high enough resolution to enable the modelling of X-ray and cryo-EM structures into the images. Gaining an understanding of the overall organisation of cells at a detailed level is an extremely important aspect of systems biology because the activity of many important functional complexes within the cell, as well as the consequences of their activity on cell behaviour, is critically dependent upon where they are located with respect to other cellular components with which they interact and on which they act. Thus, spatial information will have to be gathered and built into functional modelling approaches.
- **Integrating chemistry into structural biology.** The objective is to map the binding of small cell metabolites to the different protein complexes in an organism, and then to extend this map to include a large set of chemical compounds. The project requires computational tools to analyse and model structures and their dynamic behaviour in the context of protein interactions.

2 Large-scale characterisation of protein complexes and their interactions

Proteins do not function alone, but as components of large macromolecular machines that assemble into functional complexes in the cell. Many EMBL groups that use structural, biochemical or computational approaches are already involved in the characterisation of functional protein complexes. Several of them have now joined forces in an ambitious collaborative project that involves EMBL scientists from Heidelberg, Grenoble, Hamburg and the EBI, and aims to accomplish the genome-wide purification of as many of the stable complexes as possible from the yeast *Saccharomyces cerevisiae*, and their analysis by cryo-EM. Purification will be carried out using the tandem affinity purification (TAP)-tagging methodology developed at EMBL and will rely on the extensive experience gained in collaborations involving research groups and the EMBL spin-off company Cellzome, which have been ongoing over the past five years. The cryo-EM data will be combined with tomographic data obtained by analysis of cell sections, both directly through structural docking methods and indirectly through EM-aided large-scale complex modelling using computational approaches. It will also be possible, by bioinformatic analysis of large-scale protein–protein and complex–complex interaction data, to incorporate information on cellular interaction networks. A project of this ambition is unlikely to be completed within the period of a single EMBL Programme. It is however reasonable to expect that the analysis will have reached the point in five years where most, if not all, yeast complexes are well defined in terms of their composition. Furthermore, many of the interactions between these complexes will have been identified. The susceptibility of these complexes to purification and structural analysis, essentially determined by their stability and abundance, will be known, and good-quality cryo-EM data should be collected on all complexes that can be purified in sufficient quantity over the course of the next five years.



Hybrid methods for complex structure determination

1.1.2. The organisation and integration of structural biology at EMBL

We should now turn to the organisational plans of the three structural biology Units. As already mentioned, these have been worked out after extensive consultation. The SCB Unit in Heidelberg has, for some time, been able to take an integrated approach to structural problems as a result of its expertise in three major structural techniques (X-ray crystallography, NMR, cryo-EM) and the availability of computational approaches that bridge between platforms and scales. This will continue, although the new focus on EM, including tomography, and the availability of strong X-ray crystallography groups in Grenoble and Hamburg, means that there will be one fewer X-ray group in Heidelberg. Furthermore, although it would be a mistake to lose all capacity in NMR from the Unit, EMBL will continue to be unable to invest in top-class NMR equipment and is thus unlikely to recruit a group leader whose research depends on access to such instrumentation. However, EMBL's record shows that, even without the very best equipment, NMR can still be a powerful method to answer structural questions. In addition, there are other possibilities of using NMR that would interest us. These include recent developments in the use of NMR for *in vivo* analysis of metabolites and their turnover and micro-magnetic resonance imaging (MRI). The latter technique involves non-invasive imaging of embryos or small organisms at the single-cell level and could be a very interesting complement to light microscopy for developmental biologists. The current NMR group is scheduled to leave in the near future, and we will carefully analyse these options before making a new appointment.

Integration of chemistry will take place in the Heidelberg Unit in replacement of the X-ray crystallography group mentioned above. A team leader has just been recruited who will work on the identification of small molecules that will interact with, and disrupt the function of, macromolecular complexes. In addition, the computational groups of the Unit launched a common project to investigate various aspects of the interactions between chemicals and proteins in order to support the integration of chemical knowledge into structural biology. Recent activity in the SCB Unit in computer modelling and prediction of protein complexes will contribute to the mapping of protein complexes into EM tomographic images of entire cells. Within an EMBL-wide collaboration, those static pictures could then be combined with dynamic images acquired by light microscopy. This kind of detailed spatiotemporal information has been virtually absent to date and will be a crucial aspect of systems biology.

In Grenoble, the founding of the Partnership for Structural Biology (PSB) (Section D, Box 19) involving EMBL, ESRF, ILL and the Institute for Structural Biology (IBS), as well as the ongoing active collaboration between EMBL Grenoble and the Institut de Virologie Moléculaire et Structurale (IVMS), has not only led to the construction of the new, jointly occupied Carl-Ivar Brändén building for high-throughput structural biology but also to greatly increased cooperation between the partners. This includes much easier access for researchers at EMBL Grenoble to NMR and cryo-EM expertise located in the IBS and IVMS. These activities add to the previous collaborations with the ILL on biological neutron scattering and with the ESRF in the context of the Joint Structural Biology Group (JSBG), whose role is to design and build beamlines for structural biology and to support the visiting users of these beamlines. Thus, local collaboration can now open the way to an integrated structural biology attack on difficult research projects. The 2005 SAC review panel advised EMBL Grenoble to take advantage of this improvement in the local environment by broadening its expertise over the period of the next EMBL Programme through recruitment of a team leader in cryo-EM and a group leader (to replace Rob Ruigrok) with expertise in biochemical and cell biology methods [e.g. RNA interference (RNAi), light microscopy]. Our ability to implement these proposals will depend on a modest increase in the Outstation budget.

The Hamburg Outstation has considerable expertise in applying X-ray crystallography, SAXS and X-ray absorption spectroscopy (EXAFS) to tackle problems in the life sciences. Owing to recent developments that have occurred as a result of the research and development activity of the Hamburg Outstation, SAXS is increasingly recognised as a method of great utility in the determination of the shape of macromolecules. It is complementary to and, in some cases, can substitute for cryo-EM. SAXS is even more suited to large complexes, which will be extremely valuable given the focus of EMBL's future research activities. We expect this, together with the construction of new beamlines at Hamburg (Section D), to make a huge difference to the number of applications of SAXS methods.

Software developments across the structural board have become, and will remain, a central feature of research activity of the Hamburg Unit. The ARP/wARP suite of programmes for automated structure model building from X-ray diffraction data are very widely used and, like the ATSAS package for SAXS applications, were considered to be world-leading when evaluated by the last Outstation review panel (see Annex 2). New software developments include BEST (to help in the design of X-ray crystallography experimental strategy) and AUTO-RICKSHAW (for pipelining structure determination), as well as software for automated crystal detection in crystallisation drops and in the X-ray beamline. These will be brought to maturity and be ready for wide distribution during the next EMBL Programme.

Both structural biology Outstations are in the process of implementing high-throughput facilities for structural biology. These facilities will become an integral part of the technology platforms for structural biology. Know-how, technology and staff will be transferred between the Outstations during the construction of these platforms. Much of this work occurs in the context of European-scale and national projects (e.g. SPINE, BIOXHIT, XMTB). These provide significant external funding but bring with them a requirement for investment from EMBL. High-throughput structural biology is costly in consumables and this will need to be taken into account in planning the Outstation running budgets.

Both Outstations are active in disease-related research. Grenoble has both a considerable history and current interest in structural virology, including work on Adenoviruses, Epstein–Barr virus, Ebola virus, influenza virus and HIV. Hamburg is coordinating a project with the objective of combating tuberculosis (TB) through the characterisation of new potential drug targets using X-ray crystallography, high-throughput screening and *in vivo* validation procedures (Box 3). This project is global, with complementary activities in Europe, USA, India and New Zealand, and involves academic and industrial partners. This and related future projects are expected to be greatly facilitated by development of the high-throughput infrastructures and thus will expand in the 2007–2011 period.

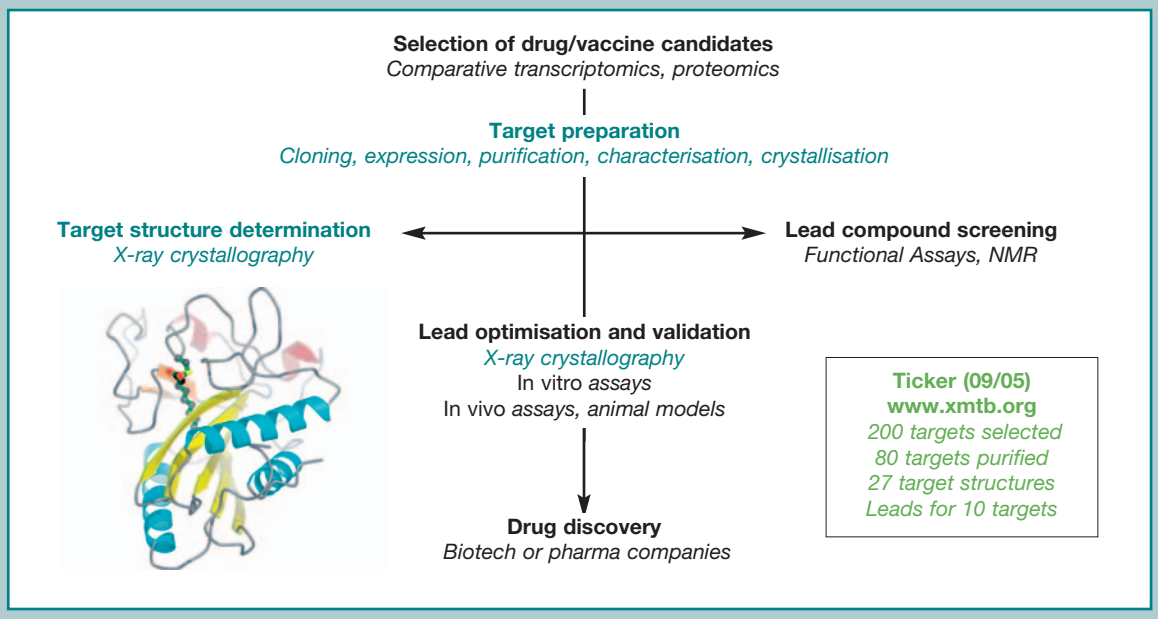
1.2. Molecular, cellular and organismal biology

The next broad area of research at EMBL that we will consider encompasses the levels of scale from the molecule to the organism, and represents four of our Units: Gene Expression, Cell Biology and Biophysics, and Developmental Biology in Heidelberg; and the Mouse Biology Outstation in Monterotondo. The research groups that are engaged in this research are listed in Annex 8. By design, these Units do not have sharp boundaries but instead have both complementary and interlocking activities in the interest of promoting inter-Unit interaction and collaboration. They have traditionally employed a mix of approaches including biochemistry, light microscopy and EM imaging, genetics and recombinant DNA technologies, although the proportionate use of these methods has been different in each Unit. Over the period of the last EMBL Programme, all of the Units adopted large-scale functional genomic approaches in a subset of their projects. First steps were also taken towards incorporating the computer modelling and simulation approaches that will be increasingly important for our move into systems biology. They are therefore well prepared for the next phase of research in molecular biology, where broad, genomic-level approaches will be

3 Structural genomics to improve human health

High-throughput structural biology methods, particularly for X-ray crystallography, have advanced enormously during recent years. The advent of third-generation synchrotron facilities, the ongoing development of fully automated pipelines for sample preparation, characterisation, crystallisation and structure interpretation, and the development of established and emerging methods, such as SAXS, high-resolution EM and EM tomography, have supported this advance. In different parts of the world, including Europe, large-scale structural genomics consortia have formed over the past five years.

The Hamburg Unit coordinates the *Mycobacterium tuberculosis* XMTB structural genomics project, funded by the Bundesministerium für Bildung und Forschung (BMBF), which is aimed at targets from *M. tuberculosis*, an organism with about 4,500 genes. *M. tuberculosis* was chosen in the light of the devastating health situation caused by TB in many countries of the developing world, aggravated by a major and increasing problem of multiple drug resistance and HIV/TB co-infection. Current anti-TB drugs originate from developments that occurred more than three decades ago. Thus, there is a strong mandate to the research community, as well as to non-governmental organisations and industrial enterprises, to carry out innovative research to help understand the underlying molecular basis of TB infection and to develop novel, rationally designed drugs to combat the disease. Within the XMTB consortium, targets are chosen that have significant potential to become useful for drug discovery, vaccines and/or diagnostics. Key tools of the consortium are: high-throughput X-ray crystallography for rapid structure determination, structure-based screening for target-specific lead compounds, and *in vitro* and *in vivo* validation. Within the first 18 months of its existence, the consortium has been able to purify more than half of the 200 targets chosen for its first phase and to determine 25 X-ray structures. Three targets have already been selected from these as a basis for future drug discovery efforts. The project will continue to elucidate structures and functions of as many potential drug targets as possible over the coming years.



Flowchart of the XMTB structural proteomics project demonstrating how structural biology techniques have been integrated into overall project objectives, structure-based drug discovery

combined with detailed analysis of individual functions and dynamic processes to obtain a deeper understanding of the properties of biological systems and their malfunctioning in disease states.

Across such a broad domain of biology, and such a large fraction of the Laboratory, it is only natural that we expect multiple research themes to be pursued over the next quinquennium. We will now describe some of these.

1.2.1. Molecular analysis

EMBL's research on the functional analysis of individual biological molecules and molecular complexes is centred on the Gene Expression Unit. However, the interest in molecular level analysis among our research groups goes far beyond the boundaries of this single Unit, as reflected by the large number of EMBL group leaders in the other Units with joint appointments in Gene Expression (see Annex 9). The underlying rationale is that biological mechanism is often best understood at the level of intermolecular interactions and their regulation. At the root of many of the projects pursued at EMBL are changes in the expression of a set of gene products and regulation of their interactions.

The pathway from gene to protein is a complex one involving changes in chromatin organisation, transcriptional activation or repression, RNA processing and turnover, RNA export from the nucleus, and protein translation. Each of these individual steps can be regulated and the expression of any individual gene is often regulated at several distinct steps on the pathway. Even after the protein product of a gene is produced, its activity is frequently regulated by post-transcriptional modification, controlled degradation or re-localisation within the cell. Most genes are subject to the effects of several *trans*-acting regulatory factors that are themselves subject to spatial and temporal expression control. Elucidating the logic of the interacting regulatory networks that control gene expression in response to changing conditions is a major goal of systems biology, and one that can realistically be approached over the next five years. Since so much of what is interesting for modern biology incorporates analysis of the processes involved in gene and genome expression, it is important that EMBL maintains the critical mass and diversity needed to study these processes at all steps in the expression pathway, as well as to understand their organisation within the context of the cell and the organism.

One broad theme under study will be the organisation of chromatin domains and the mechanisms by which genes are expressed and DNA is replicated, repaired or reorganised within the context of different organisational states of chromatin (Box 4). These studies sit at a particularly interesting interface, between the spatial organisation of the nucleus and the molecular mechanisms by which the genome is accessed and expressed.

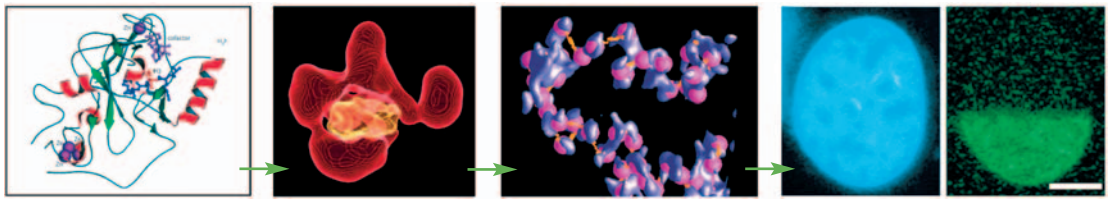
Among other things, chromatin organisation ultimately serves to alter gene transcription. Transcription mechanisms themselves will be studied on a genomic scale in various projects at EMBL (see below). In addition, a novel aspect of transcription regulation was recently discovered here following detailed study of an individual, oestrogen-dependent gene and will be further analysed. Factors that regulate this gene were found to cycle continually on and off its promoter even during periods when the gene is nominally activated. Further study of this dynamic process of transcriptional regulation is clearly warranted, and should provide more insight into the biological logic behind this unexpected behaviour. It may be that rapid on/off cycling is a requirement for all metazoan genes whose activity must be acutely regulated in response to environmental signals. Given the medical importance of hormone-dependent gene expression, the discovery of this novel mechanism also provides a new avenue for drug discovery.

Although transcription control is important, work over the past two decades in which EMBL groups played a leading role has underlined the importance and variety of regulatory mechanisms that target

4 Chromatin modifier/remodeller – from atomic structure to functional dynamics *in vivo*

In order to fit into the cell nucleus, the extremely long DNA molecules that make up chromosomes must be compactly folded and elaborately packaged in chromatin. This raises the question of how the activities that repair, modify, replicate and transcribe the DNA gain access to their substrate. This project is a concerted effort to understand how chromatin modifying and remodelling complexes, which regulate DNA access, interact with chromatin templates at different levels of resolution. The project will integrate atomic structures of individual proteins and protein complexes both on their own and together with mono- or oligonucleosomes, EM structures of larger complexes associated with defined oligonucleosomal templates and fluorescence microscopy imaging. Structural information at atomic resolution will be used for the rational design of mutant proteins with perturbed function. EM and light microscopy imaging will enable analysis of the changes in chromatin and chromosome structure brought about by these complexes. Dynamic measurements of chromatin modifier/remodeller functions in living cells and embryos by fluorescence microscopy methods will provide insight into the turnover of activities at sites where chromatin modification is occurring. Fluorescent reporters will be used to study the function of chromatin complexes at the different levels of analysis. Importantly, the use of genetic model systems will provide direct tests of the function and dynamic behaviour of mutant proteins in single cells or in a living organism.

Groups from four EMBL Units are involved in the project. Systems currently under investigation include ISWI-containing remodelling complexes and the Polycomb group complex Enhancer of Zeste. The fluorescence microscopy aspect of this project will interface directly with work on the simulation of reaction diffusion networks planned elsewhere at EMBL.



Different scales of chromatin modifiers/remodellers. From left to right: atomic structure of a SET domain of a histone methyltransferase, 3D EM structure of the Rsc/nucleosome complex, chromatin organisation of nucleosomes and DNA by 3D cryo-EM tomography and photoactivation of a GFP-tagged histone methyltransferase in a live cell

RNA. Study of mRNA metabolism will continue to be a focus of the Gene Expression and Developmental Biology Units and both the Grenoble and Heidelberg structural biology Units. Major themes will be translational regulation and fidelity, mRNA turnover, mRNA localisation, and the mechanisms of regulation of mRNA expression by short interfering (si)RNAs and microRNAs. Analysis of the role of microRNAs in *Drosophila* development is ongoing as part of a major effort to develop bioinformatic and experimental methods to define the functional roles of microRNAs in animal development and physiology. Our goal is to continue to elucidate the mechanisms that underlie post-transcriptional steps in the regulation of gene expression and to integrate this knowledge into quantitative modelling of specific biological systems.

1.2.2. Cell organisation

The processes discussed in the previous section involve the characterisation and functional dissection of dynamic multimolecular protein or nucleoprotein complexes. This understanding of biological processes is at the level of organisation that the structural biology Units of EMBL will also be tackling in the next EMBL Programme. The next higher levels of organisation – macroscopic subcellular structures and whole cell morphogenesis and dynamics – will also be very actively studied at EMBL.

This level of organisation requires a change in the way we think about biological systems. Although we easily grasp how groups of molecules with complementary structural properties can interact to form a multimolecular complex, we are much less accustomed to thinking in terms of how dynamic, large-scale cellular structures such as the nucleus, the cytoskeleton or the endomembrane systems emerge and acquire both their individual identity and a well-defined shape and organisation. The size and complexity of these structures relative to the molecular scale, their dynamic properties and the maintenance of their overall organisation in the face of continuous and frequently rapid turnover of their individual components pose challenges that require a new set of tools – conceptual, computational and experimental. Understanding the origin of subcellular and cellular shape is of utmost importance because this is the level at which we might start to grasp the essence of the physicochemical principles that confer order to living matter. Moreover, a good understanding of the principles that govern cell morphogenesis and physiology is also important to understand the next higher level of organisation in living systems – the multicellular assemblies of tissues and organs that form in the developing embryo.

In living matter, dynamics is of the essence. At the subcellular level, there are distinct sources of dynamic behaviour:

- Stochastic effects that are powered by thermal motion (Brownian movement). Temperature, viscosity and diffusion play important roles in such motion.
- Deterministic effects, driven for example by molecular motors or various types of polymerisation and de-polymerisation processes. These are powered by the consumption of high-energy phosphate, usually in the form of ATP or GTP.

Obviously, no life is possible without energy-dependent deterministic processes. Stochastic interactions of molecules of well-defined shape lead to the formation of structures that are close to thermodynamic equilibrium. Deterministic motion can drive processes that are far from thermodynamic equilibrium. Both processes act in concert to define the state of living matter and lead to the self-organisation of complex dynamic structures inside the cell.

The movement of specific biological molecules can be visualised and quantified with new methods such as fluorescence correlation spectroscopy (FCS), fluorescence resonance energy transfer (FRET)

and high-resolution 4D light microscopy. EMBL has been heavily involved in the development and use of such methods.

At the multicellular level, dynamics results from the combination of autonomous cell behaviours that themselves result from the dynamics of cell-internal elements like the cytoskeleton. Extracellular signals can trigger intracellular signalling pathways that orient cytoskeletal dynamics and produce organised collective cell behaviour. The production and control of such supracellular processes are key to embryogenesis and organogenesis. The development of new microscopes such as the single plane illumination microscope (SPIM) at EMBL will play an important role in the detailed analysis and quantitative characterisation of such processes.

In order to make progress in this new direction, the Cell Biology and Biophysics Unit has been restructured over the past five years. Traditionally, this Unit had been concerned with the molecular composition of subcellular structures. Although this remains important, new technologies have made it possible to begin to study the behaviour of cellular molecules in dynamic ways over time and in 3D. The recent introduction of physicists and biophysicists into the Unit has enabled first examples of the use of mathematical modelling and computer simulation coupled tightly to experimental approaches. A strong focus in the future will be on interdisciplinarity and particularly on the study of the physical and dynamic properties of the various compartments that constitute the cell, and of how such properties might explain the emergence of dynamic cell organisation.

The structure and dynamics of the microtubule cytoskeleton and its relation to cell shape and organ development will continue to be a focus of interest in the next five years. Dynamic molecular interactions are being examined using new technologies such as FRET or FLIM (fluorescence lifetime imaging) that allow the study of biochemical reactions *in vivo* or in complex cell extracts in real time. This has led to a new understanding of the nature of signalling pathways, how they are organised in space and time, and how they are connected with cell shape. Combining genetics and quantitative light microscopy will lead to new insights into the morphogenesis of cellular compartments, including the cytoskeleton, in whole cells and tissues.

The work on cell morphogenesis suggests that the emergence of complex dynamic shapes at the scale of tens of microns involves fundamental principles. Clearly, stereospecific recognition between molecules leads to the formation of fairly stable cellular domains (like chromatin, membrane domains, supramolecular machines). Other structures such as the cytoskeleton self-assemble but are kept out of thermodynamic equilibrium by consuming chemical energy. Both stereospecific interactions and self-organisation processes are regulated and coordinated by enzymatic networks that are spatially organised. An overall picture emerges in which reaction–diffusion processes based on enzymatic networks coupled to the self-organisation of structural elements might produce the diversity of cellular forms seen in nature.

A second fundamental problem addressed broadly in the Gene Expression and Cell Biology and Biophysics Units concerns the temporal control of nuclear envelope and microtubule cytoskeleton dynamics during the cell cycle (Box 5). We have made great strides in the past five years in these areas. Looking forward, we aim to provide insight into how these regulatory mechanisms are converted into changes in nuclear and microtubule organisation (i.e. the molecules and mechanisms directly involved in such changes).

5 Systems biology of the cell division cycle

Understanding how cells grow, segregate chromosomes and divide has important implications for comprehending proliferative diseases such as cancer. This understanding will also provide insight into fascinating problems of dynamic cellular organisation such as the mechanisms by which nuclei or mitotic spindles assemble in the right place and at the right time during cell division.

Over the past years, several groups at EMBL have made important contributions to the understanding of chromatin organisation, nuclear envelope assembly and organisation, and nucleo-cytoplasmic transport during interphase. The molecular and physical principles responsible for spindle assembly have also been largely charted, in part by EMBL researchers. The belief that drives this work is that fundamental principles of cell organisation can be understood while studying the spatiotemporal system that rules the cell cycle.

In the future, the Laboratory will aim at drafting a comprehensive list of the key elements involved in nuclear and spindle assembly. The rules that govern the interactions between these elements and the kinetic parameters associated with key interactions will be defined. Models and computer simulations based on such data will be built to establish a logical link between the individual and collective properties of the elements. This should lead to improved understanding of how complex dynamic structures like the mitotic spindle emerge from a multitude of local dynamic interactions. At the same time, an intense effort will be made to understand how one kinase, Cdk1, can transform the properties of the cytoplasm from a state where a nucleus assembles (in interphase) into a state in which a spindle forms (in mitosis).

This work will involve groups from throughout the Laboratory and a variety of technologies. These will include large-scale visual/functional screens for the identification of components, molecular cell biology, large-scale high-sensitivity mass spectrometry, biophysics and dynamic light microscopy, EM, structural biology and bioinformatics. Finally, various computer simulation methods ranging from stochastic to deterministic reaction–diffusion equations will be used. Most of these approaches have already been developed or initiated at EMBL and significant progress is expected over the next period.

A common feature of the work on the nuclear envelope and the cytoskeleton is the complete rearrangement that these structures undergo during cell division in metazoa. Whereas the nuclear envelope breaks down and disperses, the microtubule cytoskeleton undergoes an even more remarkable transition. The cytoskeletal components that have given shape to the cell and have allowed motor-driven transport through the cytoplasm during interphase are dismantled. They are then rebuilt into a completely different structure – the mitotic spindle. This cell-cycle-dependent reorganisation is a prime example of systems behaviour, where essentially the same components are used to make dissimilar structures that carry out very distinct functions under two different regulatory regimes (see Box 5).

The nuclear envelope and microtubule cytoskeleton are regulated spatially and temporally. Temporal regulation occurs downstream from the cyclin-dependent kinases that drive the cell cycle. Work carried out at EMBL during the last Programme revealed that spatial regulation of the formation of the spindle and nuclear envelope is at least in part due to the small GTPase Ran. However, how cell organisation and function is regulated during the cell cycle is a major problem about which we still know very little. This problem will be tackled by several EMBL groups in the next five years.

Methods employed will include assessing the effect of genome-wide gene depletion on cell-cycle progression in cultured human cells, using a combination of high-throughput, high-content, automated visual screening in the light microscope and siRNA-based gene silencing. Once set up, this combination will be used for several other screening projects. Also useful will be RNAi screens in whole organisms, particularly where cell division can be easily monitored, such as in *Caenorhabditis elegans* embryos.

These screens should provide us with a number of new gene products with roles in cell-cycle progression. Some of these will probably be new cell-cycle-regulatory molecules, whereas many are expected to be the molecules and complexes whose direct action is responsible for the morphological and organisational changes observed as cells progress though the division cycle, and which will be activated and inactivated at specific cell-cycle stages.

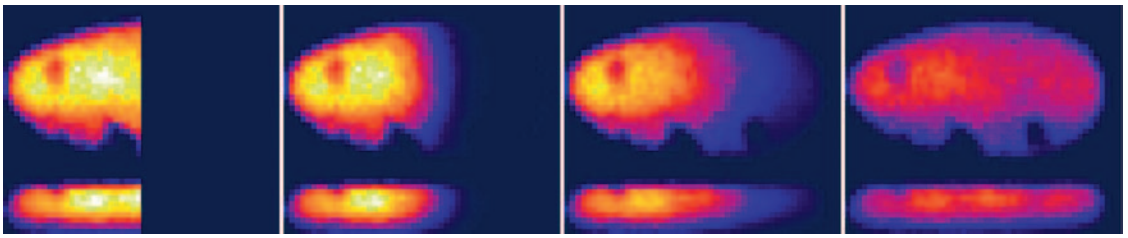
Activation/inactivation cycles in biological contexts are generally achieved by one of three mechanisms: a change in the synthesis or degradation rate of an RNA; a change in the modification state of a protein (e.g. by phosphorylation or ubiquitinylation); or a change in the association of a particular protein or RNA with a functional macromolecular complex. The latter type of change often requires one of the former two, but can also be brought about by other mechanisms, for example when nuclear envelope breakdown during mitosis allows molecules that are separated from each other during interphase to interact with one another.

Many of these changes are most easily detected by mass spectrometry (MS)-based proteomic methods and the demand for access to these in EMBL is high. Although EMBL was very active, and successful, in introducing protein purification (TAP-tagging) and protein analysis methods that enormously broadened the use of MS in biology, it currently does not have state-of-the-art equipment for proteomics. In particular, great potential benefits will accrue to many EMBL researchers if we are able to acquire new equipment to increase the sensitivity of protein detection and enable quantitation of protein levels in experimental samples. In terms of technology development, it is clear that current methods for detecting protein modifications are neither sensitive enough nor amenable to use at high throughput. For the analysis of cell-cycle progression, and many other projects, both sensitivity and rapidity of identification are highly desirable. Given the fact that our current MS research group will turn over in the near future, together with the desirability of progress in this area, we intend to search carefully for someone working on technology development for the analysis of protein modifications.

6 Functional imaging: from cell to organismal biology

Advanced light microscopy is a key technology to study intact biological systems quantitatively and is thus a prerequisite for systems biology. In the last EMBL Programme, many new approaches to functional imaging of cells were employed and/or developed, including 4D imaging, FLIM, FRET, fluorescence recovery after photobleaching (FRAP), fluorescence loss after photoactivation (FLAP) and FCS. In combination with computational analysis, these methods deliver quantitative information on molecular interactions in living cells in 4D, space and time. These studies provided novel information on the organisation and dynamics of multiprotein complexes (e.g. nuclear pore complex), cellular structures (mitotic spindle), signalling pathways (Ras signalling) and essential cellular functions (chromosome congression and segregation).

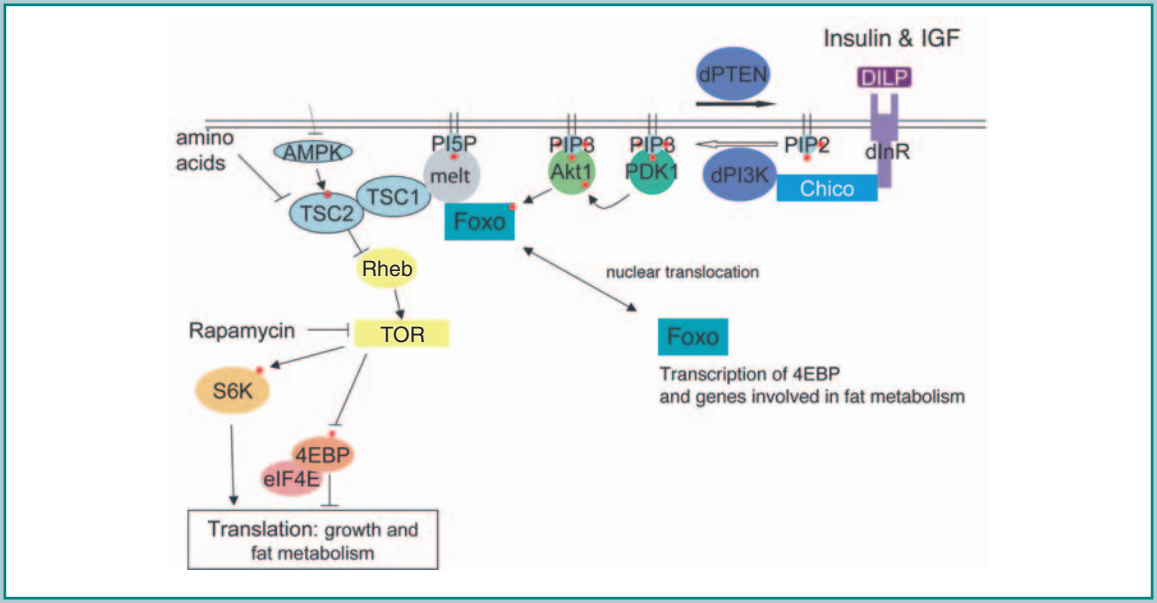
In parallel to developing ever-more powerful methods of single-cell analysis, in the next EMBL Programme we intend to move functional imaging methods into whole organisms. Here, we will take advantage of the recent EMBL development of SPIM. Using a light sheet, SPIM allows fluorescence imaging of development in living organisms up to the size of mouse embryos to a level of resolution and detail previously impossible. Importantly, SPIM can in principle be combined with the functional imaging methods already established in single cells. For example, Philippe Bastiaens and Ernst Stelzer, the developers of FLIM and SPIM, respectively, are confident that the two techniques will allow real-time imaging of intermolecular interactions in developing embryos, which would represent a very important technology breakthrough.



An image of 3D fluorescence loss after FLAP, which allows the dynamic motion of a specific protein to be monitored over time

7 How nutrient availability connects to cell growth and division

Metabolism is a complex network of interconnected modules that calls for a systems level analysis. This is as true at the level of metabolic regulation of the cell as at the level of the organism. A new interdisciplinary, inter-Unit initiative will integrate quantitative experimental biology with chemistry, bioinformatics and mathematical modelling of a key signal transduction pathway involved in control of metabolism at the cellular level. A multifaceted analysis of the insulin/TOR nutrient-sensing pathway will be a first step towards the more ambitious goal of understanding metabolism at the organismal level.



The insulin/PI3K/TOR nutrient sensing pathway

The TOR pathway has multiple inputs from growth factors and cytokines, different classes of nutrients, and multiple outputs that control transcription and translation (see figure), as well as connections that mediate cross-talk to other signalling pathways that regulate cell proliferation and cell survival. The pathway was chosen for process analysis because it is a key module of the system that lets the cell interpret its environment to decide how fast to grow and divide, and whether to store or mobilise energy reserves.

One challenge that we will tackle is to assess the flow of information through the pathway. How are the transcriptional and translational outputs balanced? *In vivo* sensors that measure signal transduction activity and output efficiency will provide new tools with which to derive quantitative information needed for modelling the pathway. Sensors will be designed for key regulatory steps, particularly at network branch points. Such sensors can then be used to identify small-molecule inhibitors that alter pathway activity, providing new tools with which to study the effects of modulating activity at different points in the network. Combined with mathematical modelling, these tools will permit an in-depth understanding of this pathway and its regulation at the cellular level. In a longer-term perspective, cell-level modelling is an essential starting point from which to extend this approach to the whole animal. The use of tractable model systems will facilitate these efforts, but the insights gained will be broadly applicable and could provide important insights into diseases of metabolism.

The ultimate goal of this research is to understand the structure of the signalling network that acts downstream of the major cell-cycle regulator Cdk1. Necessarily, as outlined at the beginning of this section, we will find that this network of enzymes affects the collective behaviour of structural molecules. Again, this will lend itself to modelling and perhaps to the discovery of new properties of spatiotemporal organisation in living matter.

1.2.3. Organismal biology

Multicellular systems are complex regulatory networks. Studies of developing organisms aim to understand how the genome is deployed to control the differentiation and organisation of single cells and, in addition, how these cells then interact to form tissues that are balanced so as to form an anatomically and physiologically coherent organism (Box 6).

The molecular underpinning of multicellularity will be studied at EMBL using genetics, functional genomics and imaging. Ongoing projects will systematically identify the genetic and regulatory networks underlying the development of specific organs and tissues. As with studies at the single-cell level, better computational analysis of genomic data is essential for progress, and modelling and simulation methods will be used to generate experimentally testable hypotheses on the functioning of regulatory modules in organism growth and development (Box 7).

Another aspect of the planned work that is strongly integrated with cellular studies is the analysis of cell migration. Groups in the Developmental Biology and Cell Biology and Biophysics Units address the cellular and molecular underpinnings of guided cell migration using a range of model systems. We aim to understand the logic underlying coordinated cell movement because this helps sculpt the 3D shape of tissues and organs. Studies of cell migration might also provide insight into tissue regeneration, during which stem cells need to move to and populate locations where regeneration is required, as well as insight into the pathological migration of metastasising tumour cells.

1.3. Genes and disease: EMBL Monterotondo

Our knowledge of how organisms function and the defects that underlie disease states is rapidly increasing, thanks to basic research in the life sciences. This brings with it the realisation that more and more of the work we do at EMBL is of direct relevance to understanding disease. We have already referred in earlier sections to work on pathogenic viruses and bacteria, and to studies of cell growth regulation and migration, both of which are relevant to cancer. These studies, like all other aspects of biology, require diverse experimental approaches if a disease state is to be well understood. Numerous computational biology projects that aim to provide insight into disease states are described in Section C.1.4. below. Genetic factors also underlie many non-pathogenic alterations in organs, systems and behaviours that will give insight into complex functional mechanisms and suggest novel therapeutic approaches.

Although numerous experimental approaches at EMBL are being applied to address gene function in disease, the most relevant avenue for extrapolation to human biology and genetic disease is to generate and analyse the phenotype of alterations in genes or sets of genes in the mouse. Nowhere is the connection between basic research and human biology so intimate as in the work done on the mouse, making it the pivotal model organism for determining the underlying genetic basis of human disease. The mouse model can also be used for testing therapeutic strategies, as well as for drug screening, because it combines a close evolutionary relationship to humans with exquisite genetic manipulability. With the completion and annotation of the human and mouse genome sequences, the field is poised to tackle one of the major challenges in genomics – the systematic determination of the function of all genes in the human genome and their respective roles

in disease. However, multigenic perturbations or 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 overexpression strategies in the mouse. Innovative chromosomal engineering and conditional approaches will be necessary to achieve spatial and temporal control of gene expression, and to augment the power of large-scale mutagenesis screens.

The international mouse genetics community has set long-term goals for achieving a comprehensive genetics-based view of mammalian physiology and disease. The EMBL Mouse Biology Unit has been active in the establishment of a systematic mouse functional genomics initiative in Europe, which currently encompasses three areas: mutagenesis, including conditional mutants in key medical traits; phenotyping, for the standardised characterisation of all mouse mutants in a medical context; and informatics, developing the necessary tools and language (e.g. ontologies) to integrate and disseminate mutagenic and phenotypic data.

Developing these areas poses challenges that are beyond the scope of an individual laboratory or centre, and yet represent key instruments for achieving the wider goal of applying genomics to health and biotechnology. Advances in the field will require the generation of custom animal models of human pathologies on specified genetic backgrounds, with clinically relevant analyses of the effect of mutations on everything from common metabolic syndromes to psychiatric disorders.

These challenges will be directly addressed by research in mouse biology at EMBL. The tremendous development of the Monterotondo Outstation over the period of the last EMBL Programme, recognised by the 2004 SAC review panel, has been enriched by links between Monterotondo and the other EMBL Units, including those (Developmental Biology, Gene Expression) where mouse work is being pursued at a more modest level. In collaboration with the large EU-funded mouse mutagenesis and phenotyping consortia that have recently put Europe in a strategic position worldwide, EMBL mouse research has developed a particular focus on generating new models of human disease. These currently include leukaemia, heart failure, diabetes, inflammatory disorders, obesity, sclerotic disease, defects in iron homeostasis and a variety of diseases related to the nervous system, including Alzheimer's disease, anxiety disorders and depression.

Human diseases can be categorised in various ways but, for EMBL's plan, the most crucial division is between monogenic and polygenic disorders. Ongoing work at EMBL in yeast, the most easily manipulated eukaryotic model organism, is providing new insight into the genetic basis of one complex trait – the ability of yeast to grow at high (human body) temperature and thus to become pathogenic. After very detailed genetic analysis, using specially developed methods that currently can only be applied in yeast, the conclusion was reached that each of the several chromosomal loci contributing to the complex trait of high-temperature growth was itself complex. In other words, more than one linked gene at each of several loci was involved in allowing yeast to grow at high temperature.

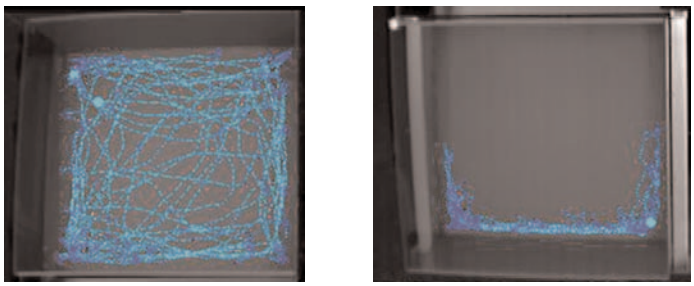
At first glance, this presents a significant challenge for analysis of human diseases, which are mostly complex traits rather than monogenic disorders, and are often only partially penetrant in the outbred human population. Moreover, multicellular organisms have developed additional genetic regulatory circuitry to generate and maintain diversity and coordination of cell function, which often cannot be extrapolated from analysis of single-cell organisms such as yeast. Therefore, new systems approaches are necessary to address complex genetic disorders both in humans and in animal models, involving the integration of analyses at many levels of regulation. Fine dissection of signal transduction of metabolic disorders in insects, or defining the molecular basis of inflammatory and proliferative defects in fish, are prime examples of how systems biology is being applied to the characterisation of human disease at EMBL. These genetically tractable systems can be effectively subjected to high-throughput screening to explore the genetic basis of the multigenic traits. Extrapolation of findings in these model organisms to humans would next involve exploration in the mouse as the mammalian organism of choice, as it allows application of systems approaches, advanced genetics, biochemical analysis, imaging technologies, and integration of datasets from

8 Identifying susceptibility genes for common diseases

Common diseases such as diabetes, cancer, heart disease, and mental illness account for a large fraction of the medical burden in developed societies. Genetic factors are important in determining who will develop these diseases, and mapping susceptibility genes is a major biomedical research goal. In these common diseases, variations in several genes act in concert to determine disease susceptibility. The mouse is the model organism of choice to study the genetics of human disease. However, the high-throughput identification of subtle disease susceptibility genes will require the analysis of large numbers of animals and will only be feasible using strategies incorporating the simultaneous screening of multiple genes.

To address this need, EMBL researchers will use chromosomal engineering technology to construct a series of 500 mouse lines carrying large, megabase-sized chromosomal deficiencies and duplications covering the mouse genome. These lines will carry either 1 or 3 copies (instead of the usual 2) of the approximately 50 genes within each chromosomal rearrangement. These mouse lines will be subjected to testing for common disease phenotypes in three areas of EMBL expertise: cognitive disorders, metabolic disease and immune dysfunction. RMCE-BAC transgenic lines tiling the chromosomal rearrangement will be used to identify the gene(s) responsible for the phenotype, and conditional targeting and additional molecular genetic techniques will be exploited to identify the cellular and molecular mechanisms underlying disease susceptibility.

The screening project will build on existing collaborations between bioinformatics experts at EBI, whole-animal imaging resources at EMBL Heidelberg, and the transgenic and phenotyping facilities at EMBL Monterotondo. To be successful, this EMBL-wide project will require augmentation of mouse housing capacity, investment in high-throughput phenotyping equipment, and development of whole-animal imaging technology. Fortunately, EMBL is already well positioned in the European mouse community and is poised to play a significant role in the current explosion in mouse functional genomics. Mouse lines developed at EMBL are deposited in the EU-funded EMMA repository where, together with lines generated by EUCOMM, the European knockout consortium, they are freely available. Phenotyping data coming from EMBL will be made available to the mouse community via the EUMORPHIA consortium. Similarly, we plan that chromosomal rearrangement lines developed at EMBL will be searchable via the Ensembl genome portal at EBI. In this way, data generated at EMBL will rapidly become available for integration with complementary efforts in the mouse and human biomedical research communities in the member states.


A B

Modelling susceptibility to depression in the mouse. When placed into a novel open arena, mice tend to avoid the centre and spend more time close to the walls, a behaviour called thigmotaxis. Studies at EMBL have revealed similarities between genetic factors that regulate this behaviour and those that determine susceptibility to depression in humans, demonstrating that genetic screening in mice can identify genes relevant to psychiatric disease in humans. Trajectories of (A) low avoidance, C57BL/6J and (B) high avoidance, BALB/cByJ mice

several experimental strategies. The interdisciplinary opportunities available at EMBL will enable several joint projects directed towards multigenic disease modelling in different organisms, based on individual and collective strengths within the institution (Box 8).

In relation to disease states caused by a single gene, animal models where the relevant gene can be mutated in the adult in a temporally regulated and cell-type-specific way that accurately reflects late-onset human disease will be another major focus of our efforts. Most of the mutations generated to date are loss-of-function (null) germline mutations. Whereas null mutations allow the earliest function of a gene to be determined, if a gene has a vital developmental role, the identification of functions later in development are often obscured. To address these limitations, EMBL researchers are collaborating with other European consortia to develop novel mouse models that provide inducible, tissue-restricted control over gene activity, using combinations of bacterial recombineering and binary recombinase-based approaches. Generating and characterising these models, and relating them to the relevant human pathologies, is labour-intensive and thus expensive both financially and in terms of time commitment.

The repair of tissue damage following injury or illness is another focus of disease-relevant research at EMBL. Regenerative biology is currently focused on the untapped potential for tissue maintenance and repair in vertebrates and the basic impediments to more-efficient regeneration in mammals. A host of cellular responses, such as cell-cycle checkpoint control, mutagenesis, and apoptosis in response to DNA damage, must be considered if we are to identify more-specific targets for the development of stem cell protectants that prolong stem cell regeneration and increase stem cell pools. The multidisciplinary environment of EMBL is ideal for addressing the increasingly complex and clinically strategic issue of regenerative biology. In our planned future work in this area, we will focus on the molecular and cellular mechanisms underlying the normal tissue repair process in mouse muscle, some of which might be successfully harnessed or altered to achieve adequate cell replacement and decrease inflammation after tissue damage in adults.

An increased appreciation for the participation of adult stem cells in mammalian tissue renewal now requires a more thorough understanding of their biology and their potential use as therapeutic agents in tissue maintenance and repair. A focus of the Mouse Biology Unit will be to define the characteristics of adult mammalian stem cells that can be employed to increase the regenerative potential of critical organs such as the heart or nervous system. The overall goal of our work in these areas is to provide the medical community with material, genetic tools, standards and knowledge to support clinical interventions, thereby enhancing the understanding of human health and treatment of disease conditions in a systematic manner.

1.4. Bioinformatics and computational biology research

The revolution in functional genomics over the past decade has led to an ever-expanding information-rich biology, in which data integration and interpretation will play an essential role to further our understanding of biological functionality. Computational biology (used here to include all aspects of bioinformatics and modelling) promises new discoveries, driven by analysis of data and context, by inference through evolutionary relationships, by integration of data from different experiments across multiple organisms and environments, and through improved modelling of biological processes. It is both an enabling technological discipline and a research direction. We therefore see computational biology as being at the heart of what EMBL wishes to achieve in the next five years.

Today, computational approaches drive discovery. Optimal solutions to biological questions are often only achieved through an iterative process, starting with computational simulations of an experiment to optimise experimental design, followed by data collection in the wet laboratory, interpretation through computational methods, including integration with relevant public domain data, and then on to designing the next experiment. Bioinformatics is involved from start to finish, from collection and processing of the data, providing in-depth annotation, further interpretation and integration, and finally making the data available

in the public domain to be fully exploited in a broader context. An important goal, requiring modelling and simulation techniques, is to transform biology from a descriptive into a quantitative predictive science. To promote this, we expect computational biology to become pervasive in all biological laboratories, analogous to the way that molecular biology techniques, such as gene cloning and overexpression, have become ubiquitous.

Modern biology today presents many "grand challenges" that will only be met through the application of computational approaches. The complexity of living systems is immense, involving huge numbers of different molecules. These molecules might perform multiple functions, tuned according to their cellular location or to the specific organisms in which they are expressed. Elucidating molecular function *in vivo* remains a major challenge for the foreseeable future, requiring extrapolation, integration and interpretation of heterogeneous data sets. However, complexity occurs not only through the number of different molecules involved, but to an even greater extent through the intricacy of their regulation and the number of interactions they form in complexes and pathways. Elucidating biological pathways and networks, especially in humans, will require extensive data collection, assimilation and simulation. These networks are dynamic, changing according to their spatial and temporal environment. Such complexity can only be handled computationally, as it involves large amounts of data coming from different research fields and derived at different scales, from the molecule to the organism and its environment. Appropriate models will be sought to describe biological processes at the appropriate level (molecule, cell, organ, organism or community). In addition, as we seek to move towards an understanding of biological systems, we will need a seamless integration of the entire information space. Currently, we are struggling to "integrate" all the molecular data, derived using different technologies, different experimental conditions and different model organisms. Increasingly, it will be necessary not only to integrate these data, but also to access information in other disciplines, such as chemical, medical and ecological data.

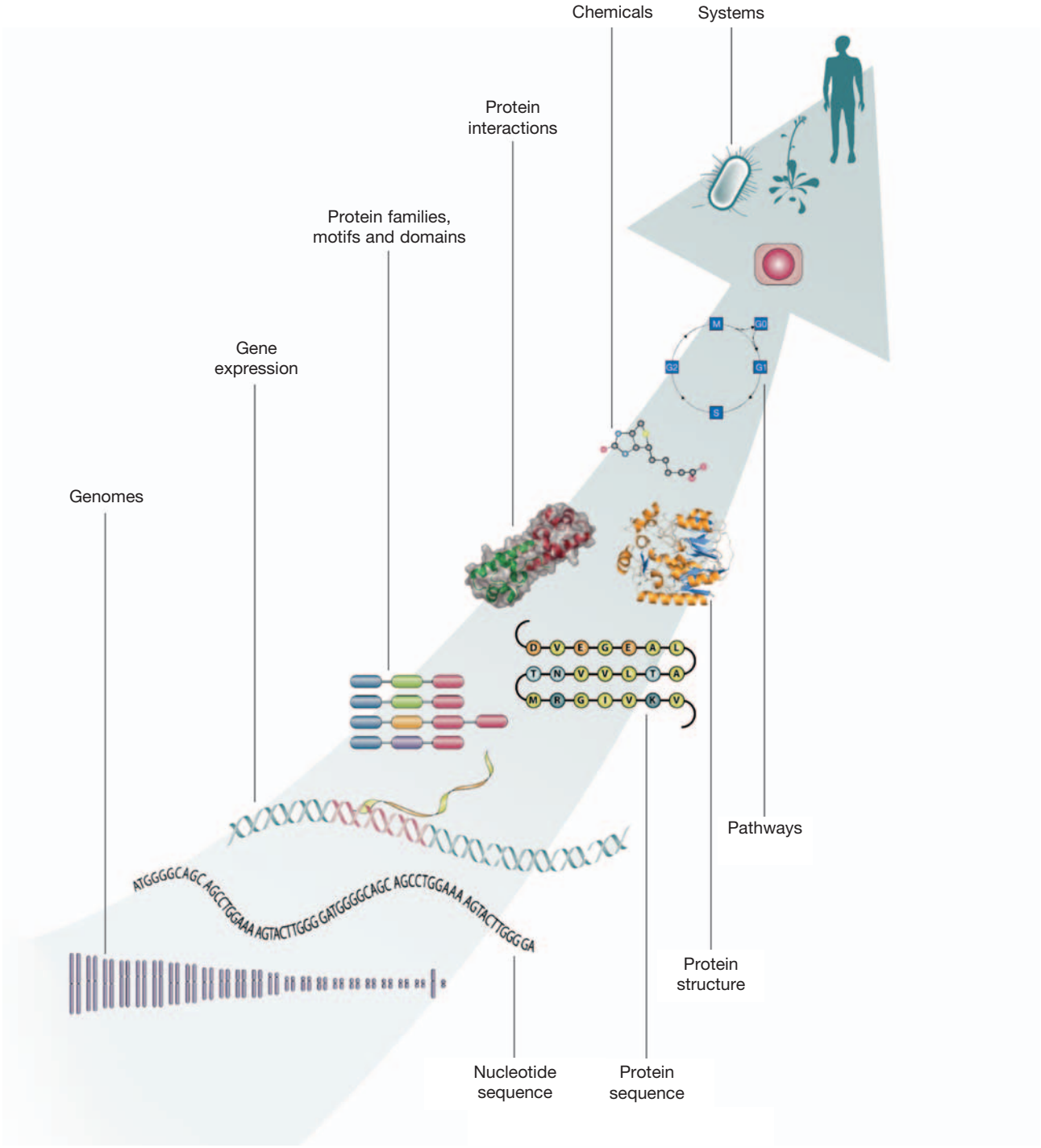
To begin to address these challenges, computational biology has changed considerably over the past five years, becoming more diverse with many sub-specialisations, according to the experimental technique used (e.g. transcriptomics/structural biology), the organism of interest (e.g. bacteria/mouse) and the "mathematical" approach employed (e.g. statistical data analysis/differential equation models). At EMBL, we have already developed strengths in many of these areas. The far-sighted decision (taken by Lennart Philipson when he was Director General) to pursue bioinformatics research on the one hand, and the construction and provision of biological databases on the other, means that EMBL now has a considerable proportion of its groups at EBI, Heidelberg and Hamburg whose focus is computational biology. We can recognise several distinct flavours of computational biology research that will all be essential as we move towards systems approaches. There is great strength in the computational biology research groups that pursue pure *in silico* research to answer biological questions, frequently by interpreting experimental data generated in-house in the context of existing knowledge. Second, with the strong technology needs of the data resources at EBI, there is extensive research and development into finding new ways to collect, annotate, standardise, present and integrate biological data. The challenges presented by biological data stimulate research into novel search methods and comparison algorithms, as well as into improved database design and integration. Similarly, the development of software for use in structural biology has become a speciality at the Hamburg Outstation. What will be necessary in our future research is not only to have these specialist bioinformatics or computational groups, but also to use their expertise to implant computational approaches into all of EMBL's Units and most of its individual research groups. Finally, groups in Heidelberg and EBI are pursuing modelling and simulation in the context of collaborative projects that involve several EMBL Units and an interdisciplinary combination of expertise.

The groups and teams engaged in this work are listed in Annex 8. The close interaction of all of these groups, which will be stimulated further by the Computational Biology Centre (see below), provides a powerful integration of data, methods and research across EMBL.

1.4.1. Future research directions

The research in computational biology at EMBL covers data at different scales, from single point mutations (e.g. single nucleotide polymorphisms) to ecosystems (e.g. metagenomes). The major direction over the next five years, in keeping with the theme of systems biology, will be the integration of this data to gain a better understanding of how a cell, organ, organism or whole ecosystem works (see Figure C.1.). This will involve a move towards quantitative biology, aiming for improved predictive power, using the appropriate model according to the scale on which the research is focused.

Figure C.1. Bioinformatics and computational biology research at EMBL



To achieve this, we will maintain our core research strengths in exploiting: genomic and protein sequence and structure relationships to infer function; comparative analysis of high-throughput data, including functional genomics and proteomics; integration and standardisation of heterogeneous data; bridging from molecular to subcellular (network) scales; and evolutionary studies at all levels, including the evolution of molecules, networks and phenotypes. In addition, we envisage a greater emphasis on: research into developing quantitative approaches to model specific biological processes, including temporal and spatial aspects; the integration of metabolite data to facilitate a better understanding of how the cell functions and the move towards chemical biology; an increasing focus on how to handle and process biological image data and their projection to molecular processes; the integration of the literature with all other types of data; and an increasing involvement in understanding the molecular basis of disease, especially how human variation influences health and disease. This plan is developed below by reference to ongoing or emerging projects that serve as examples of our future plans.

1.4.2. Inferring molecular functions: comparative genomics and evolution

Following the determination of numerous genome sequences, a major challenge for the whole biological community is to determine the function of the gene products. This is a huge task, involving both experimental and computational approaches. Several research groups in EMBL have developed new methods and data resources to address this problem. To advance current prediction schemes, a better understanding of evolutionary processes is required and computational approaches are expected to contribute considerably. At EMBL, there will be continued research into the evolution of genomes and proteins, including their structures, functions and interactions. The fundamental processes involved in sequence evolution will be modelled with increasing sophistication, helping to identify genes that have been under selection to remove deletions or mutations, or purifying selections, during recent evolution. Protein structure allows a deeper insight into the processes of molecular evolution, revealing how individual proteins duplicate and subsequently evolve, through radically altering their sequences or by changing their domain composition. This permits the evolution of novel functions or the formation of multiple different complexes, and so allows complexity to develop at the systems level. In higher organisms, many gene families include several members that are involved in multiple signalling pathways that control diverse biological processes. Comparative genome analysis of these gene families and the functional units they encode (i.e. the protein domains) allows inferences to be made about functional differences between organisms based on expansion and loss of branches of these protein families. In the next period, we will pursue the development of methods that allow better predictions to be made about the interactions and specificity of different family members. If successful, this will enable us to provide a significantly enhanced level of information on protein function.

1.4.3. Inferring cellular functions: protein interaction and regulatory networks

When moving towards systems biology, the understanding of cellular networks is essential. Groundwork is still needed to describe the topology, evolution and dynamics of such networks. In the near future, two major types of networks will be studied: protein interaction networks and regulatory networks. At EMBL Heidelberg, tools are being developed that use information derived from the sequence, genome location, phylogenetic profile, pathway data, interaction data and text mining to suggest protein interactions, and research on dynamic and evolutionary aspects of such networks is being pursued. At EMBL-EBI, methods are being developed that predict interaction information from structure. To understand regulatory networks that control gene expression, it is going to be necessary to have better methods of predicting regulatory sites in DNA or RNA sequences. Comparative methods across genomes will also be applied here because the relative simplicity of these regulatory elements makes them impossible to detect accurately in analyses of a single species. Experimental

analysis coupled with computational modelling should also enable insight to be gained into the evolution of complex subcellular structures. To provide an example, there are quite distinct types of mitotic spindle structure across evolution, and modelling approaches coupled to experimentation will be used to understand the factors that influence microtubules to produce such apparently diverse structures. Comparative analysis across evolutionary time of the genetic and regulatory networks that give rise to homologous structures in diverse organisms should allow the identification of "genetic modules" or networks that have evolved to regulate particular events at either the subcellular or organism level.

1.4.4. Towards temporal and spatial aspects: chemical biology and systems

When moving towards a systems understanding, not only are quantitative, standardised data sets needed, but also information on the temporal and spatial context in which biological processes function. This is a field to which chemical biology is ideally suited, as an understanding of the interaction of small molecules and proteins might enable controlled and time-sensitive perturbation of networks and systems. A general interest in chemistry of all the bioinformatics groups at EMBL, together with the recruitment of a new group leader in the area, is triggering exciting new developments. A systematic mapping of metabolites, together with in-house efforts to catalogue protein–compound interaction data exhaustively, should provide the basis for modelling cellular systems. Tools that consider the compartment structure of a cell and diffusion between compartments are currently being developed.

In parallel, recent work carried out at EMBL has provided examples of how it is possible to use structural information and homology modelling to infer the structure of new protein complexes, potentially even on a large scale. This process is further fuelled by worldwide structural genomics consortia, which are solving the structures of many novel proteins and protein complexes. At EMBL, novel methods are being developed to predict the function of a protein from its structure, focusing on the identification of small-molecule ligands. This is particularly relevant in families of disease-related proteins, such as the kinases or short-chain reductases, which are known drug targets and are large families with diverse substrates. Integrating these data with current knowledge on the presence and quantity of small molecules in the cell (the metabolome), together with knowledge on drug molecules, will provide testable candidate ligands to probe biological function. We are thus aiming at an integration of structural biology, chemistry and systems biology, with much input from new computational approaches.

1.4.5. Biology at different scales

As mentioned in previous sections, biological function and biological systems come at different scales. The protein itself can be thought of as a complex, dynamic, biological system composed of atoms or amino acids. Similarly, a pathway, a cell, a multicellular organism or an entire ecosystem can be the object of system-level study. Until recently, quantitative data were only available for molecules (genes, proteins); such data now exist at the genome, transcriptome and metabolome level and, with advances in imaging techniques, data on whole multicellular organisms will become available at high resolution. Furthermore, the first data on the genetic material of large ecosystems have become available (metagenomes), which allows an analysis of gene pools across species boundaries and provides, for example, information on the environmental constraints on such communities (e.g. in water and soil) and the complex interchange of nutrients between species. Computational biology at EMBL has provided leading contributions at all these scales and will continue to develop the tools required to enable an increased understanding of biological systems and their defects in disease (see Box 9).

9 Disease-related computational research

Computational research at EMBL will be increasingly relevant for understanding the molecular basis of health and disease. This is illustrated by reference to the three specific projects described below:

- **Systematic identification of coding microsatellites mutated in cancer cells.** DNA-repair-deficient cancer cells tend to incorporate frameshifts into their protein chains when the coding regions of exons are composed in part or in total of simple sequence repeats. The presence of such repeats and the frameshifted proteins that they encode can be predicted from genome information. In collaboration with a clinical group in Heidelberg, the predicted mutated versions of some such proteins could be shown to be reliable targets for specific T cells that were stimulated by the peptides corresponding to the repeat regions. Clinical trials using the first generation of peptide predictions have recently been initiated.
- **Systems biology modelling of drug target proteins in the brain.** The protein phosphatase inhibitor DARPP-32 is a major target for dopamine and glutamate signalling in the striatum. Quantitative models have been built of the signalling pathways known to mediate the effects of neurotransmitters, neuromodulators and drugs of abuse on DARPP-32 phosphorylation. The role of DARPP-32 as a signal integrator and coincidence detector was studied by dynamic simulations. Sensitivity analysis revealed that cocaine acted on the less-robust parameters of the signalling pathways involved. The model is now being extended to incorporate the effects of growth factors and gene regulation, to gain a better understanding of the mechanisms of neuroadaptation.
- **Molecular basis of ageing.** As part of a large “Functional Genomics of Ageing” consortium, centred at University College London, we are comparing expression data for ageing flies, worms and mice, and the effects of longevity-associated mutations. New methods are currently being developed to compare these mutant organisms and calorie-restricted individuals in different species. Preliminary analysis thus far suggests that detoxification processes (the “Green” theory of ageing) are involved in longevity assurance.

1.4.6. Technical research and development, including data integration

In the past few years, EMBL's bioinformatics researchers have been active in the development and dissemination of software tools that, used in conjunction with the EBI databases, allow the large and diverse community of life scientists to extract or process the information they need to support their research. Examples include tools for comparing structures, for predicting molecular interactions or for functional annotation. Methods development will continue in the next EMBL Programme and, in particular, research groups at the EBI will pursue this in close collaboration with their colleagues who work directly on the core molecular databases.

For example, tools to analyse functional genomic data (e.g. microarray or proteomic data) are still developing rapidly. This is evident from the fact that analysis of the same data by different software packages frequently produces very different results and conclusions. Groups at the EBI will work in this area, using complementary approaches and pooling biological and statistical expertise. Similarly, the development of software packages and databases for simulation methods, to be used in systems biology, will be actively pursued at EMBL. There is also a huge need to develop better methods to extract information automatically from the literature, and study of this problem is underway at the EBI. It will be evident to readers of the preceding sections that imaging is going to be a major activity at EMBL and in other life science research institutes over the next five years. We will need to be involved in the generation of tools that enable quantitative image analysis and storage in such a way that image data can be stored in databases and allow experiments carried out in different laboratories to be compared in a quantitative way.

One challenge common to all approaches in systems biology is the need for integration of many different types of experimental data, often held in different databases. For example, to understand the molecular basis of a disease will probably involve combining information on individual molecules, their interactions (protein–protein, protein–nucleic acid or protein–small molecule) and their involvement in metabolic or signalling pathways. Developing improved approaches for accessing and interpreting such data is a major research goal of EMBL. In the future, linking such information to human variation data and even patient records will be an even greater challenge. One current example that illustrates the approach needed is the BioMart project at the EBI. While BioMart has developed a novel generic solution to integrate the major molecular data resources within the EBI, solutions that involve external data resources will also be required.

There are numerous pilot projects underway in which EMBL groups provide the expertise to annotate and interpret large and diverse datasets from research consortia worldwide. These include the vast majority of the metazoan genome projects, where teams at EBI and Heidelberg took responsibility for gene prediction and comparative analysis, as well as transcriptomics and proteomics projects where EBI groups help to formulate standards and to rationalise and format the data for worldwide usage. Genetic interactions that underlie complex phenotypes such as buffering and epistasis are being studied by EMBL groups within a large consortium in order to infer genetic networks ("pathways"), using genome-wide RNAi to perturb development and signalling in *Drosophila* cells and embryos. Various reporter assays, from simple viability to complex image time series, are used as a readout (phenotype). This requires analysis and modelling, with a combination of statistics and graph theory, as well as image analysis to derive and digitalise the description of phenotypes in cell lines and fly embryos. The tools developed by EMBL's computational biologists are already in extensive use throughout the Laboratory (Box 10).

10 Integrating computational and experimental activities within EMBL

Often, problems become too complex for individual groups and there are a variety of cross-Unit projects where bioinformatics is crucial. Below we illustrate some projects that have naturally evolved in the EMBL environment, usually with sharing of personnel between groups or Units to combine wet-lab and dry-lab expertise.

- The iteration of computational identification and target prediction for microRNAs, and their experimental verification, has enabled EMBL researchers to be at the forefront of this new and emerging field.
- The annotation of linear motifs such as phosphate-binding sites in protein sequences and its combination with interaction data has led to the prediction of many novel peptide-binding sites in globular domains, several of which could be experimentally verified.
- To investigate *cis*-regulatory elements in vertebrates, computational screening techniques developed using comparative genomics methods at EBI were combined with both iterative rounds of experimental verification using transgenics and *in situ* studies in Medaka fish at Heidelberg. To date, the result of this continuing project has been the definition of over 50 specific targets that help explain retinal development in the vertebrate eye.
- High-resolution (tiling) maps of coding and non-coding transcripts in yeast are being generated and analysed. A static analysis in one experimental condition has been completed. This single study found extensive evidence for the role of non-coding RNA (3' and 5' UTRs, antisense RNAs, general untranslated RNAs) in regulation, and efforts will now turn to the dynamic picture of expression of these RNAs during the yeast cell cycle.
- Combinations of genome-wide chromatin immunoprecipitation (ChIP)-on-chip and array time series have been generated and analysed with the help of computational biology groups. New methods have been developed to interpret and integrate such data in order to study the targets of a number of transcription factors and to decipher their complex and interdependent regulatory activities during muscle development.

In addition to these tight collaborations between experimental and computational groups, many collaborations exist between computational groups within and across Units; examples include genome and proteome annotation projects, as well as coordinated work on alternative splicing.

1.4.7. Facilitating and promoting bioinformatics research across Europe

Almost all computational biology groups in EMBL are involved with some aspect of genome annotation, software tool development or provision of resources to annotate data based on information derived from the literature or from integration of heterogeneous data. EBI leads three Networks of Excellence (BioSapiens, EMBRACE, ENFIN), bringing together over 50 laboratories from across Europe to provide the infrastructure and new methods to allow biologists to easily access information generated in many separate laboratories. The computational groups across EMBL are involved in more than 40 distinct EU projects on various research topics in which they share their experience in data integration and analysis (see Annex 10).

In summary, the centrality of computational biology to EMBL's future plans is unquestioned. The difficult decisions will be to decide which areas we should tackle and which we do not have the resources or personnel to pursue. The list above of the important fields for further research is far from exhaustive but represents the areas that we see as our highest priorities.

2. Building internal integrative activities

Two of EMBL's major strengths are its Unit structure and its collaborative culture. The Units are designed to be mutually complementary and interdigitated. In order to support this organisational model, the Units must have sufficient critical mass to be able to look beyond the borders of their core competences when recruiting. The Units consist of young research groups and teams of limited size that are subject to regular turnover. This combination of properties, together with a carefully nurtured EMBL "research culture" leads naturally to groups collaborating both within and between Units to tackle projects that require a broader range of technical and intellectual expertise than can be collected in one individual research group.

The beginnings of genomic-scale studies during the 1990s and the rapid developments of high-throughput technologies to support them posed a problem for EMBL. Genomics technologies are typically expensive, to a level that exceeds the possibilities of our individual research groups, and they also require dedicated personnel with significant expertise to run and maintain them. The Laboratory met this challenge during the 2001–2005 EMBL Programme by investing a considerable amount in setting up Core Facilities, well-equipped teams with a service and training role. These facilities are available to scientists throughout EMBL and our plans for their future development will be described in more detail later.

During the preparation of the EMBL Strategic Forward Look 2006–2015 (http://www.embl.org/aboutus/news/publications/forwardlook.html), which was an exercise that involved extensive discussion both throughout the Laboratory and with representatives of the SAC, EMBL Council, EMBO and EMBC, the Laboratory expressed the opinion that the above-described measures were not sufficient to support EMBL's plans for the future optimally. The integration across disciplines that will characterise and be necessary for success in the area of systems biology requires additional strategies. The most important of these are the EMBL Centres.

2.1. EMBL Centres

The first EMBL Centres were established with minimal funding in 2003 in a pilot phase. Their purpose is to promote interdisciplinary research at EMBL. They should promote ambitious goals that go beyond "normal" EMBL collaborative projects in areas that are of interest to most, if not all, EMBL Units, such as computational biology or biological imaging. One example of possible activity is in projects that involve a significant amount of technology and instrumentation development (e.g. software, microscopes, high-throughput technology). Another is to promote collaboration in areas that will become increasingly important for EMBL, but where we are currently comparatively weak, such as working together with

clinicians. The EMBL Centres are transient structures formed as ad hoc interest groups to network and promote the exchange of ideas through seminars, conferences and training activities. We wish now to encourage the Centres to engage in the development of new interdisciplinary projects.

Four Centres have been established so far: 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 2–4 senior scientists or group leaders (see Annex 11). Participation in Centre activities is open to all EMBL scientists. The number of participants in each Centre is variable and depends on the scientists' interests.

In the initial phase, Centre activities were proposed and approved by the Heads of Units and the Director General. During this pilot phase, the Centres were funded through donations from the existing Units, supplemented by matching central funds. A variety of activities have been provided with minimal support by this mechanism, including a centre website to be used by all Centres, seminar series and symposia, software development, training in computational biology and imaging, and a project that pursues the development of a 4D gene expression database (see Box 12 below).

The pilot phase has had the purpose of promoting the flow of information and expertise between the Units of the Laboratory among people who have similar interests and needs. To succeed in their more ambitious goals, however, the Centres will require additional funding. There are personnel requirements in terms of computer engineers to develop software, databases and websites, and for efficient training. Also, it is desirable to have the possibility to provide seed funding, for example in the form of pre- or postdoctoral fellowships, to promote the interdisciplinary projects that will be carried out in the context of the Centres. Much of the funding for these activities will come from existing internal and external sources, but supplementary funding is requested in order to encourage and kick-start interdisciplinary activities.

2.1.1. Centre for Molecular and Cellular Imaging

Imaging analysis is a pervasive activity in modern life science laboratories. At the smallest scale, it begins with X-ray crystallography and NMR, moves through cryo-EM and SAXS, to EM tomography at various levels of resolution, and then on to light microscopy of cells and organisms. More and more biological experimentation relies on the acquisition and analysis of digital images. The storage, annotation and distribution of high-resolution structural biology image data are functions of the Molecular Structures Database (MSD) at the EBI. In part as a result of their activity, structural biologists are generally aware of the requirement for standardisation of methods for data acquisition, storage and sharing. This is enabling the increasing automation of image analysis (e.g. of X-ray diffraction data), an activity in which EMBL scientists have been, and are, playing a leading role.

Quantitation and comparative analysis of light microscopy images is at a much earlier stage in its development. One aim of EMBL's Imaging Centre is to promote cross-talk between people developing tools for, or carrying out image analysis at, different scales, to enable transfer of expertise and methods between what have previously been separate disciplines. A long-term ambition is to work towards integrated, quantitative image comparison across scales, so that fitting X-ray structure into EM tomograms or integration of light microscope and EM images becomes more readily feasible within individual experiments and between different laboratories.

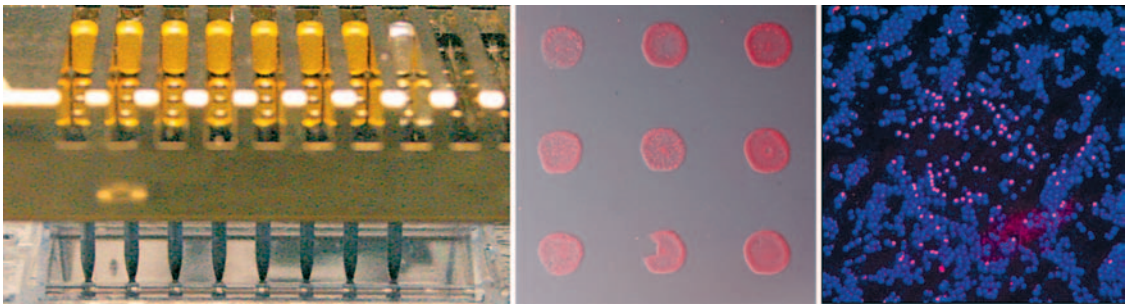
Currently, members of the Centre are organising courses on methods for live cell imaging and a workshop on EM- and synchrotron-based tomographic methods. Provided funding becomes available, they have plans to carry out several projects. One of these is described in Box 11.

11 High-content microscopy screening

The Centre for Molecular and Cellular Imaging aims to make microscopy-based screening of cellular assays available to EMBL research groups.

The phenotype of cells provides a high-content readout of functional genomics experiments such as RNAi knockdown or gene overexpression. In cultured cells, immunofluorescence or live cell analysis of appropriate fluorescent reporters provides unique phenotypic detail. Cell arrays for transfection combined with automated microscopy, both of which have been successfully tested at EMBL, will make such high-content cellular assays amenable to high throughput, allowing genome-wide screens. This technology and a large part of the infrastructure was established during the last EMBL Programme with the help of external funds, and microscopy-based, genome-wide RNAi screens to identify novel human genes involved in mitosis and protein secretion are underway. However, microscopy assays can be designed for almost any biological process and thus many EMBL groups wish to benefit from being able to perform such gene discovery experiments, either at the genome-wide level or focused on already identified gene networks. This will be a key enabling technology for systems biology.

Several resources will be required to make high-content microscopy screening available within EMBL, including genome-wide RNAi libraries of high and validated quality for several species, large-scale data storage and management infrastructure, and image processing and bioinformatics tools for the analysis of large numbers of digital images.



Production and analysis of transfected cell arrays

2.1.2. Centre for Computational Biology

Biological research is becoming more and more dependent on computational expertise. There are two main reasons for this. The first is the dramatically increasing amount of heterogeneous data available on biological molecules and processes. Efficient storage and handling of the data is essential to be able to progress to the most important step – data interpretation. This, in turn, requires the data to be compared with existing knowledge, which is fragmented and stored in various forms at many different locations.

The second reason is that biology rapidly becomes more quantitative. As we move towards the understanding of complex systems, it will no longer be sufficient to provide cartoon depictions of biological pathways and networks. We will need quantitative descriptions of the connectivities between the components of the networks and consideration of spatial and temporal constraints. Quantitative measurements and sophisticated statistical analysis will be crucial in modelling the functional modules that make up biological systems.

It is for these reasons that we are convinced that integrating computational methods and expertise into all of EMBL's research is a necessary ambition for the next EMBL Programme. The Centre for Computational Biology will play an important role in the realisation of this ambition, focusing on three major goals: (1) the development of platforms for the intellectual support of experimental groups as well as for electronic information dissemination and exchange; (2) the organisation of various training activities such as in-house courses, interest group meetings and exchange of personnel; and (3) the support of a few ambitious interdisciplinary projects in which computational biology should play an essential and enabling role (see Box 12 for an example). Several other projects are ready to be started within the Centre, including: "Pipelines in molecular and structural biology", "From recent grid developments towards self-learning approaches", and "Modelling reaction–diffusion at different scales and in different geometries".

Much of the collaborative work to be carried out at EMBL over the next quinquennium will involve participation of groups with computational expertise. Although EMBL has a variety of successful and experienced computational biology groups, these are currently scattered, and the Centre as a central port of call will help everyone to find the appropriate collaborator with the required experience and knowledge. As many experimental groups either already have, or plan to hire in the near future, personnel that mainly work with computers, these people need to be brought together to gather experience and share problems. In particular, new recruits with a background in mathematics, physics or computer science need to be integrated with our biologically trained staff to avoid their isolation. Common meetings and provision of training will thus be one major activity for the Centre, and the support of interdisciplinary projects with different computational aspects will be another.

2.1.3. Centre for High-Throughput Functional Genomics

As with the imaging and computational biology Centres, the High-Throughput Functional Genomics (HTFG) Centre is based around state-of-the-art technologies that need to be made available to researchers in all EMBL Units. High-throughput experimentation is being applied to many problems at EMBL. To list but a selection, these include: the characterisation of a complex genetic trait (the ability of yeast to grow at high temperature) through novel genetic approaches; the analysis of the regulatory circuits that lead to the differentiation of muscle types in *Drosophila* (Box 13); the elucidation of the components and functions of the nonsense-mediated decay pathway; the characterisation of the mitochondrial proteome as a precursor to a systems analysis of mitochondria (Box 14); the screening for small-chemical inhibitors of specific steps of mitosis; and the systematic search for soluble derivatives of proteins for structural analysis.

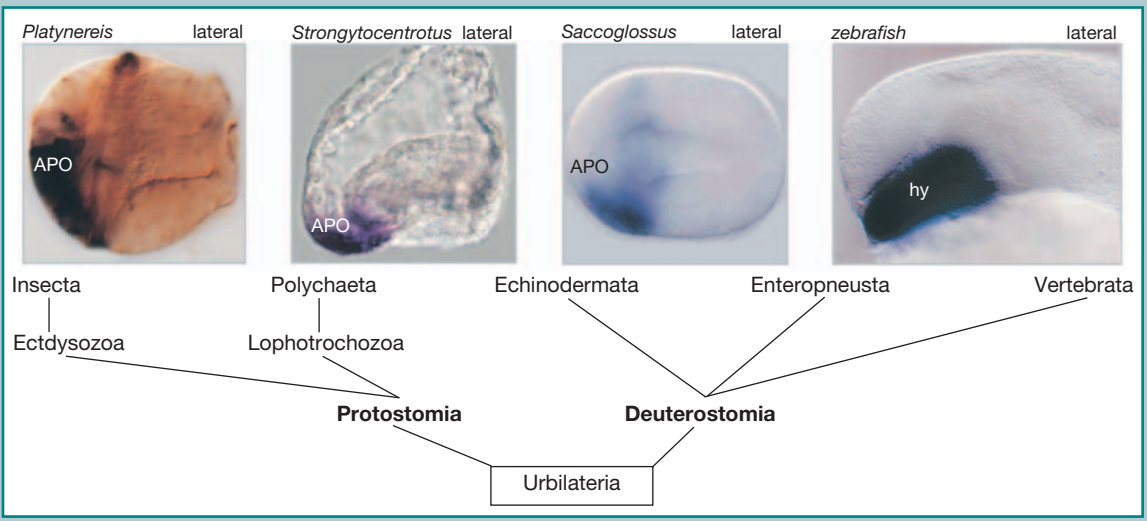
12 Comparative gene and protein expression in 4D

We aim at a comparative analysis of gene expression between species with both temporal and spatial resolution. This will allow the tracing of cell types throughout animal evolution and the identification of conserved gene regulatory networks that function at key positions in development and disease. At the moment, only fragmentary data is available in different formats for different species in different databases. The aim of the project is to combine gene and protein expression data from various sources such as expressed sequence tags (ESTs), microarrays and whole-mount *in situ* approaches to generate a resource that provides qualitative and quantitative information on gene expression in the context of the developing organism within the framework of bilaterian evolution (4D). This will contribute to our understanding of what defines a distinct cell type within a given species in terms of gene expression.

We identify homologous cell types in different animals based on their molecular fingerprint. As an example, one cell type we focus on in different species are the photoreceptors, as there is already a considerable amount of existing data. We will use our expression in 4D database to compare gene regulatory networks that specify and control photoreceptor cell differentiation, morphology and physiology. The eye is a composite structure composed of many different cell types, making it a good model for the integration of diverse cell types into a complex, functional system.

We also intend to carry out an analysis of the temporal and spatial expression of genes in an entire germ layer, the mesoderm, in a cross-species approach. This will reveal important conserved gene expression networks that govern mesoderm development and also species-specific differences.

The final aim is to apply this information, mostly derived form model organisms, to human organ development and disease. There are several computational challenges in this work, ranging from the development of dedicated databases, to tool development for comparative analysis, and the requirement for novel techniques for the interpretation of complex and diverse results. The nature of the project imposes a requirement for tight communication between experimental and computational groups scattered across EMBL. There is also a strategic aspect to this research that is relevant to the European scale and mission of EMBL. The comparative analysis imposes the necessity for standards, and database and data production pipeline design, at a level beyond an individual species (and research group).

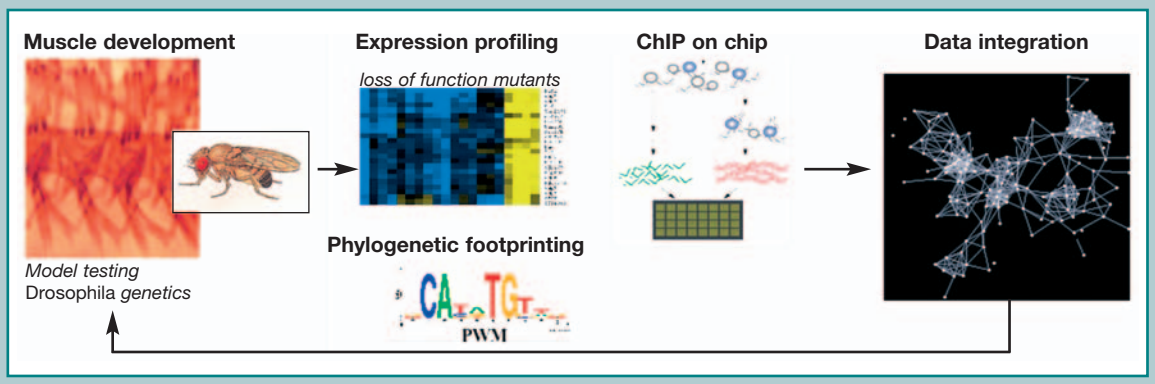


Spatially restricted expression of nk2.1 orthologous genes in different species by whole mount in situ *hybridisation*

13 Understanding transcription networks, noise and buffering

With the development of functional genomic approaches such as ChIP-on-chip analysis and gene expression profiling in different genetic backgrounds, it is now possible to obtain a global view of transcriptional regulation. We are interested in understanding the topology and dynamics of transcriptional networks within the context of embryonic development. *Drosophila* muscle development was chosen as a model system because it has several clear advantages. First, the transcription factors regulating muscle development are known, are highly conserved from flies to humans and are well characterised. Second, as there is only one member of each of these transcription factors present in *Drosophila*, our network will not have to deal with the complexity of functional redundancy between family members. Third, the robustness and buffering of this network can be tested in this genetically tractable system.

Efforts are currently underway to identify the direct target genes of six interconnected transcription factors involved in muscle formation, using ChIP-on-chip analysis. Expression profiling data on embryos deficient in each transcription factor is being used to assess if the direct target genes are activated or repressed. The selected transcription factors are known to regulate the expression of each other, as well as that of some common target genes. The overlapping nature of the datasets serves both to validate the network connections and to provide new insights into the modular components of the network and their regulatory code.



Dissecting the logics of transcriptional networks during development: loss of function mutations in Drosophila *for all essential regulators of myogenesis are being used as an entry point to identify the transcriptional program required for muscle development.*

As network topology alone is not sufficient to determine network behaviour, the dynamics of the *cis*-regulatory responses will also be assessed. ChIP-on-chip experiments and expression profiling is being performed at consecutive developmental time points. Although this will not reveal the real-time kinetics of transcription factor binding, it is a first step to assess the dynamics of network behaviour within the context of a developing tissue. By continuing to apply these multiple experimental strategies, we will be able to construct a detailed network of the circuitry that directs different stages of muscle development and will begin to dissect the logic of individual regulatory components. Of particular interest are components that serve to decrease “transcriptional noise”, which would otherwise lead to developmental defects. Key predictions will be tested by generating perturbations in network components using genetics and/or RNAi.

Myogenesis, like many other developmental events, is also regulated by epigenetic mechanisms. Experiments are currently underway to examine the control of transcriptional silencing by Polycomb group genes during development. These data, along with published genome-wide studies on acetylation and methylation, will be integrated into the myogenesis transcriptional network. This will provide information on the chromatin context of both the transcriptional regulators themselves and their direct target genes.

As all high-throughput genomics projects involve large numbers of repetitive experimental steps, these require automation. This technology is unfamiliar ground for most biologists. It requires not only expertise in robotics but also in management of large quantities of experimental samples and large-scale data analysis. Normal bench-top experiments must often be streamlined and reduced in scale to make it feasible to carry them out at genomic scale. On the basis of the cost of materials alone, this miniaturisation step is frequently essential to allow genome-scale experimentation. All of these essential steps in moving from gene-scale to genome-scale experimentation require specific expertise of the sort that is unavailable in most individual molecular biology laboratories.

The purpose of the HTFG Centre is to catalyse the use of high-throughput methods further, in part by aiding in the organisation of service-based Core Facilities. It will also serve as a forum for the collection and distribution of intellectual and practical expertise between the groups who wish to become active in high-throughput experimentation in the next EMBL Programme. The centre is organising an internal functional genomics discussion group that will meet twice a month, to promote interactions and information exchange within EMBL. The centre intends to organise an international "Functional Genomics to Systems Biology" meeting, to be held at EMBL in October 2006; if funded, this will be the third such symposium that has been organised in Heidelberg.

2.1.4. Centre for Disease Mechanisms

A major challenge for life sciences in the next decade will be to generate the basic knowledge required to design new diagnostics and therapeutics for a large number of human diseases. The understanding of the mechanisms leading to disease pathogenesis (molecular pathophysiology) will be fundamental in this effort by providing new targets for therapeutic intervention. This task will require an interdisciplinary approach, bringing together different expertise across the borders of basic research and clinical science. As a result, the borders between basic and applied medical research will coalesce. EMBL, with its unique structure and position in European research, is already playing a role as a collaborating partner in this effort. The rapidly increasing level of information on the functioning and malfunctioning of biological systems means that an ever-larger proportion of the research carried out at EMBL has a direct connection to disease states. The logical progression of much of this research will be to uncover its potential relevance to human disease and its alleviation. EMBL is not, and has no pretension to become, a clinical research centre. Rather, EMBL research on disease mechanisms will be pursued in a collaborative manner with clinical partners from member states, to ensure rapid dissemination and medical application of our basic discoveries for medical application.

The EMBL Centre for Disease Mechanisms (CDM) was established to support medically relevant research at EMBL by expanding the interface between EMBL groups and clinical research activities at Europe's medical institutions. The CDM aims to promote the increasingly important application of basic research to the understanding, diagnosis and treatment of human disorders through molecular and biomedical exploration. To achieve its goal, the CDM will organise and sponsor a number of activities.

Dissemination and integration activities will constitute a main focus of the CDM. Such activities will provide better information within EMBL on available projects and expertise, as well as visibility to the outside by highlighting preclinical research within EMBL, thereby contributing to the organisation of more collaborations on medically relevant research topics. There is currently one EMBL partnership, the Molecular Medicine Partnership Unit (MMPU) (Section F), operating in this area and we see significant scope for an increased number.

The CDM website will provide an up-to-date resource of EMBL research on disease mechanisms, announcements of meetings and symposia on translational research topics, as well as links to members' websites. The CDM website will list CDM research projects, cooperative initiatives and requests for external expertise.

14 Towards systems biology of mitochondria

System-level characterisation has been carried out on single pathways and in defining metabolic networks in bacteria. There is now much interest in applying a similar systematic understanding to eukaryotic organisms and their organelles. Our aim is to construct a protein network of mitochondria as a first step towards a comprehensive, systematic characterisation of the organelle.

Mitochondria are highly conserved among eukaryotes and are well characterised by single-gene studies, yet a quarter of their component proteins remain unknown. All but nine yeast mitochondrial proteins are encoded in the nucleus and transported to mitochondria, making the mitochondrial organelle dependent on its cellular environment.

Four years ago, this project started by characterising genes involved in mitochondrial function through a systematic yeast gene deletion approach. More recently, high-throughput proteomics identified 550 proteins that co-purify with the organelle. We integrated 20 additional large-scale datasets to derive a combined list of 691 proteins for which we can adduce strong evidence for involvement in mitochondrial function. Interaction measures for each pair of proteins, based on genomic context, protein fusion, evolutionary conservation, co-expression, literature mining, database entries, and physical protein interaction evidence, now form the raw material for constructing a protein network.

Well-characterised mitochondrial modules, like the respiratory chain complexes, will be used as examples to test the accuracy and predictive power of the network. The connectivity, regulation, evolution and disease susceptibility of the mitochondrial system will be dissected with a systematic approach. The paradigm of systems biology is to improve a model continually until it accurately represents the real system. Given that our current understanding of mitochondria involves detailed understanding of a few complexes and a parts list, it is likely that the first model will only approximate the actual protein interaction network of mitochondria. Continued experimentation and model refinement will be performed.

In humans, 20% of the known mitochondrial proteins are implicated in genetic disease. It is likely that this number, though high, is an underestimate because the mutations that cause the much larger class of complex mitochondrial disorders are unknown. The dataset of yeast proteins and their functional interconnections will prove useful in defining new candidate genes whose mutation can give rise to human disease.

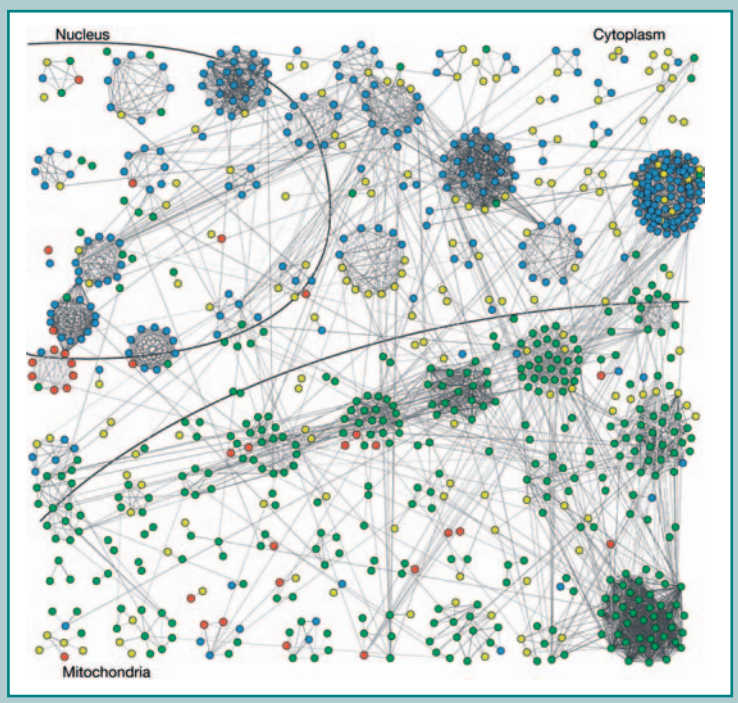


Figure depicting a protein network of mitochondria, also showing the interconnectivity of mitochondrial modules with those outside the organelle. Each node represents a protein: green and red are mitochondrial proteins verified by single-gene studies; yellow are proteins predicted by integration of systematic datasets to be associated with mitochondria; and blue are proteins connected by network predictions with any of the above

The CDM Forum will continue to host the popular EMBL-organised "Minisymposia on Molecular Medicine", which foster contacts between clinicians and basic researchers. The Forum will host roughly two symposia per year distributed between the different sites of EMBL.

We wish to add to these activities in the future by providing seed funding from central sources that will facilitate contacts with clinical researchers. The CDM MD–postdoc programme aims to attract and provide funding for young MDs that wish to join EMBL labs to work on disease mechanisms. This programme will target highly qualified young MDs from many European countries who wish to obtain training in molecular research to complement their medical studies. Many of these MDs currently choose to move to the USA for their postdoctoral training, where they have made a significant contribution to the recent increase in medically relevant projects being performed in some of the best basic research laboratories across the Atlantic. These MD–postdocs bring their medical expertise into the basic research lab and, taking advantage of the expertise and technology available, can open entirely new research projects targeted to investigate the mechanisms leading to human diseases. EMBL's outstanding research laboratories provide an excellent environment for ambitious MD–postdocs to engage successfully in biomedical research projects.

By analogy to our experience of targeted fellowships in the International PhD Programme (e.g. the Louis-Jeantet Foundation – supported fellowships for PhD students from eastern Europe), we are confident that the existence of a small number of funded positions in the Programme will seed a significantly greater activity by generating publicity for the Programme and will lead to many more applications for externally funded fellowships to be held at EMBL. As these MDs will go back to the clinic, this programme will result in the enrichment of European medical faculties and institutions with MDs highly qualified in molecular research. Furthermore, MDs trained at EMBL will maintain their links with the Laboratory and thus provide strong contact points between EMBL and clinical institutions.

Similarly, we intend to provide a limited amount of project funding, to "kick-start" pilot-stage collaborative projects through CDM grants. The CDM grants will provide a seed fund to sponsor the initial set-up of clinically oriented collaborations. These will be awarded on a competitive basis to support research initiatives for a limited time (up to one year). This start-up funding will allow preliminary exploration on the basis of which joint grant applications involving the participating EMBL and clinical groups will be formulated.

As is evident, the CDM has a rather different focus than our other Centres, being primarily concerned with enabling EMBL scientists to structure their research on disease mechanisms in the optimal way to allow the investigation of its relevance to human disease.

2.2. Chemical biology

Chemistry research at EMBL has been, until recently, largely devoted to synthetic and analytical chemistry of nucleotides and peptides. The highlight of this activity was the production, in the late 1980s and early 1990s, of a series of modified oligoribonucleotides that were useful for the structural and functional analysis of ribonucleoproteins, as well as for their efficient purification. However, in recent years, the interface between chemistry and biology has become more active. An increasing number of chemists tackle biological problems with the help of chemistry-based approaches. To date, however, chemical biology research in Europe is mostly located in chemistry departments, often distant from biological expertise and collaborators. We believe that chemical biology should be part of biology. Therefore, at EMBL, the chemical biology groups will be embedded in the existing Units to ensure maximal contact with wet-lab biology. At present, there is a single group in Gene Expression performing synthetic work, plus the Chemical Biology Core Facility (Section C.2.3.5). We are however convinced that EMBL should increase its investment in the field of chemical biology in the future.

There are two major areas where we wish to make use of chemistry in our research. The first is in the generation of tools that can be used to perturb and manipulate the biological systems we study. Small-molecule-based inhibition of biological processes is extremely useful as a complement to genetic and reverse genetic approaches. Chemical tools enable changes to be introduced to *in vitro* or *in vivo* biological assays in an acute manner, with a temporal resolution that cannot be achieved using genetic manipulations. Furthermore, genetic manipulations, although extremely powerful, bring with them the inevitable disadvantages of the possibility of causing indirect effects or inducing compensatory changes as a result of the fact that both forward and reverse genetics can only be used in experiments that have a long time course.

The development of chemical probes that can be visualised (e.g. by a coupled fluorescent reporter) is another important area. Such probes can provide both temporal and spatial information on the biological process under study. If several such probes are used simultaneously (e.g. in multiparameter imaging experiments), complex cellular processes can be investigated, thus complementing other systems biology approaches. Finally, dose–response analysis allows quantitative information on the importance of a single molecule or reaction to be obtained. Temporal and spatial resolution, coupled with quantitative analysis, are prerequisites for the systems-level analysis that EMBL will be pursuing in the next five years.

Two developments elsewhere at EMBL have increased the interest of our Units in the recruitment of additional chemistry expertise. The first is the adoption of high-throughput assay methods in various parts of the Laboratory. The miniaturisation and automation of assays for the activities of interest to us makes them ideal for use in screens for small-molecule inhibitors. Thus, small-molecule screening has moved out of the laboratories of "big Pharma" and into those of research institutes. In Europe, EMBL and the German Cancer Research Centre (DKFZ) has created one of the first such facilities (Section C.2.3.5). This facility presently allows roughly five such screens per year for EMBL groups. The tools generated in this way are conceived primarily for use in basic research, but some might also be useful in the context of technology transfer as they might be suitable lead compounds in drug discovery.

The second development is in live cell imaging. The use of cell-permeant chemical agents either as reporters of activity or to perturb activities is an extremely powerful approach to understanding biological processes. The one current EMBL research group in chemical biology is in the process of making fluorescent reporter probes or "caged" probes – molecules that are inert until activated (e.g. by exposure to UV light) – to enable multiple steps of intracellular signalling to be visualised and perturbed with spatial and temporal resolution. The goal of this research is in the understanding of cystic fibrosis and potentially in the design of strategies to ameliorate this condition. However, having this expertise in-house has led to a flood of requests for collaboration from other EMBL groups. The requests are various: for inhibitors or activators of specific processes; for artificial amino acids that can be incorporated into proteins using engineered tRNAs; for caged nucleotides, oligonucleotides and proteins; for labelled small molecules that serve as partners for endogenous or tagged proteins in cells; or for methods that will render proteins cell permeant.

These are all areas of active research in overseas institutes, and the positive effect on EMBL's research capabilities of having a single chemistry group active in some of those areas leads us to the view that we should recruit an additional, complimentary, chemical biology group as soon as possible. The plan is not to create a Unit for chemical biology but rather, as in the case of the existing chemistry group, to integrate the additional group into one of the Units, with Cell Biology and Biophysics or Developmental Biology being the most obvious candidates. This will ensure that the new group's efforts will have a biological focus rather than being exclusively devoted to method development. As described in Section 1.1.2, we will also incorporate a chemical biology group interested in interactions between small molecules and proteins into the Structural and Computational Biology Unit. A team leader in this area has just been hired at the time of writing.

The final area of chemical biology where EMBL intends to be active in the course of the next five years is in the construction and eventual public dissemination of better informatic databases on chemicals with

biological activity. There is currently very little organised information on this topic. The ChEBI (Chemical Entities of Biological Interest) database of the EBI is a first step in this direction. What is needed is the incorporation of much more information on chemical–biological macromolecule interaction at all levels, from the atomic structures of such complexes to lists of target molecules (including likely off-target interactions), for all chemicals that have biological activity. Like the other EBI databases, such an information resource would be of very widespread utility to life scientists from many disciplines.

2.3. Core Facilities

There are three reasons underlying the reorganisation over the past five years of certain EMBL activities into Core Facilities. The first is related to the high cost of equipment and materials required for certain experimental methods that have become a standard part in the repertoire of many EMBL groups, such as microarrays, MS of proteins, high-throughput robotic sample handling, and so on.

The second is that EMBL's research groups are typically of modest size, both in terms of the number of people and the size of their budget. Since these groups follow ambitious research programmes that require diverse technologies, it would be impossible for them to attract the capital and personnel investment required to carry out all their planned projects if they had to do so independently.

The third is a result of EMBL's experience with a previous model for providing access to critical, widely used technologies, which was to incorporate them into one of our research groups. This model proved unsatisfactory because groups, and group leaders, were placed in a constant conflict in having to divide their time and energy between, on the one hand, carrying out research and development projects (on whose basis they would largely be judged on leaving EMBL) and, on the other hand, providing service to their colleagues. Particularly for techniques where supply can virtually never meet demand, such as mass spectrometry or EM, the situation of the mixed service and research service groups was very problematic.

For all these reasons, it was cost effective and efficient for EMBL to inaugurate the Core Facilities to provide access to technologies that are either expensive to set up/maintain or require considerable expertise. The establishment of the Core Facilities was announced in the previous EMBL Scientific Programme (2001–2005) and approved by Council. Council supported the idea by giving the resources necessary for the first phase of establishing the Core Facilities. The current facilities are: Genomics; Proteomics; Protein Expression and Purification; Monoclonal Antibody; and Chemical Biology. All are highly dependent on facilities for robotic liquid-handling systems, advanced light microscopy and EM (both of which have tight links to the Centre for Molecular and Cellular Imaging), flow cytometry and transgenic production. We have asked SAC to organise expert review of these facilities in the same way as they do for our research Units, and the first such review will take place in Spring 2006.

2.3.1. Genomics Core Facility

The Genomics Core Facility provides two categories of service to EMBL groups. The first is sample based and comprises DNA sequencing and quantitative PCR. Although the capacity of these services is limited compared with that of dedicated DNA-sequencing laboratories, over 27,500 DNA samples were sequenced in the facility between July 2004 and June 2005.

The second category of service is project and automation based. The Genomics Core Facility helps research groups to miniaturise and automate their experiments and adapt them for robotic handling. These projects are mainly based on the use of microarrays. EMBL has access to several commercial microarray platforms and there are also projects for which research groups, together with the Genomics Core Facility, manufacture "designer" arrays at a level of roughly 2,500 slides per year. For just one of the commercial platforms, Affymetrix, current EMBL use is greater than 500 arrays per

year. As mentioned above, robotics and automation are a feature of several of the facilities and are also used elsewhere in EMBL, for example in a genome-wide RNAi screen for gene products required during mitosis. The Genomics Core Facility provides advice, expertise and in many cases the equipment required for the realisation of these projects. A further, much more-limited service provided is in support of serial analysis of gene expression (SAGE) projects. The SAGE approach has some advantages over microarray analysis, particularly in providing a quantitative analysis of transcripts over a wide range of abundance, and is therefore still used where this quantitative information is of the essence.

The Genomics Core Facility will continue to develop. The predictable challenges for the next five years include:

- Adaptation to novel nucleic acid arrays and methods. The advent of tiling arrays (arrays of overlapping fragments representing whole chromosomes or genomes) opens many new possibilities. Detailed mapping of all the transcripts produced from the entire genome under a given condition becomes possible. Genome-wide mapping of chromatin proteins, including transcription factor and histone-modifying activities, will be possible by ChIP-on-chip experiments. With certain types of array, based on short oligonucleotides, tiling arrays will become the method of choice for analysis of genomic variation. Such projects are planned at EMBL. They will require a significant upgrade of the computing infrastructure used for data handling in the Genomics Core Facility (both hardware and software) compared with what is currently available.
- Adaptation to new types of microarray platform. For example, requests have already come from EMBL groups who wish to analyse all the protein targets of specific modifying enzymes (e.g. kinases, glycosyltransferases) and we expect protein arrays to become the method of choice for such experiments within the next five years.
- Adaptation to the need for array and quantitative polymerase chain reaction (PCR) experiments on samples consisting of a few cells or small tissue samples. This will require further miniaturisation of the Genomics Core Facility experimentation and investment in equipment.

2.3.2. Proteomics Core Facility

The Proteomics Core Facility provides a range of protein and peptide identification services. The facility is a cooperation between EMBL and two industrial sponsors, Waters and Bio-Rad, who provide the majority of the equipment used in the facility in exchange for the right of access. A significant activity of the facility is in providing training and demonstrations for external users. The major techniques offered are:

- Protein separation and comparative sample analysis by 2D gel electrophoresis.
- Identification of proteins from 1D or 2D gels by MALDI-MS and ESI-Q-TOF-MS.
- Identification of proteins and peptides from samples purified by other means (fractionation methods, affinity purification or immunoprecipitation).

In 2004, the Proteomics Core Facility ran almost 700 2D gels (very close to the current capacity of the facility) and analysed over 3500 protein samples by mass spectrometry. The major users of MS in the facility, accounting for roughly two-thirds of the total use, are the structural biology Units. This use reflects the requirement for careful characterisation of the purified proteins used for structural analysis. However, it also illuminates the current limitation of the Proteomics Core Facility with respect to other major users of MS. The current equipment in the facility lacks the sensitivity required to identify low abundance proteins (quantities that can only be silver stained) or to identify post-translational modifications that are virtually always present at low abundance in purified protein samples.

Projects that produce samples requiring high sensitivity and that require identification for post-translational modification are common in the molecular, cellular and organismal biology Units. These can currently only be pursued in collaboration with the MS research and development group of Matthias Wilm. The equipment required for both of these techniques, as well as to increase throughput, has to be implemented in the Proteomic Core Facility during the next EMBL Programme. Additionally, EMBL will need to consider adding new technologies such as high-throughput quantitative protein expression analysis and identification in the next five years.

2.3.3. Protein Expression and Purification Core Facility (PEPCF)

Virtually every research group at EMBL must clone and express recombinant proteins, the exceptions being those solely engaged in computational analysis. Although protein expression and purification is widespread in all molecular biology laboratories, there is scarcely an aspect of our work that is as idiosyncratic as protein expression. Each protein and protein fragment interacts differently with the expression host, and finding conditions to express a protein at reasonable yield in soluble form is often the rate-limiting step in a project.

The PEPCF provides invaluable advice and help to users at each step of the expression process. It has collections of vectors and host strains for expression of proteins in prokaryotes, and advises users on which to try first, and under which conditions, depending upon the properties of their specific protein. The PEPCF can also assay, at small scale, a limited matrix of expression conditions and will provide further hands-on help in difficult cases.

Further activities of PEPCF include production of several proteins that are in use in many laboratories. Together with the monoclonal antibody facility, it helps produce and purify stock anti-protein tag antibodies that are in use throughout EMBL for biochemistry and immunofluorescence analysis. PEPCF is in the process of establishing yeast fermentation for the large-scale production of yeast complexes for biochemical and structural analysis (Section C.1.1.1., Box 2). Finally, PEPCF is responsible for the maintenance of the equipment required for a diverse set of biophysical analytical techniques, and for training and advising EMBL scientists on their use.

In the future, a major development will be the provision of eukaryotic cells (e.g. yeast, Chinese hamster ovary cells) and vectors for expression and, together with the Genomics Core Facility, automated screening for optimal expression conditions.

Part of the protein expression and purification pipeline at EMBL Grenoble is the ESPRIT (expression of soluble proteins by random incremental truncation) platform. This supports automated screening for soluble protein fragments. Our intention is to develop this with other protein expression activities (as well as the appropriate other aspects of automation of protein X-ray crystallography) in close integration across the Heidelberg, Grenoble and Hamburg sites.

We plan to extend the ESPRIT platform in the next EMBL Programme to incorporate an automated gene-cloning service. This service would be of tremendous utility to those engaged in structural and functional genomics projects. It was the most frequently requested addition to the Core Facilities when we surveyed our research staff in Summer 2005.

2.3.4. Monoclonal Antibody Core Facility (MACF)

The utility of antibodies in modern biology is unquestioned. For instance, they are used to test for the presence and quantity of a given protein in a specific sample or cell, to localise proteins (by light microscopy or EM), for native purification of a protein either alone or together with its functional partners, and to determine the conformational or modification state of a protein.

For many of these purposes, monoclonal antibodies are the tools of choice. They have the additional advantage over polyclonal antisera of being derived from an immortal cell line and thus being inexhaustible. The EMBL MACF was created to streamline the production of monoclonal antibodies in-house, and also to offer its services to external users. In the past 18 months, just over 300 projects have been initiated by the staff of four, divided almost 1:1 between internal and external users. Of the 219 projects completed in that time, 85% produced monoclonal antibodies that were fully satisfactory to the users. Anyone familiar with monoclonal antibody production will appreciate that these are very favourable statistics.

In general, users are happy with monoclonal antibodies produced from an enriched mixture of hybridoma cells. MACF, however, will need to invest in equipment and staff that will allow cloning to monoclonality in a larger proportion of cases in the future, as well as making the scaling up of antibody production more readily available.

Other future plans include:

- The use of genetic immunisation to replace protein or peptide injection.
- The use of mutant spleen cells, from mice with a Robertsonian translocation, for cell fusion. This will allow direct selection for antibody-producing hybridomas.
- The use of immunosuppression to reduce the likelihood of producing antibodies against an undesirable epitope (e.g. a non-phosphorylated peptide if the desired antibody should recognise the phosphorylated species).

2.3.5. Chemical Biology Core Facility (CBCF)

As discussed previously, the use of small-molecule inhibitors of biological processes can be of great utility in basic research. In addition, such molecules can sometimes provide the starting point in the development of a therapeutically useful compound. EMBL groups wish to identify small-molecule inhibitors for both these reasons.

The purpose of the CBCF is to aid research groups in modifying bench assays such that they can be used as the basis to screen for an inhibitor of the process under study. This always involves scaling assays down, and usually requires the modification of the assay to provide a readout that can be easily monitored in a multi-well assay format.

The main compound library used for screening at EMBL comes from a commercial partner, although it is supplemented by smaller libraries from academic laboratories. The CBCF is a joint venture with the DKFZ, and current capacity allows for roughly five full-scale screens from each institute to be carried out each year. Although the CBCF is still at a very early stage, the first results have been very satisfactory.

In the future, the CBCF will need to extend its compound library, which will be a significant cost. Furthermore, in common with the other chemical biology activities at EMBL, increased access to cheminformatic resources is a necessity. There are plans for collaboration with the Genomics Core Facility, to combine small-molecule inhibition with microarray analysis, and with both the PEPCF and the Heidelberg SCB Unit to develop and implement methods to assay interactions of small molecules with their targets.

2.3.6. Advanced Light Microscopy Facility (ALMF)

The ALMF was the first of EMBL's Core Facilities. It started as a bottom-up initiative within the Cell Biology and Biophysics Unit, whose goal was to modernise the light microscopy equipment available

to EMBL researchers. It represents cooperation between EMBL and a considerable number of industrial partners, who install and maintain their latest equipment in EMBL because they know that biologists from throughout Europe look to EMBL for advice and help with problems in this field.

Roughly 25% of the currently 22,500 hours per year of use of the ALMF microscopes is by external scientists through the Visitors Programme. Visitors usually come in small groups of two or three to try out the equipment and to carry out experiments. In addition, the ALMF is the founding member of the European Light Microscopy Initiative (ELMI), a consortium of laboratories with a common interest in advanced light microscopy. ELMI organises courses and workshops to train users in the techniques of light microscopy and the ALMF participates in both this training and in organising courses (as part of the Centre for Molecular and Cellular Imaging) for in-house researchers.

Aside from providing a wide variety of straightforward microscopic tools and the accompanying expertise to users throughout the Laboratory, the ALMF allows sophisticated experimentation. Examples are multicolour 3D still imaging and time-lapse microscopy and a variety of methods for manipulating fluorescent probe molecules in live cells such as FRAP, FLIP, FRET, photoactivation and correlation spectroscopy.

The heavy use made of the ALMF equipment referred to above, and the reliance of groups from throughout EMBL's Units on the advice of the ALMF, makes it extraordinarily valuable to the Laboratory. However, like the other facilities, the ALMF constantly has to look to the future in order to maintain its competence to support cutting-edge research.

The following are a selection of the future plans for the ALMF:

- Quantitative imaging of individual biological molecules and of interactions between them in complex systems.
- Imaging biochemical reactions in cells and organisms.
- Working, in the context of the Centre for Molecular and Cellular Imaging, to integrate light microscopy and EM images (the Electron Microscopy Core Facility is described below).
- Setting up and participating in high-content, high-throughput screening activities.
- Incorporating EMBL developments, such as SPIM and FLIM, into the ALMF to allow their wide use.
- Providing advice on image processing and analysis, in the context of the Centre for Molecular and Cellular Imaging.

2.3.7. Electron Microscopy Core Facility (EMCF)

EMBL has a considerable history of research and development in EM, particularly in the development of cryofixation methods and in EM hardware development. In recent years, however, our capacity to carry out EM work for cell biology had become sub-critical and has been restricted to individual research groups. Because a high general demand for EM access remained, this led to the setting up of the EMCF in early 2004.

The first goals of the EMCF have been achieved. With three staff members, the EMCF has established high-pressure cryofixation and cryosubstitution. More standard cryosectioning and plastic embedding methods are also currently available in the facility. The EMCF works together with users to optimise protocols to be adapted to the variety of sample types under study, and to provide the possibility of immunolocalisation on optimally preserved samples. EM is of course extremely labour intensive, and users must participate in the analysis of their samples after being trained by the EMCF staff.

The EMCF will also work on the development of image analysis software within the context of the Centre for Molecular and Cellular Imaging. An additional goal is the establishment, with the ALMF, of correlative light microscopy and EM. This is essential to understand the changes in ultrastructure underlying the events visualised in normal and experimentally modified cells by fluorescence microscopy.

2.3.8. Flow Cytometry Core Facility (FCCF)

The smallest Core Facility at EMBL is the FCCF. Flow cytometry was previously carried out by one of the research groups in Biochemical Instrumentation. With the dissolution of that Unit, a new solution had to be found. In Spring 2004, the FCCF was founded with a single person to run it. In the following year, over 200 cell-analysis experiments, 150 cell-sorting experiments and 15 embryo-sorting experiments were carried out. However, approximately half of the time was used for equipment renovation and maintenance and for user training.

Uses range from the sorting of small cell populations (e.g. the migrating border cells in a *Drosophila* egg chamber for microarray analysis), to the examination of the effects of gene mutation on the production of specific cell types within a mouse tissue or organ. As in the case of the other Core Facilities, user surveys suggest a considerable increase in the demand for cell sorting and analysis will occur over the next EMBL Programme.

In addition, new projects for the future include:

- Development of novel FRET methods for use with the cell sorter. This would have two significant advantages over current FRET methodology to measure intermolecular interactions. First, the FRET signal would be measured in each individual cell in a population, rather than representing a population average. Second, cell populations with FRET signals of different intensity could be fractionated and assayed further to see the consequence for the cell of the differences in the level of biomolecular interaction revealed by the FRET assay.
- Adaptation of flow cytometry together with other Core Facilities such as the ALMF and CBCF, for use at high throughput. There is currently no commercial equipment capable of supporting this application. If successfully developed, this could for example be used in conjunction with genome-wide RNAi or chemical compound screens to look for modifiers of cell division and growth.

2.3.9. Transgenic Core Facilities (TCFs)

EMBL provides three facilities for production of transgenic animals.

The *Drosophila* transgenic service has been in operation in Heidelberg for over ten years. Since 2003, it has also provided an efficient low-cost service producing transgenic *Drosophila* to both the internal and external scientific community.

Facilities for production of transgenic mice are located in Heidelberg and in Monterotondo. Both provide the capacity to produce genetically modified cells and animals for EMBL researchers. The TCFs also aim to serve the European scientific community as capacity permits, providing limited access to standard mouse transgenic and gene-knockout production, and rederivation and cryopreservation services. In addition, the TCFs offer biologists working in other areas an accessible and affordable opportunity to expand their research into mouse models without prior experience in transgenic mouse production. Future projects at the Monterotondo TCF will be focused on generating more accurate models of human pathologies and multigenic disorders, and will include the continued refinement of gene regulation through the development of conditional and inducible mutagenesis schemes.

2.4. IT infrastructure and services

The IT infrastructure required to support basic molecular biology research has increased dramatically in size and complexity over the past 20 years. EMBL-EBI has been at the forefront of the development of IT infrastructure for bioinformatics because of its important role in scientific service provision. IT infrastructure at the other EMBL sites has been less prominent; however, this is changing rapidly. Scientists are collecting data not only in traditional individual laboratory experiments but also increasingly using large-scale high-throughput approaches. Technologies such as high-throughput microscopy, small-animal imaging or electron tomography, which are now applied regularly throughout EMBL for answering biological questions, generate very large quantities of data. We expect data collected by EMBL scientists to be in the petabyte-plus range within the next few years. These data not only have to be stored reliably, they also need to be easily retrievable for analysis and further processing. These developments result in three main requirements for IT infrastructure at EMBL:

- to establish a capable and user-friendly data exchange platform and intranet;
- to provide a highly scalable and reliable infrastructure for high-performance storage and computing;
- to simplify and reduce complexity by implementing new solutions and processes.

IT Services and the Structural and Computational Biology Unit operate the largest IT infrastructures in Heidelberg; Monterotondo is fully supported from Heidelberg, whereas the Outstations in Hamburg and Grenoble currently operate their own IT infrastructure. Simplification and standardisation across these sites will be important for implementing new solutions and processes. We will have to integrate existing resources better, replacing the existing home-grown solutions by industry standard components to reduce costs and increase efficiency, and to work towards a shift of resources into a central IT budget.

Improving service provision to the EMBL community will require improvement in communication processes, service quality and increased transparency of EMBL's IT organisation. The goals are to build trust with the users, to be recognised as reliable and expert service partners, and finally to achieve general user satisfaction.

These challenges apply to all EMBL sites. IT specialists throughout EMBL will work towards a unified approach whenever possible to gain maximum benefit for the whole organisation. The special requirements for, and demands on, IT services at EMBL-EBI will however continue to make this Outstation an exception with regard to IT services.

2.4.1. EMBL collaboration platform and intranet

The need for an institution-wide collaboration infrastructure derives from EMBL's plans to move into systems biology, for which efficient interdisciplinary inter-Unit research is essential. A new collaboration portal will cover the key aspects for projects, including easy file sharing, full text search, central calendaring, contacts and task management, desktop conferencing and real-time messaging. In addition to EMBL-wide collaborations, it will allow collaboration with external partners in the member states.

The planned collaboration portal can also serve as the central information resource for EMBL organisational data and form the technical basis for a future EMBL intranet. A dedicated intranet currently does not exist within EMBL. However, we consider an intranet portal to be important in facilitating future collaborative work. Tight integration of the intranet and collaboration portal will help to avoid redundancy and content management overheads.

2.4.2. High-performance storage and computing

The increase in high-throughput/high-content readout technologies used in large-scale experimentation will generate higher demands for storage, compute power and network bandwidth.

Data storage and backup

Demands for data storage are expected to double every 12–15 months. We will develop a plan for storage and backup architecture that meets this challenge and that can keep up with the pace expected from the research groups. The system will of course be designed to find a balance that optimises performance, scalability and reliability with the total costs for investment and maintenance of that infrastructure.

High-speed processing of huge amounts of data (e.g. high-throughput image processing in a cluster) uncovers a problem that increasingly affects EMBL and similar high-performance computing sites. The current file system technology is too slow for cluster approaches, creating an Input/Output (I/O) bottleneck. This issue needs to be addressed in order to take full advantage of available compute power. Appropriate solutions could be the use of parallel file systems, which is a new technology that will soon be in a mature state. Storage technologies such as this will be evaluated in close collaboration between IT Services and EBI storage experts.

Data processing

The expected growth rate for compute demand is likely to follow the rate of storage growth (i.e. it is expected to double every 12–15 months). Matching this demand requires a threefold approach:

- Individual clusters should be merged into one logical infrastructure (one queuing system) in order to take maximum advantage of the setup. Depending on agreement of the different users, this could be accomplished in a step-by-step process over a certain period of time.
- In-house cluster facilities must be regularly renewed with a turnover time of 2–3 years maximum, otherwise the overhead costs for power consumption (primary and cooling), and for management of otherwise increasing numbers of machines, will unbalance the total cost.
- Remote computing will have to take increasing volumes of the total computing load. This can occur within Grid computing approaches, partnership projects with supercomputing centres or showcase projects with different hardware vendors. Apart from appropriate funding, the major prerequisites for remote computing will be to build and manage the partnerships on the business side and to establish a high-bandwidth network link to the remote computing site.

Geant2, a virtual campus network infrastructure, is an example of such a collaborative effort. It is perfectly designed for experiments that generate large amounts of data and are likely to create an increasing demand to move huge data volumes between EMBL sites, or to and from external project partners, respectively. Funded by the EU, it provides an infrastructure to create pan-European high-speed links of up to 10 Gigabit. This could form the basis for a virtual high-speed inter-campus network and would enable efficient sharing of high-performance computing infrastructures across EMBL sites. Additionally, it allows the construction of data pipelines, for example from producer to analysis sites or to outbound content provider sites such as EMBL-EBI. Since Geant2 infrastructure already exists, it could be used in scientific collaborations very soon. The financial prerequisites for EMBL to use Geant2 would involve investments in networking hardware, the costs for Geant2 connectivity, and the internal costs for connectivity set-up and maintenance.

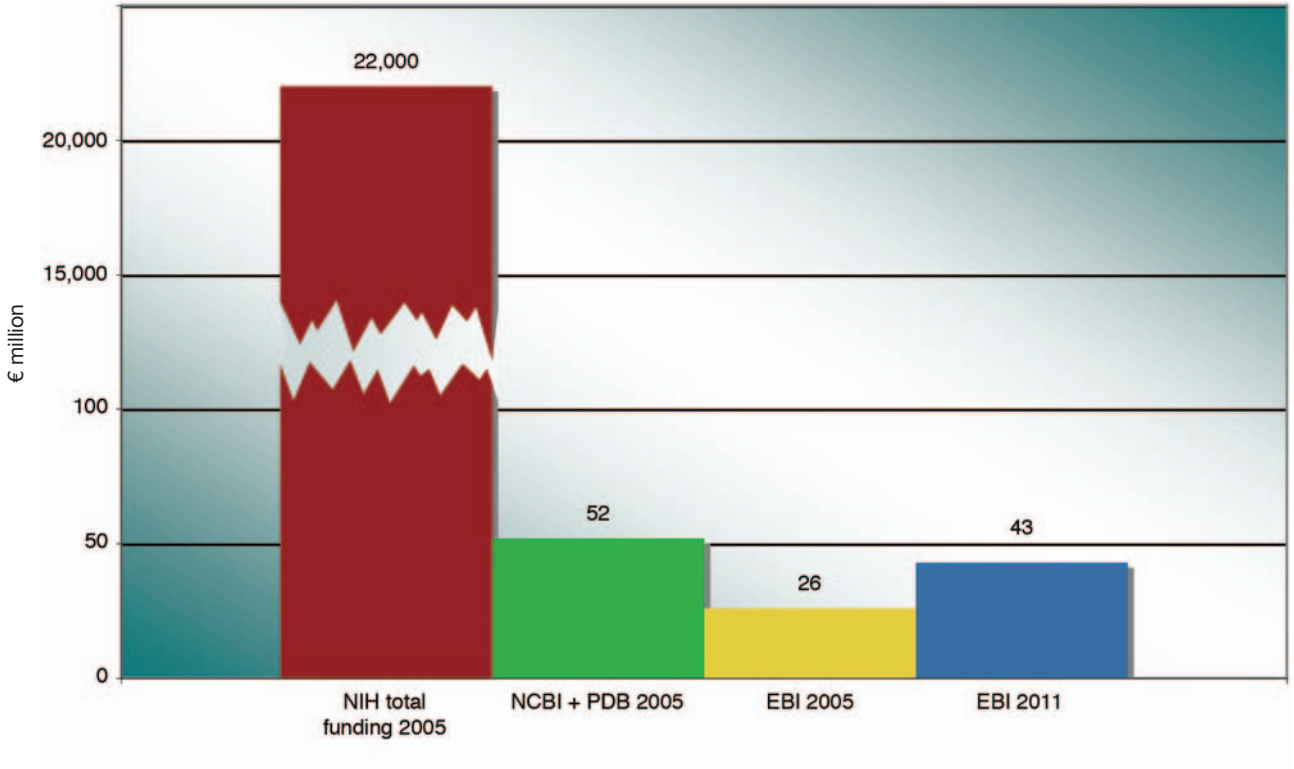
D. SERVICES TO THE MEMBER STATES

1. Service provision by the European Bioinformatics Institute

EMBL-EBI provides the core biomolecular resources in Europe. Its collections include nucleotide and genome sequences, information on protein sequences and functions, macromolecular structures and data on gene expression, proteomics, metabolic pathways etc. The EBI bioinformatics services have captured data from experiments on which many billion euros have been spent over the past two or three decades. As a single example, the US National Institutes of Health (NIH) spends €22 billion per year on biomedical research. Regenerating the existing data in today's most efficient high-throughput labs would be extremely expensive were it to be lost and it is therefore essential that the databases are stably and continuously maintained. The data resources held by the EBI support all facets of life science, a large and growing area. The total user community served by the EBI is of the order of a million users (estimated from web logs) and roughly 200,000 of these are regular users. Bioinformatics is employed extremely widely and supports the work of scientists in wet laboratories by reducing tasks that would take months or even years of laboratory work to hours of computer work. Research that is supported by the databases encompasses all facets of basic research in biology and medicine, as well as more distant fields ranging from the development of new drugs to agriculture, nutrition, forestry and fishery. EMBL-EBI represents Europe in international database consortia to ensure that European scientists retain access to the largest and most up-to-date data collections from around the world.

The amount of data that needs to be captured has been growing exponentially since the beginning of the Human Genome Project and will continue to grow in the same way in the foreseeable future. As described above, the cost of data acquisition is very high compared with the cost of capturing that data and making it available to users in a sustainable way (Figure D.1.). EBI's funding compared with its nearest US equivalent, the National Center for Biotechnology Information (NCBI), is modest: NCBI has a budget of €48 million per year today which is already more than we are asking for the EBI in 2011; in fact, the EBI's mandate includes the macromolecular structure database, whose US equivalent currently receives additional funding of roughly €4 million.

Figure D.1. Comparison of the annual cost of data acquisition with the annual budgets of NCBI including PDB, EBI in 2005 and EBI in 2011



1.1. The paradox of EBI funding

The services provided by the EBI are simultaneously one of the easiest aspects of EMBL to justify and the most difficult to fund. This apparent contradiction is based on the following facts. Roughly 200,000 scientists make regular use of the EBI's information resources, and, as already mentioned, the total number of users is roughly a million. The users access over 2 million pages of information per day (Figure D.2.). Most users are from EMBL's member states, making the data resources an essential tool that enables scientists in a variety of life science disciplines to carry out their daily work. The funding problems stem from the fact that the EBI was set up on the basis that direct contributions to the EBI from the member states, through the core EMBL budget, would cover only a fraction of the total costs, with other sources providing the majority of the funding. The member states envisaged that they would provide much of the additional funding indirectly, through the EU.

In 1998/99 EBI lost about one third of its staff following a discontinuity in EU funding and the member states had to be asked for an exceptional contribution in 2000. This situation was very damaging to both EBI function and staff morale, and recovery from it was slow and demanding – we cannot let it happen again. The EBI services are an essential infrastructure for supporting European research and, as such, the EBI's mission enjoys substantial EU goodwill. However, the existing EU funding mechanisms were not designed to support ongoing service activities such as those of the EBI and it currently seems unlikely that the EBI will be in a position to receive more stable funding from the EU in Framework Programme 7 to support the biomolecular data resources. The recent EU budget decisions make it very likely that research is not going to receive the priority it deserves and the report of the Competiveness Council from November 2005 holds out little hope for major new spending on research infrastructures within FP7. We therefore need to make sure that sufficient funding is provided by the EMBL member states to ensure the continued operation, maintenance and upgrading of EBI's data resources. Currently, EBI funding is almost equally split between EMBL and a variety of external funding organisations (Figure D.3.).

Figure D.2. Use of the EBI data resources

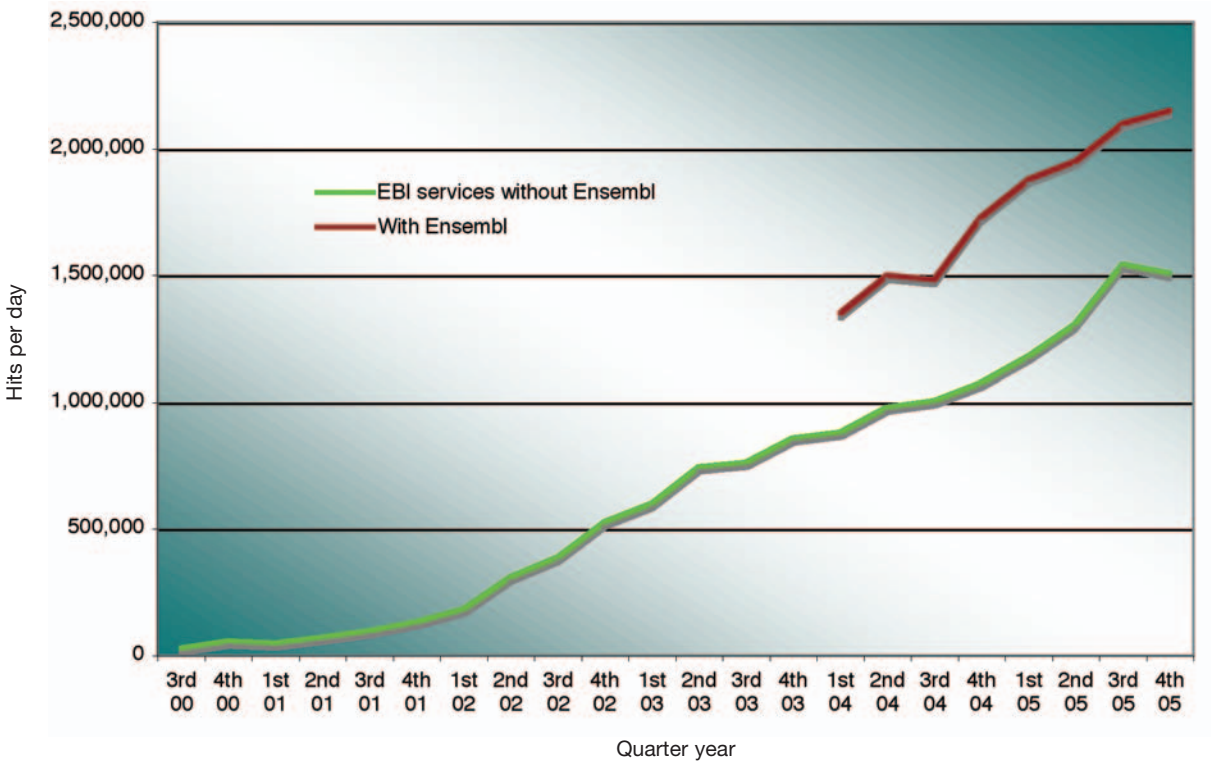
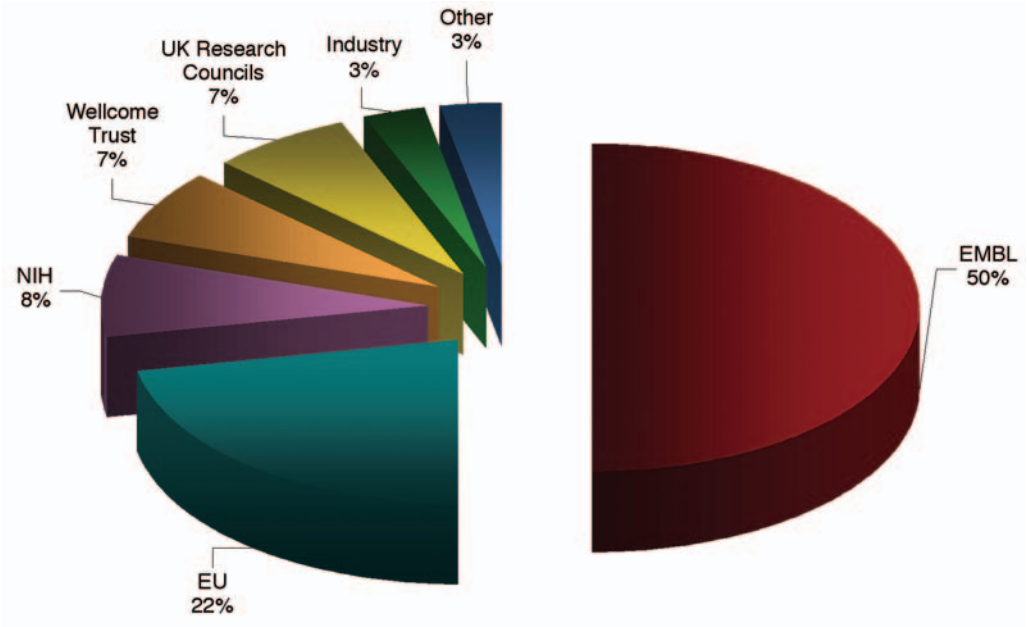


Figure D.3. EBI funding 2005



1.2. The EBI service mission

The EBI collects, organises and distributes a wide range of core biomolecular information in the form of the following range of complementary databases:

Ensembl	Genome information
EMBL-Bank	The DNA sequence archive
UniProt	The global protein sequence database
eMSD	Macromolecular structure information
ArrayExpress	The transcription database
IntAct	Protein–protein interactions
IntEnz	International standard for enzyme nomenclature
Reactome	Molecular processes and pathways
InterPro	An aggregation of different databases on protein functional motifs
GO	Gene ontology
Integr8	A database that provides integration of gene-related information from other component databases
ChEBI	Chemical entities of biological interest

Many of these databases are large, and most are still growing exponentially. For example, a new nucleotide sequence is currently deposited into the EMBL-Bank about once per second. The structure and sequence databases are global collaborations and data are exchanged regularly (in some cases daily) with both the USA

and Japan. Ensembl is a collaboration between the EBI and the Sanger Institute, and has strong international links. ArrayExpress was the first public microarray data repository. Its founders were prime movers in establishing standardised procedures for microarray experimentation – a prerequisite not only for comparing data from different sources but also for establishing the conditions in which a central data collection can be useful.

Some of the more recently initiated databases, though smaller, are growing rapidly in importance. These include IntAct, the protein–protein interaction database, and Reactome, which contains information on pathways and networks. Very recent initiatives include PRIDE (a resource for proteomic data) and BioModels (a prototype database of modelling methods for use in systems biology). In the course of the next five-year programme, we envisage a need to add further resources serving data on, for example, metabolites, pathways and electron microscopy (EM) tomography. New demands for data resources from the scientific community arise regularly and must be judged in the light of both their merit and the limited capacity of the EBI resources to support additional activities.

Offering any one of the databases is a very substantial and complex task involving several stages.

- **Designing and building databases.** This is scientifically demanding as it requires an in-depth understanding of the research fields addressed, and thus draws heavily on relationships with the research groups in the EBI and the rest of EMBL. It is also technically demanding and has required that the EBI assemble formidable expertise in IT.
- **Collecting and curating the scientific content of the databases.** Data curation involves intensive hands-on work from scientists who understand the fine detail of the findings represented and who also have the computational skills necessary to ensure whole database integrity, for example in the consistent use of terminology or the creation of correct cross-references between databases.
- **Serving the data to users.** Typically, any single database must actually build a separate representation of the data that has been optimised for user service based on the master production database, which itself is optimised for data integrity. These data are then served to users through web pages and computer programs developed or installed by a service development team. A complex range of mechanisms enables user sites to maintain mirror copies of databases in synchrony with the EBI's master copies.
- **Outreach and training.** Exploiting the full range and complexity of data requires considerable understanding from the users, which is supported by the provision of high-level training, extensive documentation and responsive help-desk support.

Box 15 describes the work of the UniProt database as a single case study.

15 The Universal Protein Resource (UniProt)

The amount of information available about proteins continues to increase at a rapid pace. Protein interactions, expression profiles and structures are being discovered on a large scale, and completely sequenced genomes cover the taxonomic tree with both breadth and depth. Biological and biochemical functions of individual proteins continue to be elucidated. Furthermore, improved analytical tools are available to make intelligent predictions about function, localisation, secondary structure and other important protein properties.

The ability to store and interconnect this expanding universe of protein information is crucial to modern biological research. The Universal Protein Resource (UniProt) provides the central resource on protein sequences and functional annotation, and is an extension of work previously carried out by the Swiss-Prot + TrEMBL and PIR-PSD protein databases.

The primary mission of the UniProt consortium is to support biomedical research by maintaining high-quality protein sequence and function databases that are freely accessible to the scientific community. The core activities in UniProt include sequence archiving, manual curation of protein sequences assisted by automated annotation, development of a user-friendly UniProt web site, and interaction with other protein-related databases to expand cross-references. The UniProt Knowledgebase (UniProtKB), comprising the manually annotated UniProtKB/Swiss-Prot section and the automatically annotated UniProtKB/TrEMBL section, is the pre-eminent storehouse of protein functional annotation. The extensive cross-references to about 100 molecular databases, together with functional and feature annotations, and literature-based evidence attribution, enable scientists to identify and analyse proteins and pose questions about their functions across databases with complementary information. The UniProt Reference Clusters (UniRef) speed similarity searches through sequence space compression, which is achieved by merging sequences and sub-sequences that are 100% (UniRef100), 90% (UniRef90), or 50% (UniRef50) identical. Finally, the UniProt Archive (UniParc) stores all publicly available protein sequences, containing the history of sequence data with links to the source databases.

These developments benefit scientists in Europe and worldwide in several direct and indirect ways:

- by providing a stable and comprehensive resource for information on proteins, their sequences and their functions;
- by enabling scientists to use these data to identify and analyse genes and their products, and to make queries across databases containing complementary information;
- by providing efficient and unencumbered access to the databases.

Previously, Swiss-Prot + TrEMBL and PIR-PSD coexisted as protein databases with differing sequence coverage and annotation priorities. In 2002, the Swiss-Prot + TrEMBL groups at the Swiss Institute of Bioinformatics (SIB) and EBI, and the Protein Information Resource (PIR) group at Georgetown University Medical Center and the National Biomedical Research Foundation, joined forces as the NIH-funded UniProt consortium (http://www.uniprot.org). In recognition of the strength of SIB and EBI in protein function databases, 75% of the current NIH grant has been awarded to the European consortium members, with the EBI being the Principal Investigator site. Although the ability of EBI to attract NIH funding for one of its flagship database activities is a clear sign of the international reputation of EBI and EMBL as a whole, it should be of some concern that NIH now has certain rights over one of Europe's most prestigious database projects and that Europe was not able to find its own long-term funding solution for Swiss-Prot + TrEMBL.

1.3. The user community

As we have indicated, the EBI's user community is numerically large. However, to demonstrate the utility of the EBI, it is also extremely informative to look at the diversity of the community that it reaches. Today we see users from several different areas:

- **Fundamental biological research.** The molecular processes of life, many of them universal, can only be understood with reference to the information held by the EBI. Essentially, every laboratory conducting biological research will use information that is present in the EBI data resources. This group represents a majority of the users.
- **Medicine.** The mechanisms of disease are often understood by studying abnormalities at the molecular level. Sequence databases including information for instance on the chromosomal localisation, sequence and structure of genes (as provided by Ensembl) are crucial for clinical research. Such medical research already exploits our services heavily and, as clinical research becomes increasingly molecular, our resources will be exploited in the clinic for diagnosis and prognosis.
- **Pharmaceutical research.** Drug molecules work by interacting with biological systems. By referring to databases of structures, we can understand the structural and biochemical nature of these interactions. By comparing global gene expression patterns in the presence and absence of the drug, the specificity of its effects can be estimated. Databases used in these ways yield insight into drug action and clues to the development of better drugs with new mechanisms of action. The vast majority of Europe's major pharmaceutical companies take advantage of the EBI Industry Programme to enhance their bioinformatics capabilities. The newer, smaller biotechnology companies, often working at the forefront of innovation, also exploit our services and participate in our forum for small-to-medium enterprises (SME forum).
- **Nutrition.** The effects of different nutrients can be investigated analogously to those of drugs. Companies involved in nutrition research are moving towards such molecular approaches and we expect an increasing number to join the Industry Programme.
- **Agriculture and animal farming.** Understanding the genetics of crop plants and farm animals enables exploitation of the natural mechanisms that confer disease resistance and generate high yield. Investigation of the molecular mechanisms of agro-chemicals enables them to be better targeted to their purpose. Such science, crucially dependent on the underlying data, is widespread from conventional agriculture, to fish farming, to forestry. The well-being of humans in marginal agricultural environments can also be much improved by the development of crops whose genetic make-up enables them to withstand harsh (e.g. salty) environments.

These diverse users access data and bioinformatics tools of various sorts. Roughly 90% of the data deposition to the EBI is from Europe, and roughly 65% of its data provision is to European users.

Users are not only distinguished by their research field, we also serve members of the scientific community in different ways depending on their level of experience with bioinformatics databases and tools. The general scientific community (academic and commercial) submits their data (which is often required for publication in peer-reviewed journals, e.g. DNA and protein sequences, macromolecular structures, gene expression data, protein interaction data etc.), downloads information through the EBI website (more than 2 million hits per day), performs analysis over the net using tools provided by EBI and needs training in the use of the EBI resources. User communities are actively involved in developing standards and providing input for the development of the next generation of data resources (e.g. proteomics community, microarray community). EBI has been a driving force in organising these scientific communities and in helping them develop standards and data repositories.

More sophisticated users are "power users" who often maintain local copies of whole databases kept up-to-date using EBI services. Such users produce their own tools and databases. The EMBRACE project for example, which is an EU-funded Network of Excellence, will develop standards for data analysis tools using local copies of EBI databases as their raw material. Another such project involves the development of standard graphical language for describing pathways and processes.

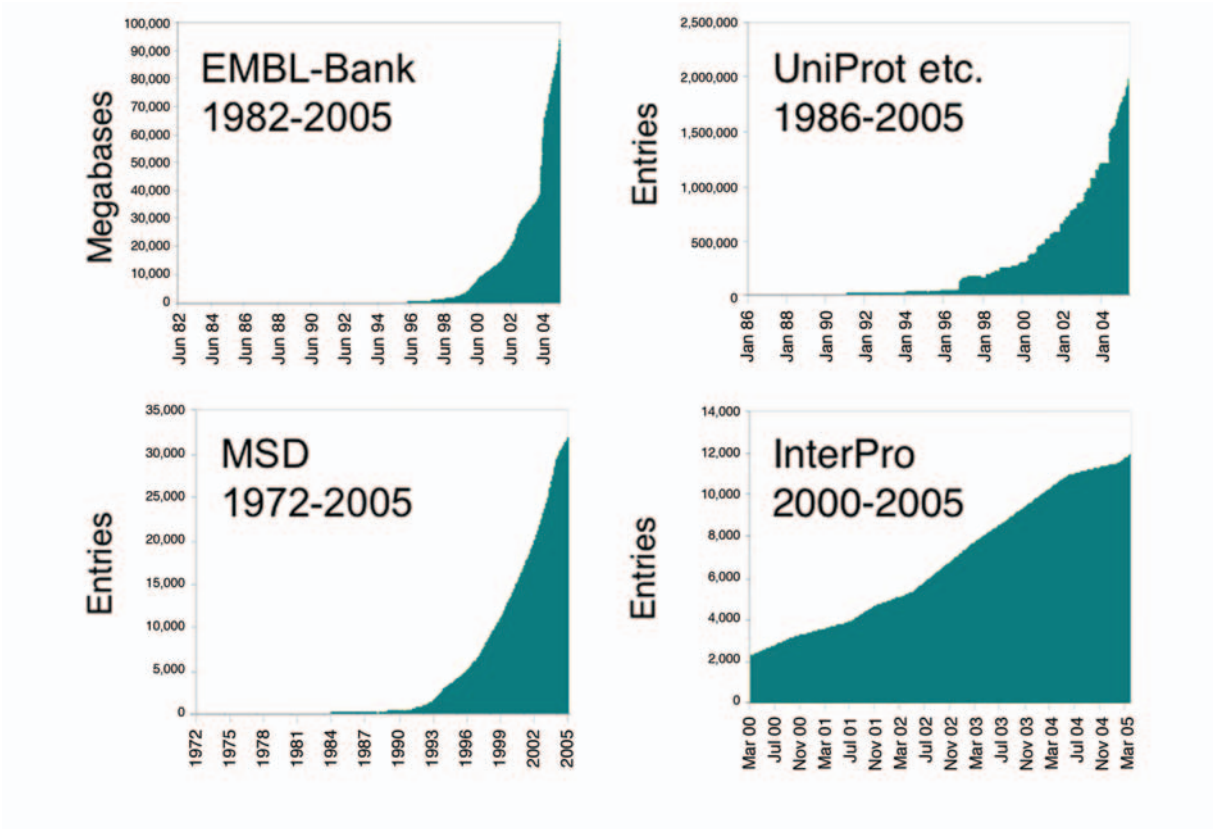
1.4. Growth of the EBI

In all the fields of EBI service activity, data deposition is growing rapidly (Figure D.4.). But this is only one of several reasons why the EBI needs to increase both its personnel and computational resources.

Sheer volume of data is in itself a challenge, but equally substantial challenges come from the inherent complexity of the data, for instance from macromolecular structures or microarray expression experiments, and from the need to create interconnections between databases representing different aspects of biological processes. Increasingly, we also need to develop sophisticated tools for harvesting data automatically and directly as the experiment is carried out.

An additional growth area is in the demand for high-quality annotation. Users want the DNA or protein sequence they find, or the genes whose expression is upregulated under the particular experimental conditions that interest them, to be seamlessly linked to as much functional data as possible. Although annotation of databases can be automated to some extent, such automation typically relies on similarities between new data and data on related molecules or processes that has been collected by careful manual annotation, drawn from information found scattered through the published literature. This renders investment in manual annotation doubly important, because its quality, good or bad, is amplified when it is propagated to related systems by automated annotation mechanisms.

Figure D.4. Growth of EBI data deposition



As the size of all the databases increases, the computational demands required to search, handle and integrate the data increase even more rapidly. For example, a common bioinformatics task is to compare two sets of data, or one set against a whole database. The computational power required to do this increases with the square of the size of the sets of data to be compared; this, combined with the burgeoning size of the databases, creates ever-increasing demands for computational capacity.

Finally, demand for new data resources is very high. To name but a selection, life scientists want databases that hold information on chemicals and metabolites, human variation and phenotype, and light and electron microscope images. The connection of all these types of data to each other and to the scientific literature is central to the research record, and the EBI has long been involved in the development and incorporation of such links into its information resources. Recent developments include the provision of basic literature search mechanisms at the EBI. As the availability of electronic full-text becomes widespread, the provision of literature mining tools in combination with the existing databases will become a core activity of the EBI.

1.5. Resources and focus

The demands on the EBI will exceed what is possible, and we will need to select carefully so as to maintain our capability to deliver on our primary mission – to provide the core biomolecular data. However, even this will be impossible without significant growth. The 2003 SAC review panel, the annual Bioinformatics Advisory Committee (BAC) meetings and the UK research funding organisations have all been in agreement that the needs outlined here merit an expansion from the current 250 staff to roughly 400. EMBL is very grateful to the Wellcome Trust and the UK Medical and Biotechnology and Biological Science Research Councils for funding a building extension to accommodate this growth. The necessary percentage increase in computational infrastructure demanded by data growth and complexity is even greater, as explained above.

An essential feature of the strategy for the EBI has to be a mechanism for taking decisions on whether to adopt a new activity or to discontinue an existing one. Until now, over the first decade of the EBI's life, this has not caused major difficulty, with most such decisions being "obvious". Major sequence and structure databases, for example, have experienced growth in deposition and usage that has rendered the need for their continuation beyond doubt, whereas databases such as those of radiation hybrids or primer sequences used for specific experimental procedures diminished in relevance and usage, and were discontinued without controversy.

However, our hitherto relatively informal mechanisms for determining the palette of databases and services will be challenged by the ever-increasing flood of diverse data. We intend to adopt a more rigorous approach in the next five years. To do this, we will invoke the input of the BAC by preparing, prior to each annual meeting, analyses of existing and proposed databases and services. For each, this will provide:

- the scientific mission;
- the resources (personnel and equipment) required and the possible sources of funding;
- the usage (existing and envisaged);
- the role of the database or service in contributing to other related resources;
- the appropriateness of the EBI (as opposed to any other institution) as the potential host.

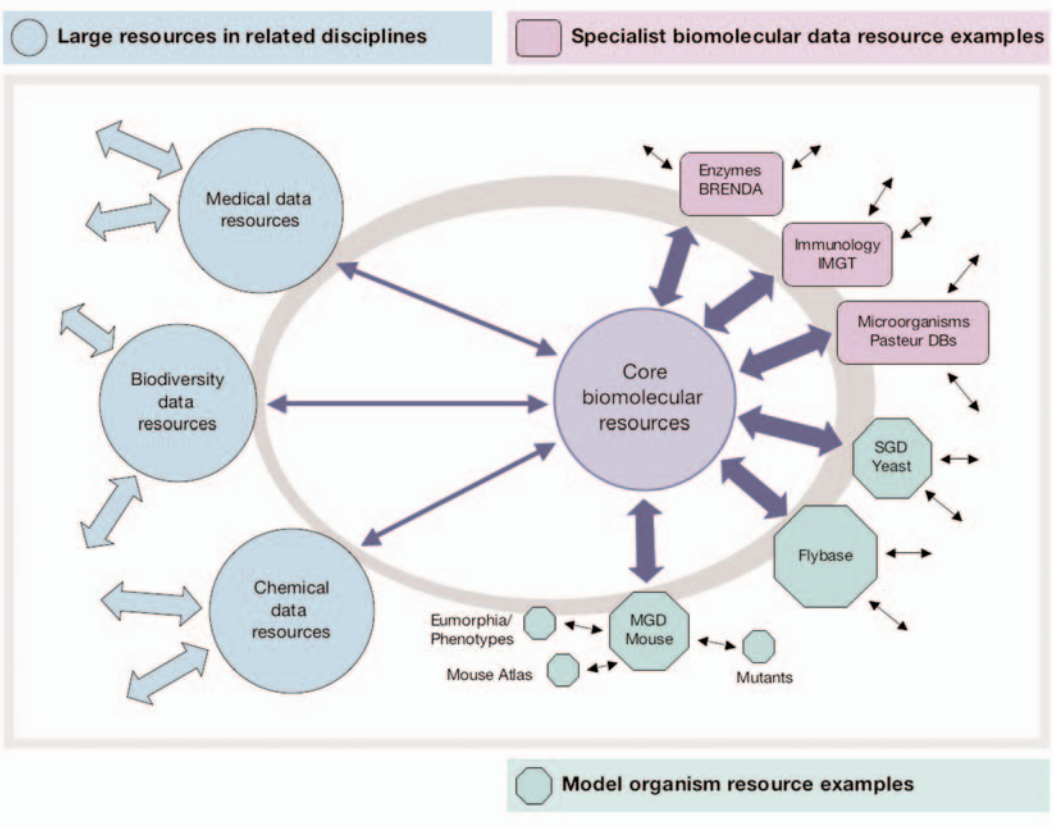
On the basis of BAC input, service activities will be prudently initiated or discontinued with a view to ensuring that the EBI maximises the comprehensiveness, relevance and quality of the services it provides within the capacity of available resources.

1.6. Integration

Optimal scientific exploitation of the EBI databases demands that users can navigate between different data collections, ideally without even noticing, irrespective of whether they are hosted at the EBI or elsewhere. Although the EBI has been pro-active in creating in excess of 25 million links between its databases, the tools it offers to users to navigate even between EBI databases are far from ideal. The creation of tools to provide seamless integrative access to all data resources is a major goal of the next EMBL Programme.

Given the impossibility of one centre holding all the data useful for life scientists, external integration is equally important. The EBI is a natural partner to many research groups and institutes, not only locally (Box 16) but also, for example, via very active participation in European and intercontinental research network activities. One can also easily discern the benefit of the core biomolecular databases being linked to a variety of other resources. Some of these are specialist organism or field-limited scientific topic databases, which typically contain much more detailed information on a specialised field than do the core databases. Others are databases of information outside of the EBI's main biomolecular focus, but directly relevant to it, such as medical records or chemistry data. These resources can be much larger than the EBI databases. Linking to them in order to ask questions about human genetic variation and health, or about interactions between biological entities and chemical entities, is an essential requirement for systems biology and therefore is clearly of high priority. A simplified diagrammatic representation of this galaxy of data resources from an EBI perspective (Figure D.5.) depicts EMBL's vision of a range of interlinked autonomous information sources configured to make science easier for everyone. To facilitate such integration, the EBI already coordinates three European networks of excellence in annotation, software integration and bioinformatics in systems biology. Other developments that will help efforts aimed at linkage and integration are the recent developments in GRID and eSCIENCE projects. The EBI is involved in both, with the goal of ensuring that simple-to-use computational tools will optimally serve Europe's life science community.

Figure D.5. Data resources from an EBI perspective



16 EMBL-EBI/Sanger Institute collaborations

The Hinxton campus has two academic institutions: the EBI and the Wellcome Trust Sanger Institute. Although each has a very distinct focus – biomolecular data for the EBI and large-scale genomic and genetic experimentation at the Sanger Institute – there is also considerable scope for collaboration in bioinformatics. The campus has the single largest community of bioinformaticians, with around 250 scientists at the EBI and 200 at the Sanger Institute. This provides a rich environment to foster sophisticated bioinformatics approaches and for the development of experimental/computational initiatives.

One of the most extensive current collaborations between the two institutes is the Ensembl database. Ensembl is a joint project between the EBI and the Sanger Institute, predominantly funded by the Wellcome Trust. Ensembl stores, manages and interprets vertebrate genomic data, including the human genome, and has become an essential part of modern molecular approaches to human biology. The collaboration between the two institutes has provided complementary skills in database development and engineering, along with detailed biological expertise for genome annotation.

As the Sanger Institute was one of the major international sites behind both human and mouse sequencing, Ensembl has benefited from rapid access to the data and to worldwide scientific expertise. Similarly, the role of the EBI in data provision has facilitated the integration into Ensembl of other important datasets and has given a strong European flavour to Ensembl. The result has been a heavily used resource and a very successful collaboration that will continue in the forseeable future.

There are other valuable collaborations between the Sanger Institute and the EBI, including (1) the long-standing integration of the Sanger's Pfam protein domain resource into the EBI-led Interpro collaboration, and (2) the collaboration between the microarray group at EBI, responsible for the ArrayExpress database, and the microarray facility at the Sanger Institute. These groups collaborate at database, analysis and biological interpretation levels. Similarly, at a structural level, the e-Family project has brought together several structural databases, including the MSD of the EBI and the Pfam group at the Sanger Institute, and has led to the development of standards important for the unified description of structural domains.

The network of both formal and informal collaborations between the two institutions has directly contributed to making the Hinxton campus an extremely exciting environment for cutting-edge bioinformatics research and development. This has benefited not only both the Hinxton institutions but also the external scientific community in Europe and worldwide. The commitment to open exchange of data and access to analysis tools at both institutes aids in the provision of biomolecular data to the broader community. Bioinformatics will be crucial to further our understanding of living systems during the next decade of molecular biology, and the collaborations between the EBI and the Sanger Institute will continue to play an important role in these developments.

1.7. The EBI and industry

We are often asked why we do not charge industry for access to EBI services. There are several reasons for this: first, we do not own the data that is submitted to the databases. It is owned by the scientists who generated it or their institutions, and it would be difficult to convince many of the owners to submit data to a repository which is used for commercial purposes. This policy also reflects global strategy. The long-term plan for research infrastructures of the Office of Science (US Department of Energ) requires open access to all data that is being generated and deposited. The Japanese government follows a similar strategy. If Europe wants to continue to be part of the global network of biomolecular databases, it has to apply the same standards and rules.

Another reason for not charging industry is that the rules of some of EBI's most important funders such as NIH, who currently provide €2.8 million per year (2004) to the EBI, require open access to the data resources funded by their grants. Finally, commercial users will only agree to support resources financially if they obtain a return for their money. Realistically, this would mean carving out certain areas within the databases and making them exclusively available to a small number of commercial clients. This would disrupt the overall integrity of our databases and, in the longer term, be very costly to implement, because of its labour-intensiveness.

One example for involving industry and providing something of value to them for which they are willing to pay is the EBI Industry Programme which was initiated when EBI was first established. We intend to continue to build on this Programme, which provides a lively forum for large companies that exploit bioinformatics and is funded by subscription income from those companies. Although these are mostly pharmaceutical companies, the membership increasingly includes companies from the agrochemical, personal care and nutrition sectors (see Box 17). We provide extensive training and information transfer, introducing our industry members from across Europe to frontier developments in bioinformatics. IT solutions developed at the EBI have been exploited in many of these companies, promoting best practice and facilitating data mining. Many technology advances, especially in high-throughput biology and combinatorial chemistry, have been led by industry, only later migrating to academia. Therefore, close interactions with industry also provide an essential view of future directions, make us responsive to the needs of such industries, and permit the EBI to develop in new directions that benefit the life sciences as a whole. The Industry Programme is financed by collecting a significant annual subscription from the member companies and developing a programme of activities in discussion with the members. Unsurprisingly, the provision of training has been their highest priority.

In addition, we have initiated an industry support forum for SMEs, which are expected to be predominantly from Europe's biotechnology sector. We anticipate this forum will develop and strengthen our links with SMEs in the same manner as our existing large industry club has done in that sector, and will promote both the uptake and use of bioinformatics services and research within this important community. This forum will be linked to training activities supported by the BioSapiens Network of Excellence, which is coordinated by the EBI.

In the computer sector, we will continue to work together with database, software and hardware providers, to gain access to new technologies and to facilitate the development of bioinformatics-friendly products through advice and collaborations. Since the establishment of the EBI, we have enjoyed excellent relationships with many such companies, gaining preferential discounts and advance notice of future developments. These interactions are important as we seek to improve our computer resources for services and research.

We will also continue to work closely with industrial colleagues to develop specific resources, consortia and research collaborations that meet their needs and are consistent with our public domain mission.

17 Current membership of the EBI Industry Programme

AstraZeneca

Bayer Healthcare

Bristol-Myers Squibb

Boehringer Ingelheim

F. Hoffmann-La Roche

GlaxoSmithKline

Johnson & Johnson Pharmaceutical Research & Development

Merck KGaA

Nestlé Research Centre

Philips Research

Procter & Gamble

Sanofi-Aventis

Schering AG

Serono S.A.

Syngenta

Unilever Research & Development

1.8. Training and outreach

The provision of any data resource is futile unless we help to ensure that biomedical researchers learn of its existence and can find out how to use it to best advantage. The demand for training in bioinformatics far exceeds the supply, and we see a clear responsibility for EBI to expand European training activities substantially in the coming five years. We will continue our present emphasis on high-level training, postgraduate and above, especially "training the trainers". We are currently running 70 training events per year for various databases and applications. For details, see Box 18.

18 Training at EBI

In 2005, EBI staff carried out a total of 70 training events, 85% of which were short, practical courses run over one, two or three days. Of these, nine were three-day training events for the EBI Industry Programme (eight events) and the SME Industry Programme (one event) in the following categories:

- Interactions and Pathways
- Macromolecular Structures Database (MSD)
- Protein Annotation
- Comparative Genomics in Ensembl
- Genome-Wide RNAi: from Model Systems to Drug Discovery
- BioMart
- Microarray Data Analysis and Normalisation
- Database Administration and Management (SME)

With the exception of the Industry Programme events (all of which were held at EBI), most events were off-site, in 15 European countries, USA and Canada.

Major training areas were Comparative Genomics (22), Functional Genomics (10) and Introductory Bioinformatics (8), with additional training in the following areas:

- Text Mining and Ontologies (6)
- Proteomics (3)
- Sequence Databases
- Structural Genomics
- EMBOSS
- Neurobiology
- Molecular Evolution

In addition to these formal training events, numerous individual lectures or conference talks in which EBI scientists introduced one or more of the data resources served as training opportunities during large conferences and workshops. Young scientists have been introduced to bioinformatics during two Open Days, one for bioinformatics Masters students (approx. attendance 70) and one for undergraduates (approx. attendance 40).

The EBI Industry Programme and SME Industry Programmes will run about 12 workshops in 2006, of which about eight will be bioinformatics training events. These will be externally funded as part of our Industry Programme. With a few exceptions, such as a roadshow run by the MSD group throughout 2003 (made possible by funding from the EU) and training in the Ensembl genome annotation system (funded by the Wellcome Trust), all other training events are funded from EBI core funds, supplemented by cost recovery charges whenever these are feasible.

The EBI's programme of short courses is clearly a success, and it must be expanded and extended to include many more academic researchers. We need (and are seeking) additional funding to run workshops at a cost that will be manageable for academic researchers.

Expanding our training provision is challenging. Thus far, it has depended almost entirely on the efforts of EBI staff, who can of course only devote a certain amount of their time to training. We coordinate such courses through a scientific training officer and will now identify funding to allow us to expand training provision and to develop new courses where needed and possible. This will include provision of online help, documentation and training guides for each of our resources, outreach at conferences through talks and software demonstrations, and the facilitation of provision of courses throughout Europe. We hope, by providing scientific vision and high-quality logistic support, to be able to draw on experts throughout Europe who will join forces with the EBI to organise and teach courses both at the EBI and locally. A further benefit of this approach will be to bring together trainers from all over Europe to allow them to exchange approaches and ideas.

As already mentioned, there is a need for sharply focused training in optimal exploitation of the EBI databases. This training has to be carried out by people who have detailed understanding of the particular resource, but such individuals are usually so embedded in production activities that it is hard to release them for training. To combat this, we will seek to raise funds for one scientist per data resource to be responsible for coordinated training and outreach, including provision of online help documentation and a training guide for the resource.

Through these measures, we hope to double the number of courses organised by the EBI each year. These courses will develop as the field of bioinformatics matures. For example, to address the move towards systems biology, we are currently seeking funds to support a course in Bioinformatics for Integrative Biology, to provide the training to access the diverse data resources whose combined use is needed for an integrative approach.

We are also keen to promote bilateral interactions with member states, whereby they send research scientists to spend an agreed period of time in the EBI (perhaps two years) before returning to a post in their own country. Such a scheme would be mutually beneficial and indeed exceptionally valuable for the partner member states.

In addition to training scientists it is our responsibility to communicate our achievements to the general public because they ultimately pay for the EBI. We need to increase our efforts to convey to everybody that bioinformatics is essential for all biomedical research and has applications in many other areas of research and development (agriculture, food etc), that are of direct relevance and benefit to the people of Europe.

2. Structural biology facilities

EMBL provides high levels of support to the structural biology communities in the member states through its Grenoble and Hamburg Outstations. These were established adjacent to world-leading large-scale facilities, the ESRF and the ILL in Grenoble, and the German Electron Sychrotron Research Centre (DESY) in Hamburg. Acting either closely together with ESRF/ILL in Grenoble, or by the independent provision of beamlines with applications in life sciences in Hamburg, the two EMBL Units have taken a major responsibility to provide state-of-the-art facilities for a large number of structure determination projects. Annex 12 shows a reflection of the use made of these sites by documenting how many depositions into the protein database (PDB) arose from structural analysis at the Hamburg and Grenoble synchrotron sites. The numbers demonstrate the major impact on the field resulting from the user support provided for data collection in Grenoble and Hamburg.

Both structural biology Outstations are of proven utility and complementarity. In terms of recent developments, Hamburg has become a leader in providing software that allows automatic interpretation of data collected (for instance, ARP/wARP, ATSAS) and has stimulated the renaissance of biological small-angle X-ray scattering (SAXS). Owing to the more recent growth of the ESRF beamlines, Grenoble currently has leading expertise in beamline construction and automation for high-intensity undulator beamlines. This expertise will be shared as much as possible with Hamburg to benefit the planned PETRA beamline projects (see Section D.2.2.). The Heads of the Hamburg and Grenoble Outstations will be jointly responsible, and report directly to the Director General, for developing a strategic plan for the provision of synchrotron-related services to the user community.

In accordance with current demands in structural biology, both Outstations are diversifying their facilities to include high-throughput expression, purification and crystallisation platforms, as well as implementing laboratory information management systems (LIMS) for sample tracking. Close coordination between the two Outstations and with the EBI, a central player in the development of standard protocols, databases and bioinformatics for structural biology, is already guiding these developments, aided by the involvement of all three EMBL Units in the pan-European integrated projects SPINE and BIOXHIT.

2.1. EMBL Grenoble

In a very close collaboration with ESRF, organised through the Joint Structural Biology Group (JSBG), EMBL has been and continues to be involved in the design, construction, maintenance, development and operation of seven macromolecular crystallography (MX) beamlines. In addition, the CRG beamline BM14 is run jointly by the UK, through the Medical Research Council, and EMBL. Four of these eight beamlines were initially built and commissioned under the leadership of EMBL scientists, and four current EMBL scientists each have responsibility, either alone or with an ESRF colleague, for one of the beamlines that is in routine operation.

Modern MX beamlines need to be designed for efficient operation to maximise the utility of high-intensity beams, and this is only possible with automation. In recent years, EMBL scientists have designed a suite of automated sample-handling instrumentation for third-generation synchrotron beamlines, including micro- and mini-diffractometers with novel on-axis visualisation and an automatic frozen crystal sample changer coupled with automatic sample centering for positioning the crystal in the beam. In addition, a prototype mini-kappa goniometer for use with single- or multi-wavelength anomalous dispersion (SAD/MAD) experiments has been developed in EMBL. In a long-term programme of automation involving EMBL and ESRF, all the public MX beamlines are being equipped with these devices together with the control software, databases and LIMS necessary to interface seamlessly with users upstream (high-throughput protein expression and crystallisation) and downstream (data processing, structure determination).

Other projects that will continue to be a focus include design and operation of the new microfocus beamline ID23-2, manipulation of microcrystals, and development of methods both to reduce and to make productive use of the radiation damage caused by the extremely intense X-ray beams. EMBL Grenoble will also closely follow the upgrades in the ESRF ring (higher current, higher brilliance and the use of canted undulators to provide more beamlines) and be involved in opportunities that will arise for structural biology. These upgrades are currently under discussion but no detailed information is available at the time of writing.

On a more day-to-day level, EMBL Grenoble staff participate in the JSBG beamline maintenance team; junior staff act as hosts to visiting biologists and provide instruction on how to utilise the beamlines and, if required, access to biochemical facilities. At Grenoble during 2004, there were 1821 users on public MX beamlines at the ESRF, 258 users on the CRG beamline BM14 and 30 users of the ILL neutron instruments DB21 and LADI (see Annex 13).

With respect to the neutron beamlines at ILL, the Grenoble site is the only one in Europe where neutron diffraction experiments, ideal for visualising biologically important hydrogen atoms and water molecules, can be efficiently carried out on crystals of biological macromolecules. The world-leading LADI beamline was designed and built under the supervision of an EMBL scientist. Maximal exploitation of this technique requires use of per-deuterated proteins and expertise in growing large crystals. In the next funding period, EMBL will continue to support the per-deuteration facility (D-lab) established together with the ILL but, on the advice of the 2005 SAC review panel of the Grenoble Outstation, will reduce resources put into neutron beamline design, construction or maintenance, or into basic research using neutron crystallography. In the context of the Grenoble Partnership for Structural Biology (PSB) (see Box 19), the review panel felt that these aspects would be carried out more appropriately by ILL.

19 The Partnership for Structural Biology

The Partnership for Structural Biology (PSB) was formally established under an agreement that was signed in November 2002 by the Director Generals of EMBL, ESRF, ILL and the Institut de Biologie Structurale (IBS) (represented by the Commissariat à l'Energie Atomique (CEA), the Centre National de la Récherche Scientifique (CNRS) and Grenoble Université Joseph Fourier). The partners agreed to collaborate to promote multidisciplinary structural biology, especially that related to human health. The Institut de Virologie Moléculaire et Structurale (IVMS) is not formally part of the PSB but will share a new jointly funded building that will house some PSB activities and the IVMS. The 3,600m^2 Carl-Ivar Brändén Building, which is connected to the EMBL Grenoble building, was completed in October 2005. It will be the physical focus of the new integrated activities and will house several of the new technology platforms to promote state-of-the-art structural biology:

- high-throughput expression cloning and protein expression facilities (EMBL, IBS);
- protein characterisation and quality control facilities including nuclear magnetic resonance (NMR), mass spectrometry and isothermal titration calorimetry (ITC) (IBS);
- a high-throughput nanovolume crystallisation facility (EMBL, IBS);
- a new state-of-the-art, automated, two-endstation beamline for macromolecular crystallography at the ESRF (ESRF, EMBL);
- a biological isotope/deuteration laboratory for neutron scattering and NMR (ILL, EMBL, IBS);
- an upgraded diffractometer for high-resolution neutron protein crystallography (ILL);
- a comprehensive electron microscopy platform for negative staining, cryo-electron microscopy and image reconstruction (IVMS, IBS).

The PSB has already stimulated inter-institutional cooperation and scientific collaboration to an unprecedented extent, as exemplified at regular six-monthly PSB Science Days. It also represents a unified face to the exterior, making it a powerful force on the European scene, as witnessed by involvement in several large EU-funded projects (SPINE, BIOXHIT, neutron I3, synchrotron I3, 3D-Repertoire). The PSB-IVMS project has itself received funding from the sixth Framework Programme and from French authorities.

These developments will benefit local and European scientists in several ways. The partner institutes and IVMS will benefit from enhanced critical mass and access to state-of-the-art multidisciplinary platforms. The new shared infrastructures and increased integration will further enhance the site as a centre of excellence in structural biology, as well as a leading training centre for European scientists in multidisciplinary structural biology. Strong contacts with biotechnology and pharmaceutical companies will also be developed.

2.2. EMBL Hamburg

The Hamburg Outstation has more than 30 years of experience in providing synchrotron beamline facilities to the international scientific community. DESY generously provides the core infrastructures that are presently located at the DORIS-III storage ring, but will soon migrate to the PETRA-III storage ring as dedicated devices. EMBL Hamburg operates five beamlines in protein crystallography, one biological SAXS beamline, and one beamline for applications in X-ray absorption spectroscopy (EXAFS). In addition, one of the largest high-throughput crystallisation facilities in the world is under construction and will be available to external research groups in 2006. Staff from EMBL Hamburg provide a complete service package to users, including the administration of visits, scientific supervision during experiments, and the provision of infrastructure for increasingly automated processing and interpretation of data. At present, the EMBL Hamburg facilities are used by over 400 scientists per year (see Annex 14). The available beamtime is distributed by a selection committee composed of 11 leading scientists. It is presently chaired by Prof. Dino Moras (Illkirch, France).

Since the DORIS-III storage ring is no longer state-of-the-art, DESY has decided to convert the 2.3 km PETRA ring into a dedicated synchrotron source and has invited EMBL to organise, build, maintain and provide access to beamline facilities for structural biologists. The optical parameters of PETRA-III in terms of emittance, beam divergence and energy will make this source at least equal and possibly superior to other leading synchrotron facilities.

After consultation with the 2003 SAC review panel, which provided enthusiastic support for the plan, EMBL developed an integrated proposal for structural biology activities at PETRA-III that is described below. This proposal has been exposed to scrutiny by two expert panels, one that oversees the entire development plan at the DESY site and a second that was assembled specifically for the purpose of examining the EMBL plan (see Annex 16). In addition, EMBL's SAC and Council have been informed and have discussed these plans. In spite of the background situation with new national facilities in some EMBL member states, these panels all judged that the EMBL Hamburg Outstation will continue to be required in order to meet the ever-growing demand for access to synchrotron radiation sources from throughout EMBL's member states and beyond.

On the basis of this advice, we will proceed with our part of the PETRA-III plans, provided that funding, the bulk of which is likely to be provided by the host country, Germany, is forthcoming. At present, all the signs are positive and we expect a final funding decision before the start of the next EMBL Programme.

2.2.1. The structural biology facilities at PETRA-III

EMBL Hamburg is proposing Integrated Structural Biology Facilities at PETRA-III. The proposal is driven by the unique opportunities offered by PETRA-III to build synchrotron radiation beamlines to serve the international scientific community for the most demanding applications in life sciences. The proposal reflects the needs expressed by a large number of research groups from across Europe (based on a survey of 159 external research groups active in structural biology; see Annex 15) and was reviewed very positively by an external advisory council (see Annex 16). This survey documented clear demand for additional opportunities at state-of-the-art beamlines and for their integration with joint sample preparation and online data-processing facilities. The proposed facilities will include:

- Two MX beamlines; the first will be tuned for microfocusing to allow data acquisition from extremely small crystals of biological macromolecules; the second will be tuned for applications over a large energy range to allow data acquisition at the absorption edges of multiple different elements to allow experimental phase determination, and specific applications such as structures at ultra-high resolution, which require specific energy regimes. In addition, EMBL has also offered

assistance in the coordination of planning, construction and operation of a third MX beamline whose focus will be on high-throughput applications, to be funded by other research organisations.

- One BioSAXS beamline for large-scale shape and quaternary structure analysis of individual macromolecules and functional complexes, and for cutting-edge applications such as ultra-fast kinetic studies.
- A joint sample preparation and data-processing area, allowing provision of a complete pipeline for structural biology experiments using synchrotron radiation. A high-throughput crystallisation facility (currently under construction, externally funded) will be integrated into this area. Although such a facility will clearly have a limited capacity, our plan is to make as much time as possible available to external users.

The Integrated Structural Biology Facilities will be located at the last two straight sections on PETRA-III. The close proximity of the proposed components will allow us to establish direct pipelines ranging from the preparation and characterisation of samples, their transfer to beamlines, X-ray data acquisition and online data processing and interpretation, with options for remote experiment monitoring and operation. All endstations will be equipped with state-of-the-art instruments to provide a user-friendly and highly automated experimental environment, permitting a high throughput of experiments at the future PETRA-III beamlines.

EMBL Hamburg's proposed Integrated Structural Biology Facility at PETRA-III is intended to be open to other research organisations. We are aware of interest from the Helmholtz Society and other scientific organisations for future research activities on the DESY campus in Hamburg, and their interest in carrying them out in collaboration with EMBL. Such inter-institutional partnering could make Hamburg an internationally competitive centre in structural biology research as well as service.

Additional developments at DESY include the construction of two novel photon sources, the far-UV free electron laser that is under construction and the X-ray free electron laser that will be operational in 2010. Although scepticism exists, because of the inevitability of radiation damage in such intense beamlines, it is worth recalling that many structural biologists were also sceptical about the practicability of using synchrotron radiation sources 30 years ago when the Outstation was founded. It was the work of Ken Holmes and his group in Hamburg at that time that provided the proof of utility. Structural biology has not been the same since. The FEL facilities will provide unprecedented opportunities for research in physics, chemistry and life sciences. Because of their time structure and specific optical properties, particularly coherence, they could allow the study of dynamic processes at up to femto-second timescales or permit structure determinations of single particles, overcoming present requirements for crystallisation. Although the design and outcome of many experiments at these new instruments is still largely unknown, the EMBL Hamburg Outstation is keen to participate in the development and implementation of future experiments, which will evaluate if and how these new sources can be used with biological materials.

3. Visitors Programme

The EMBL Visitors Programme enables scientists and undergraduate students from all over the world to spend time at all five EMBL sites. Each year over 3,200 visitors come to EMBL to work on collaborative projects, conduct their own research, receive training in a specific area, participate in long-term collaborations or take advantage of the Laboratory's cutting-edge facilities. The vast majority of these visitors are from EMBL's member states. A detailed breakdown of the 2004 visitors is provided at Annex 17.

Visitors fall into four different categories:

- **Scientific collaborators** – scientists who are integrated into a research group for the duration of their visit in order to perform joint experiments or discuss collaborative projects with individual members of the EMBL faculty. They might receive training in the process of this research. New collaborations are frequently initiated with additional EMBL research groups as a result of such visits.

- **Sabbatical visitors** – senior scientists who wish to be associated with a specific group or Unit, or the Laboratory as a whole, for a period of study, reflection, writing and research.

- **Facility users** – scientists who use the structural biology facilities in Hamburg and Grenoble, or receive training in specialised facilities of the Laboratory in the other Units, including the Core Facilities and the EBI data resources. They often evaluate technology for their own institution, or perform a specific project that requires specialised equipment that is available at EMBL.

- **External PhD students, diploma students and trainees** – students who are registered at external institutions and who come to the EMBL for a part of their education. External PhD Students must carry out less than half of their PhD work at the EMBL; Diploma Students do practical work at the EMBL as stipulated by their home institution. Trainees are undergraduate students that spend a period of training at EMBL during which they benefit from the experience of working in a dynamic, international research environment.

The Visitors Programme will continue to provide services to scientists who are interested in visiting EMBL for any of the purposes described above and will be closely associated with the EMBL International Centre for Advanced Training (EICAT). The construction of a training building will provide new facilities that can be made available to visitors, and will include dedicated laboratory and office space. This represents a major improvement on the current situation whereby visitors have to be hosted in the overcrowded labs of research groups.

We also welcome a significant number of non-scientists to EMBL each year. These individuals are not registered as scientific visitors, but are nevertheless important for the Laboratory. Such visitors include government officials from EMBL member states and other countries, as well as representatives from research institutes who often visit EMBL in order to compare our organisation and structures with their own. We also host several groups of visitors with different interests, ranging from teachers and school children to students and pensioners. EMBL Heidelberg and the Outstations have hosted Open Days for the general public. These opportunities to visit EMBL have all been heavily oversubscribed and will be repeated in the future. We will organise specific events and visits depending on the target audiences. For example, for school children, this could be simple hands-on activities to let them experience "lab work" for the first time; by contrast, a group of students visiting EMBL would be presented with the research activities at EMBL, meet scientists and pre- and postdoctoral fellows, and visit the scientific facilities. The Visitors Programme and the Office of Information and Public Affairs (OIPA) will continue to organise these visits and events.

4. Technology Transfer

4.1. Current status

Since it became fully operational in late 2000, the development of EMBLEM, the wholly owned technology transfer subsidiary of EMBL, has been a tremendous success and technology transfer has become an integral component of EMBL.

EMBLEM's "Innovation Works" is a technology transfer model that combines a pro-active IP-sourcing strategy that maintains the free dissemination of knowledge within the basic research environment of EMBL with professional technology assessment, protection, development and commercialisation tools, as well as education and teaching initiatives. Added value is created by bundling technologies with scientific consultancy services and collaborations. Creation of incubator/accelerator space close to the main Laboratory, as well as access to the associated venture capital seed fund (EMBL Technology Fund) provides potential founders of start-up companies with all the necessary technical and financial tools and support to develop and deploy their ideas rapidly.

The number of inventions, patents and copyrights in the IP portfolio, as well as the annual turnover and number of concluded licence and collaboration agreements, grows steadily year by year. Licensing income has quadrupled since 2000 and comes in the form of a mixture of fixed cash revenue (milestone or upfront payments), equity stakes and royalty, to ensure that the technology transfer activities have long-term sustainability. Five years after its inception, EMBLEM is already at the point of breaking even, which is a remarkable achievement given the state of the markets over EMBLEM's lifetime. It is all the more praiseworthy since technology transfer activities usually need at least 8–12 years to break even, and many never do. This success supports EMBL's thesis that support for basic research is an effective way of ensuring continued innovation.

Owing to the efforts over the last five years, industrial and commercial partners recognise EMBLEM as an established and respected name in technology transfer in the member states and beyond, and EMBLEM is considered the benchmark for technology transfer from an international academic setting.

4.2. Future developments

The high-quality innovations from the Laboratory provide a healthy IP portfolio for licensing to established companies or, in some cases, to EMBL start-up companies, of which there are currently eight. EMBLEM manages an IP right(s) (IPR) portfolio of over 200 granted patents and patent applications, more than 50 copyrights and trademarks, and a licensing portfolio in excess of 250 active licence and consultancy contracts, with more than 160 licensees worldwide (see Annex 18). The technology transfer activities of EMBLEM thus have a solid basis to grow from.

Over the next five-year period, EMBLEM will continue with its policy of "sustainable growth". The EMBL IPR is expected to grow by 50–75%, approximately 300 new licensing agreements will be concluded, and the first significant royalty returns from licensed technologies are envisaged in this period.

EMBLEM has already extended its services to researchers who are not based at the Laboratory, for example to EMBL alumni across Europe, and has entered into limited partnerships with other EIROforum members such as the European Space Agency (ESA). These services and collaborations will be intensified, and new strategic partnerships will be established with EIROforum members as well as with other academic research institutes and entities.

The overall aim of the activities in the forthcoming five years is threefold: first, to provide vital services in the life sciences to EMBL and member states; second, to export the success of EMBLEM's "Innovation Works" technology transfer model to a broad European base; and, third, to speed-up the development of basic research innovations into marketable products with medical relevance, in order to bring benefit to society at large in our member states.

E. TRAINING

One of the core missions of EMBL from its inception has been the provision of advanced training. In fulfilling this mission over the past three decades, the Laboratory has become a beacon, attracting biologists at all stages of their career. Biologists from Europe and further afield are attracted to courses, conferences, practical workshops, sabbaticals, and postdoctoral or graduate studies. The Laboratory is a hub where scientists from all branches of the biological sciences converge to gain new experience, teach and exchange ideas. The high profile of the Laboratory as a top-class research institute, combined with the advanced training activities taking place at the five EMBL sites, have provided inspiration for the establishment of several research centres throughout the world. For example, EMBL, with its strength in basic research and its scientific culture, has been extensively used as a model in the planning of Janelia Farm, the Howard Hughes Medical Institute's new campus. Consistent with the EMBL philosophy of interdisciplinarity, renewal and change, the advanced training activities offered by the Laboratory continuously evolve with new scientific research and technological development.

Reflecting its commitment to advanced training in the life sciences, EMBL is in the process of establishing the EMBL International Centre for Advanced Training (EICAT). EICAT will coordinate all existing EMBL advanced training activities, complement them with those that are missing and create new ones. Although advanced training is conducted at all five EMBL sites, the main laboratory in Heidelberg plays an especially active role. EICAT will orchestrate the diverse training activities in all EMBL sites and Units, facilitating and catalysing the full participation of the Outstations in advanced training activities, without compromising their ability to act locally. By integrating the experience and know-how of the different EMBL sites, EICAT will create synergies and increase the cost effectiveness of the training the Laboratory offers to scientists working at or visiting EMBL.

With the increasing pace and interdisciplinarity of the life sciences, the number of EMBL training activities has expanded to meet the demand from scientists all over Europe for first-rate courses, conferences and workshops. EMBL's sister institution EMBO, which shares the European scale of mission with EMBL, is a frequent sponsor of these meetings. Scientists who wish to organise new courses that bring the newest ideas and methods to the attention of others, but do not have the means to do so, regularly appeal for assistance to EMBL as the main site in Europe for advanced training in the biological sciences. This welcome development has challenged EMBL's capacities. The existing facilities, from lecture halls to training laboratories, are insufficient. To face this challenge, and to meet and uphold its commitment to continue to develop and provide what is considered to be among the world's best training in the biological sciences, EMBL Council approved the construction of a Multipurpose Building in Heidelberg that, among other things, has space for laboratory courses and for visiting scientists. The host country, Germany, generously funded this initiative. However, this building does not by itself solve all of the problems. The current auditorium, the Operon, does not have sufficient capacity to allow us to accept all the applicants who wish to attend conferences at EMBL. Furthermore, poster sessions are an essential part of scientific meetings and EMBL lacks adequate poster presentation facilities. In addition, EMBL's current catering facilities are severely stretched when conference participants are added to the EMBL staff as customers. The Laboratory has been offered a generous contribution towards building a state-of-the-art conference and training facility, the ATC, which could house all of the EICAT activities (see below). The ATC would provide the space and environment necessary for EMBL/EICAT to continue providing and enabling the development of new training activities in the future.

EICAT will promote a unified external visibility of EMBL advanced training, intensify strategic activities with external partners (e.g. the EMBL partner universities) and establish new strategic collaborations to promote top-level advanced training in Europe. EICAT will also serve as an active interface with EMBO, helping the two organisations to leverage the mutual benefit of their complementary capacities and functions as well as their shared interests, experience and commitment to training in the life sciences in Europe.

EICAT, founded on the success and experience of the diverse training programmes it will oversee, will coordinate activities within a cooperative structure that gives support to training activities. The EMBL training programmes that EICAT will coordinate include:

- EIPP, the EMBL International PhD Programme;
- the new Postdoctoral Programme;
- vocational training for group and team leaders (together with the EMBO Young Investigator Programme and EIROforum), as well as for technical and administrative staff in collaboration with the Personnel Section;
- ELLS, an education facility for high school teachers, whose mission is to bridge the gap between research and classrooms;
- the Collaborative Training Programme;
- courses, conferences and workshops;
- the Scholars' Programme and the Visitors Programme, which offer principal investigators, as well as postdoctoral fellows and graduate students from other institutions, the opportunity to associate with a specific group or Unit of the Laboratory for a period of study, reflection, writing, collaborative research, etc.

EICAT will thus pursue a dual mission: to provide first-rate training for scientists working at EMBL, and to serve as a European hub of advanced training for those who primarily work elsewhere. It will aim to promote external visibility of EMBL advanced training to member states' scientists, actively engage in fundraising, and work towards creation of a state-of-the-art conference and training facility.

1. EMBL International PhD Programme

The EIPP was founded in 1983 and is today a centrepiece of advanced training activities at EMBL. It is a world-renowned programme that attracts outstanding students from all of our member states, elsewhere in Europe and, to a more limited extent, the rest of the world. Owing to the success and outstanding quality of the programme, EMBL was granted the right to award its own PhD in 1997. Collaborative agreements to establish joint PhD programmes have been signed with 24 universities in 17 countries and we are close to reaching our goal of including at least one partner university from each of the EMBL member states. Currently, 180 students from 30 countries are enrolled in the EIPP, with representation from all EMBL member states. See Annex 19 for a complete list of partner universities.

The quality of the pool of applicants to the EIPP is outstanding and the admittance rate is around 10%. Once accepted, students are trained for up to four years to complete their PhD. All students are guided by a thesis advisory committee, which usually consists of three EMBL group leaders and one university professor. Students can defend their thesis and obtain their PhD from EMBL, or jointly from EMBL and one of its partner universities, or from national institutions (EMBL has not exercised the right to award the PhD degree by itself. It has used this right to establish the "Joint PhD" options – see below). About 45% of students are funded by EMBL, 7% are funded by the EU, 5% receive fellowships from the Louis Jeantet Foundation, 2% from the Darwin Trust and 43% from other external organisations (these figures are for 2004).

The EIPP is hallmarked by a structured training and mentoring system that includes a compulsory core course in molecular biology, a bioinformatics training course and a thesis advisory committee for each student. The EIPP strives to foster student independence in the context of prudent guidance. The

collaborations with top-level universities have revolutionised the granting of PhD degrees, by providing a well-functioning mechanism for Joint PhD degrees that reflect the contributions of both EMBL and the national universities. This system enables EMBL PhD students to maintain contact with a university in their home country, which can be very valuable in their future career. Joint PhD degrees based on EMBL's statutes are also a rare example of converging academic standards for PhD training in the life sciences in Europe.

The past few years have witnessed the emergence of international PhD programmes in numerous locations in Europe. This is a very positive development. It is gratifying to see that the example set by EMBL has been used as a model, often with an open acknowledgement. At the same time, these developments challenge the EIPP to maintain its attractiveness, and to continue to set the highest standards and to serve as an example, advisor and resource.

The following aspects represent priorities of the EIPP within the next funding cycle.

- Serving the needs of an evolving EMBL.
    - PhD students are possibly the most crucial human resource to newly recruited group leaders, who find it more difficult to attract postdoctoral fellows until their research programmes are better established; PhD students are also very important contributors within more-established groups. Traditionally, the demand for PhD students has exceeded EMBL's ability to offer fellowships, although the number of highly qualified applicants has not been limiting. However, the number of fellowships that EMBL can fund and the number of students who can be trained within the available central training facilities (teaching lab, seminar rooms, etc.) are limiting factors. This situation will be addressed in the next EMBL Programme by increased external fundraising for the EIPP, as well as by upgrading the teaching facilities.
    - To support EMBL's move towards systems biology, the EIPP will attempt to further enhance its interdisciplinarity, especially in the areas of physics and mathematics for modelling and simulation. Targeted recruitment efforts will address this need.
    - The areas of "biology" and "medicine" are also converging. Within its efforts to contribute to this exciting frontier, the EIPP will develop, together with its partner universities, the establishment of MD/PhD degrees.
- Maintaining a leadership position among European PhD programmes.
    - Although the EIPP is seen and used as a model by many, we need to safeguard the attractiveness of the programme and its quality by constant improvements. This includes the complete representation of all EMBL member states among our partner universities. This will enable us to offer Joint PhD degrees within all of our member countries. An enhancement of "additional skill training" (e.g. in mathematics/statistics, bioinformatics, presentation and writing skills) will help to attract the most highly qualified candidates (including those from Eastern Europe and Asia) to apply to the EIPP.
- Sharing expertise and resources with others in Europe and building a European network based on our partner university network. We also want to expand our usefulness to outside institutions, particularly from the member states. Within the context of the EIPP and EICAT, the network of existing and future partner universities can be seen as a resource to both the EMBL and the European training landscape. The following activities will be developed:
    - joint training networks (including joint applications for external funding);
    - improved access to the outstanding EIPP applicant pool, because EIPP attracts a greater number of highly qualified applicants than it can train.

2. Postdoctoral Programme

There are currently as many postdoctoral fellows at EMBL as PhD students. EMBL provides an exciting environment for postdoctoral fellows, with outstanding research laboratories and facilities, high-quality seminar programmes and a vibrant international atmosphere, all of which ensure that postdoctoral fellows have access to an optimal scientific environment at this critical career stage. The diversity of biological research being conducted at EMBL provides opportunities to pursue interdisciplinary research and to acquire a broad perspective on biological problems and the technological approaches that can be used to address them. EMBL postdoctoral fellows are likely to become independent scientists in EMBL member states. It is therefore important that EMBL continues to be able to attract the best candidates to these positions and to provide an environment that can compete with research institutions in the USA, where many of the best European postdoctoral fellows still prefer to go.

We have recognised the special needs and requirements for training and career development support, and we are in the process of developing a Postdoctoral Programme at EMBL. Most of the training that postdoctoral fellows receive at EMBL is within the research group that they have joined, but additional centrally organised training will be provided in the future. This will include transferable skills, such as presentation and writing skills, which will help in the transition to a more senior position. Mentoring is also an important aspect for these young scientists, and we are planning to promote networks between EMBL postdoctoral fellows and EMBL alumni to help the fellows during their transition from EMBL to research institutes in member states.

Recently, a postdoctoral association was formed as an informal association of all EMBL postdoctoral fellows in a bottom-up initiative. The EMBL postdoctoral association strives to enrich the experience of its members at EMBL, as well as to enhance the contributions of postdoctoral fellows within the context of EMBL. With backing from the scientific staff and the administration, EMBL wants to promote this association and encourage the postdoctoral fellows to self-organise activities that will further enhance the value to them of their time at EMBL. Specifically, the association aims to create a network of scientific and social interactions among postdoctoral fellows across EMBL Units, to facilitate communication with the EMBL faculty and administration, to help provide information on funding and career development, and to organise courses that will enable the postdoctoral fellows to share their specialist scientific expertise (e.g. statistics, mathematical modelling, microscopy techniques) with their colleagues.

The Postdoctoral Programme will be organised by a senior scientist, with the active participation of the EMBL postdoctoral association. Most EMBL postdoctoral fellows are funded by external grants and only a modest increase in resources for organising retreats, training activities and networking events will be required to achieve these goals.

3. Vocational training

EMBL considers vocational training to be an important part of its role as an international employer, particularly since the time-limited employment policy requires that staff are trained not only to perform their work at EMBL, but also to be attractive candidates when the time comes to leave EMBL and take up a new position. Vocational training is already provided to EMBL staff; however, vocational training opportunities have so far been limited, and have not been organised in a uniform, easily accessible way throughout the Laboratory. Demand for staff training has been rising, and the complexity of EMBL with its

five different sites requires that vocational training activities have to be both well structured and offered to EMBL staff in a consistent high-quality programme.

Skills that EMBL staff need to acquire include management and leadership, and technical and software/programing, as well as "soft" skills such as communication and presentation. We expect that some training can be offered by internal staff but that other types of training will require external experts. We are currently evaluating whether it will be possible to organise staff training activities together with our EIROforum partners, some of whom (e.g. CERN) already have well-established internal training programmes that they would open to participants from smaller EIROforum organisations such as EMBL.

By establishing a structured training programme, we will be able to ensure that all members of personnel will further strengthen their knowledge and develop the additional skills required to move to the next step in their careers after their employment at EMBL. Since our staff members generally leave EMBL to take up positions in our member states, this training will ultimately be to their benefit. Expanding our training activities will require additional resources for personnel, budgets for training, and time to teach and participate at training courses.

4. The European Learning Laboratory for the Life Sciences

Biology is an integral part of life science curricula throughout Europe. Although many biology teachers are aware that a revolution is occurring in life science, most do not have a clear idea of its true scope or of its potential to have dramatic effects on society during our lifetimes. Most educational systems have not developed effective mechanisms for keeping teachers up-to-date with a science moving at the pace of modern biology. In 2003, EMBL launched a dedicated education facility, ELLS, within the framework of an EU-funded project coordinated by EMBO. The goal of ELLS is to provide professional development for biology teachers in secondary schools, a group recognised as crucial to many aspects of future European success in the area of research and development.

We are committed to continue the activities of ELLS beyond the end of this project in close collaboration with EMBO. ELLS will continue to organise and run regular courses for teachers in Heidelberg and elsewhere, and will support EMBO in the organisation of an annual teachers' workshop. Both organisations will apply for grants to fund these and new education and science communication activities in the future.

ELLS training activities for teachers are also a useful way to provide science communication training to EMBL scientists, by involving them in the development of teaching materials and organising courses for teachers and other events for school students, such as summer schools. The development of teaching material and its presentation during workshops will be at the core of the communication training. The courses will allow scientists to interact with non-specialists and to test and revise the material they have developed. Communication tools can include written documents, presentations, web-based modules and other means. ELLS will also provide communication training in the context of the vocational training programme at EMBL.

ELLS is part of EMBL's outreach activities that also include the active Science and Society Programme as well as special events organised by the Office of Information and Public Affairs (OIPA) (see Section F) and a new journal for science teachers in Europe, *Science in School*, which will be published by EMBL in the context of an EU-funded EIROforum project (NUCLEUS) starting in 2006.

5. Collaborative training

EMBL established the EIPP in 1983 and has had the right to award its own PhD degree since 1997. The EIPP is regarded as a model throughout Europe and the world. This success and our experience in the area of advanced training, as well as EMBL's internationality and its network of partners, can benefit other high-quality European efforts for advanced training.

Programmes that support and benefit from EMBL's philosophy are considered on a case-by-case basis for inclusion as a "Collaborative Training Programme". They are in general modelled on the successful EIPP, but are operated independently from EMBL. EU-funded networks of excellence and integrated projects often establish PhD training programmes, and EMBL can help these programmes by advice and example.

6. Courses, conferences and workshops

EMBL-organised courses, conferences and workshops are recognised internationally as venues for training, discussions and the exchange of ideas in all fields related to the life sciences (see Annex 7). This is an excellent example of the complementarity of EMBL and EMBO. EMBO evaluates applications from all over Europe from those who wish to organise scientific meetings of all sorts. They fund a wide range of activities by selection from these applications. EMBL scientists are very active applicants to these programmes, and are regularly successful in obtaining funding. In the past four years EMBL has organised over 100 courses, conferences and workshops, with over 9,000 participants from more than 70 nations. Demand for participation has been consistently high and, owing to the limited capacity of the facilities currently available, practical courses have been oversubscribed by as much as five times the number of places available. Indeed, participation at most events is necessarily restricted because of the lack of space in the auditorium and teaching laboratory. Better serving the demands and needs of the user community of scientists from our member states will have to be postponed until appropriate facilities and staff become available.

7. The Advanced Training Centre: an opportunity for European science

Advanced training has always been an important deliverable of EMBL's mission. The excellent quality of EMBL's training opportunities has been achieved by attention to the needs of our scientific colleagues together with feedback-driven continuous improvements. EMBL will remain committed to this responsive mode of offering a steadily improving portfolio of advanced training opportunities in the future. With the proposal of the ATC, we have identified a unique opportunity for EMBL and Europe to take an ambitious step forward.

Training young generations of scientists and providing meeting points for top-level scientific exchange are key defining factors for the development of competitive, knowledge-based societies. The existence of several world-class conference programmes and sites has established the USA as the centre for scientific exchange, with the Cold Spring Harbor Laboratory (CSHL) Courses and Conferences Program being recognised as the world leader in the life sciences. The regular provision of courses, conferences and practical workshops via CSHL, Keystone or the Gordon Conferences in the US represents a powerful advantage for young US researchers in comparison with their European colleagues, who have to travel across the Atlantic to take advantage of these programmes. Europe itself currently lacks a training centre of comparable scientific breadth and quality to CSHL. Similar to CSHL, EMBL Heidelberg is a world-class laboratory situated in a campus setting and in close proximity to a hub of global air traffic. Similar to CSHL, EMBL and EMBO have

(often jointly) offered top-quality conferences and practical workshops. In contrast to CSHL, EMBL's activity has been restricted by its limited facilities, in terms of auditorium space, poster exhibition space and teaching laboratory space. The realisation of the "ATC project" would provide Europe with a world-class training centre and thereby address a pressing need felt by many European scientists. The financial feasibility of the ATC project is greatly aided by an offer of a generous contribution by a German foundation.

The main purpose of the ATC will be to provide the infrastructure for a regular, world-class programme of courses, conferences and practical workshops. International conferences covering the spectrum of the modern life sciences would be organised during at least 15–20 weeks of the year by EMBL scientists, EMBL alumni, EMBO members and other qualified scientists. In addition, practical workshops would be offered throughout most of the year.

If approved, the ATC will be built on the EMBL/EMBO campus in Heidelberg. It would replace the plans for the already approved Multipurpose Building that was discussed in the introduction to Section E, and would incorporate the components that were planned for the Multipurpose Building. The ATC would however also include a state-of-the-art auditorium with a capacity of 500 seats, purpose-designed poster exhibition areas for up to 300 posters, and teaching laboratories for up to 60 participants. In addition, the ATC would offer smaller seminar rooms, office space for EICAT, and would provide offices for visiting scientists. Finally, it would include a catering and communication area for conference attendees. Suitably located and priced accommodation for attendees is available in Heidelberg and includes the EMBL guesthouse and the ISG Hotel, which have enough capacity to serve for smaller meetings.

Scientists from EMBL's member states often express their need for a world-class scientific conference site in Europe. This need exists both for opportunities to attend conferences and for the possibility to organise conferences at a suitable venue with the support of professional staff. The ATC could ideally meet these demands, provide a unique service to our member states, and foster strong synergies with our sister organisation EMBO.

F. EXTERNAL INTEGRATION

1. EMBL MEMBER STATES

EMBL currently has 19 member states: Austria, Belgium, Croatia, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, The Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the UK.

It is our goal to continue to provide outstanding life sciences services to our member states. Important mechanisms for this include: the Visitors Programme; advanced training for pre- and postdoctoral fellows; courses and conferences organised by EMBL scientists; the development of new technologies, institutional partnerships and technology transfer. The single most-significant impact is probably made by the EMBL alumni, 88% of whom take up positions in the member states and thereby help to spread the "EMBL model" within their national systems.

EMBL will continue to maintain and improve its excellent relations with the member states and welcomes scientists and government officials from all member states to visit. EMBL's activities will be actively promoted through publicity material, presentations and roadshows that illustrate EMBL's research and service activities, the PhD Programme and other training opportunities for young scientists. Up-to-date information about EMBL's activities is also provided to the member states through Council papers, the Annual Report and other publications. EMBL organises targeted outreach activities such as the Science and Society Programme and courses for European science teachers on a regular basis.

1.1. Partnerships with member state institutions

The SFL concluded that EMBL should further strengthen its current Units over the period of the 2007–2011 Programme, and further recommended that: "EMBL will use properly structured partnerships to promote the development of the life sciences in Europe, by leveraging the unique features and competence of EMBL, together with investments made at the national level" to develop carefully selected locally funded partnerships with national institutions. The principles of EMBL partnerships were described in the Scientific Programme 2001–2005 and approved by EMBL Council. Partnerships will be established for a defined period, usually between five and ten years, and can be extended after positive review. It is envisaged that the most successful partnership activities might develop into EMBL Outstations in the longer term.

As stated in the SFL: "EMBL partnerships are close cooperative affiliations between EMBL and external institutions of comparable standard, vision and international orientation. They are working relationships at the institutional level, not a matter of certification or of standard collaborative links between individual scientists. They are based on shared institutional goals and require synergy or complementarity. Their aim is to leverage the successful EMBL model and competences, together with the strengths of the partners, to create an interlinked system of excellent institutions and thus enhance the development of the molecular life sciences in Europe and the world."

EMBL will make special contributions to each partnership in the following areas as appropriate:

- involvement in setting up the scientific management of the partner institute or its advisory system;
- recruitment of locally funded international groups at the partner institutes in appropriate fields;
- help in the organisation of an international PhD programme;
- development of complementary and reciprocally accessible facilities at both institutions;
- short-term personnel exchanges;
- complementary collaborations.

EMBL has to make strategic decisions about the areas of the life sciences that it pursues. The partnerships have to be of mutual benefit and therefore partners focusing on complementary research activities are preferred. Partnerships have already been developed in the following areas: marine biology, molecular medicine and structural biology. Other possible areas of mutual interest are, among others, translational research for medical applications, epidemiology, regenerative medicine and systems biology. Complementarity might be achieved by covering a research area that is currently not pursued by EMBL or by contributing to an area in which EMBL is already involved. An example of the latter is the Partnership for Structural Biology (PSB) in Grenoble, which combines the diverse strengths of several partner institutes to build an integrated centre for structural biology that covers a broad range of technologies and expertise. Currently, three of EMBL's five sites are also locations for partnerships. It is likely that the strong local links on the EBI and Monterotondo sites will also eventually lead to more formal partnership agreements with our local collaborators, but partnerships will by no means be restricted to local activities near the EMBL sites.

EMBL partnerships encourage bilateral collaborations, participation of both partners in larger networks for fundraising, the organisation of joint conferences, the exchange of staff, and access to facilities and services. Each partnership should be with a national centre of excellence that wishes to train young scientists for a limited period of time before they move on to other national institutions.

Scientific excellence will be the guiding principle of all activities within the partnership. If the partner has not already established such a system, it is essential that it commits to a high-level, regular international evaluation of its activities with consequences for tenure and funding. Such evaluation can be modelled on the system of regular review of EMBL by expert panels on behalf of the SAC. The review board can be established with EMBL's support and, if desired, with its participation.

One of the core principles of EMBL is the staff turnover system. This creates flexibility to pursue new research fields and ensures that members of EMBL staff move on to national institutions. Even though we acknowledge that it is challenging to establish such a system at the national level, a partnership with EMBL should be used to achieve a degree of flexibility in the staff composition of the partner institution.

The popularity of this programme has exceeded our expectations and we are in the process of discussing five new partnerships with institutions in the member states. We intend to pursue the ongoing discussions vigorously and aim to establish several additional partnerships in the course of the next five years. However, a factor that will rapidly compromise our ability to engage with further partners is the very limited leadership structure at EMBL, and more administrative support for this programme will certainly be necessary if it continues to grow.

Four partnerships have been established, three of which are described below. The PSB in Grenoble is described in detail in Section D, Box 19.

1.1.1. The Molecular Medicine Partnership Unit

The first EMBL partnership in the area of molecular medicine was established in early 2002 with the Medical Faculty of the University of Heidelberg. During the initial pilot phase of three years, the Molecular Medicine Partnership Unit (MMPU) was co-directed by Matthias Hentze, EMBL, and Andreas Kulozik, Head of the University Department of Paediatric Oncology, Haematology and Immunology. Research focused on genetic diseases related to mRNA metabolism such as the haematological disorders beta-thalassaemia and thrombophilia. Joint training activities through the exchange of PhD students and postdoctoral fellows, as well as joint seminars and the co-organisation of conferences, were an essential component of the partnership from the beginning.

Following a very positive scientific review in 2004, EMBL and the University of Heidelberg mutually agreed to extend the partnership for ten years to strengthen the links between the two institutions. There will be an increased focus on recruiting additional research groups from each organisation, as well as intensifying activities in postgraduate training. New scientific areas that will be pursued include the investigation of:

- haematological and degenerative diseases of iron mismanagement;
- early cancer detection, diagnosis and prevention;
- pathophysiology and treatment of cystic fibrosis.

The MMPU offers a unique opportunity for training pre- and postdoctoral fellows. Those who are at EMBL and are interested in translational research can participate in projects at the MMPU, and medical professionals who have completed their MD can pursue research projects at EMBL. It is planned that they will be able to obtain a PhD through affiliation with the EIPP and successful completion of a research project of EMBL PhD standard.

In the medium term, the University of Heidelberg plans to provide a physical location to house a larger MMPU in order to create an interdisciplinary unit with significant critical mass.

1.1.2. Partnership with the Sars International Centre for Marine Molecular Biology

Life processes in the ocean and their evolution represent fascinating opportunities for discoveries concerning fundamental questions in biology, including the provision of new model systems collected from the sea. The constant progress of molecular techniques, which today permit genome-wide studies of biological function, allows molecular biology to spread from classical terrestrial model organisms to a variety of marine species and communities. Synergistic efforts in these directions are the aim of a partnership between the EMBL and the Sars International Centre for Marine Molecular Biology, which was established in June 2003.

Since then, joint scientific meetings have been held to bring together scientists from the Sars Centre and EMBL for in-depth discussions and scientific exchange. These contacts have already led to collaborative projects and personnel exchange. Future meetings will build on this initial success and, in the longer term, will contribute to the assembly of a scientific community of critical mass interested in marine molecular biology. The institutes will also encourage joint funding applications.

The two institutes provide mutual access to facilities and instrumentation for visiting scientists, as well as to databases. These include: the marine biology infrastructures of the Bergen area; EMBL's state-of-the-art facilities for genomics, proteomics and advanced light microscopy in Heidelberg; and specialised Outstation facilities such as structural biology beamlines at Hamburg and Grenoble, and the core databases at the EBI.

The Sars Centre has established an international PhD programme in Marine Molecular Biology. Students will receive research training at the Sars Centre, whereas the University of Bergen and EMBL, along with other international institutes, will actively collaborate in the planning of the programme's theoretical and practical course work. An agreement for joint PhD degrees from the University of Bergen and EMBL has been concluded and has joined the growing list of EMBL collaborations with universities throughout its member states.

EMBL has taken an active role in advising and evaluating the development and progress of the Sars Centre, by participating in its Scientific Advisory Board, helping the Sars Centre recruit new talent

internationally and supporting the Sars Center in promoting faculty turnover. This partnership has now fulfilled most of its initial aims and promises continued benefit to both institutions.

1.1.3. Partnership with DESY

The EMBL Outstation in Hamburg was established in 1974 on the DESY campus in order to promote the use of synchrotron radiation for structural biology applications in the life sciences. Over the past 30 years, EMBL Hamburg has developed a unique combination of strong on-site structural biology groups and synchrotron beamlines for macromolecular crystallography (MX), small angle X-ray scattering (SAXS) and X-ray absorption spectroscopy (EXAFS).

In May 2004, DESY and EMBL formalised a new interdisciplinary partnership whose goal is to construct and provide new generations of top-quality experimental facilities for life sciences at the future DESY infrastructures. The key components are the upgrade of the PETRA ring into a dedicated synchrotron radiation source (start of operation 2009/10), the construction of a vacuum-ultraviolet free electon laser (VUV-FEL) facility (start of operation 2005/06), and the construction of the European X-ray FEL (start of operation 2012/13). The partnership agreement with DESY will allow EMBL to participate in making these cutting-edge facilities available for applications in life sciences.

The future PETRA-III storage ring will provide an ideal opportunity for the European structural biology community to gain access to world-class synchrotron beamlines for structural biology. They will be tailored for future challenges in structural biology and imaging.

Through an integrated centre for structural biology, the services will be made broadly available to the European structural biology community, as well as used internally by EMBL and DESY scientists. DESY and EMBL will collaborate on joint research and development projects in life sciences and related fields of physics and technology. The cooperation will also extend to the common organisation of joint seminars, symposia or workshops and other scientific events, as well as joint training activities.

1.2. New and proposed member states

The integration of the ten new countries that joined the EU in 2004 is an essential step towards fulfilling EMBL's mission of being an inclusive organisation dedicated to spreading expertise in molecular biology throughout Europe. We invite all EU member states to join EMBL, including the new and prospective EU members. Special provisions that have previously been discussed by EMBL Council are designed to allow those countries easier access to EMBL. The following countries have joined EMBC, EMBO's governing body, but not EMBL: Czech Republic, Estonia, Hungary, Luxembourg, Poland, Slovenia and Turkey. Initial contacts have been made between EMBL and all of these countries except Turkey, and they have all expressed their interest in joining EMBL in the future. Croatia became the first of these countries to join EMBL when its application was approved by EMBL Council in November 2005.

Partnerships with EMBL are an attractive option for scientific institutions in new EMBL member states, and our plan is to establish selected partnerships once a relationship between EMBL and the scientific community in the new member state has had time to develop.

1.3. Associate members

EMBL's achievements have received wide recognition in the international research community. This status has prompted interest from non-member states regarding involvement in certain EMBL activities. Such involvement will be beneficial to EMBL, as well as to European science, provided it does not restrict or otherwise compromise the participation of the member states in EMBL. Two options for participation have

been approved by EMBL Council: the Junior Training Scheme (JTS) and the Associate Membership and Faculty Development Programme.

One country has shown interest in joining EMBL as an Associate Member and one country has approached EMBL to participate in the JTS. We expect a small number of these associations with EMBL to be established in the course of the 2007–2011 Programme. These will be carefully evaluated prior to further expansion of the programme.

2. Collaborations with institutions in non-member states

EMBL employs scientists from more than 60 nations and our links already reach far beyond Europe. We have good relations with institutions in several countries that involve scientific collaboration and the co-organisation of courses and conferences by EMBL scientists. For example, several light microscopy courses involving EMBL staff have been organised in Singapore, India, Japan, South Africa and Brazil. These collaborations will help to strengthen the links between Europe and the rest of the world. We are planning to set up only a few initially and to evaluate them carefully.

In a few cases, this has led to more formal collaborative links with research institutions in non-member states. A memorandum of understanding was co-signed by EMBL and Warsaw University in 2004 whose goals were to work towards Poland's membership of EMBL and to the establishment of a Centre for Biotechnology, Applied Informatics and Medicine on the University campus. The first non-European collaboration was established in 2005 with one of the leading molecular biology research institutes in India, the National Centre for Biological Sciences (TIFR) in Bangalore, to collaborate, exchange staff and organise meetings and workshops, such as the Functional Imaging Course that was organised in 2004 and funded by EMBO.

A collaboration was established with the Japanese National Institute for Basic Biology (NIBB) which is part of the National Institutes of Natural Sciences. This collaboration aims to promote joint research activities, exchange of staff, organisation of conferences and exchange of know-how. The NIBB would like to help increase the number of international visitors and participants in scientific meetings in Japan and has agreed to make travel grants available to EMBL scientists who wish to visit the NIBB and participate in scientific meetings.

We will pursue a small number of such collaborations and evaluate their usefulness in strengthening links between European and non-European scientists.

3. EMBL alumni: seeding Europe with top scientists

Part of the rationale for establishing EMBL was to create a centralised laboratory within Europe that could attract leading junior scientists from across the world and help establish them in Europe. The Laboratory has fulfilled this function very actively, attracting not only staff members of exceptional quality but also providing training to scientists at all levels and stages of their careers. While working at EMBL, scientists usually develop important collaborations and strong ties to the extensive European research network and usually choose to remain in Europe after leaving EMBL.

EMBL's policy of staff turnover not only gives the Laboratory regular opportunities to hire exceptional young researchers, it also creates a permanent outflow of highly qualified scientists. Upon leaving EMBL, the

majority (88%) move on to the member states, and many obtain important positions in universities, research institutes and industries in their national systems (see Annex 6). They do so despite attractive working conditions in the USA and other parts of the world. Helping to train and nurture top scientists, and to inculcate a desire to strengthen European research, is an important and unique service that EMBL can offer its member states and European science.

In the three decades of EMBL's existence, the body of alumni has grown to include more than 3,600 former scientists, students and support staff spread across Europe and the world. They are important both as individuals and as a network of people who have worked in a unique, international, interdisciplinary environment. The alumni form an important group that can help EMBL and strengthen European science in many different ways.

3.1. The EMBL Alumni Association – building a network

Most people who have spent part of their career at EMBL still feel a strong attachment to the Laboratory and keep in touch, either by continuing scientific collaborations with our research groups or by using our facilities. They also take advantage of workshops and other training opportunities, which help to disseminate knowledge, expertise and technology. Although many alumni keep up with what is going on through visits or by subscribing to the newsletter, others would like to take a more active role in EMBL's future growth and development. As a group, EMBL alumni have enormous potential to promote European science. In 1999, the Laboratory therefore founded an official Alumni Association and began contacting former colleagues. To date, nearly 1,000 alumni have joined the EMBL Alumni Association.

Membership is open to former staff, students or research fellows who have worked at any of the EMBL laboratories for at least six months, or who have served on the EMBL Council or SAC. Those individuals who have had an association with EMBL but who do not fulfil the criteria for full membership can be considered for associate membership on an ad hoc basis, with the goal of making the Association as inclusive as possible. The Association is governed by a board of 12 members who meet twice yearly to decide on policies and actions, as well as to organise alumni activities and events of interest to the extended EMBL community.

The goal of the EMBL Alumni Association is to promote scientific exchange in the field of the life sciences and related areas throughout Europe. The methods include establishing and maintaining a network between the members and the Laboratory, organising meetings, promoting scientific publications, and promoting young scientists. In 2003, the EMBL Alumni Association obtained charitable status and became authorised to collect tax-deductable donations for these purposes.

Specific activities and initiatives organised to date include the following:

- **Reunions and scientific gatherings.** An alumni reunion organised by the Alumni Association was held at EMBL Heidelberg in November 2004, providing an ideal opportunity for former staff to meet researchers currently at the Laboratory – the alumni of tomorrow – and learn more about how EMBL's research activities have evolved. It also let them catch up with old friends and make new ones. The reunion hosted scientific talks from past and present EMBL staff, poster sessions, practical discussions about how alumni can help each other, and a discussion on science and society.
- **Alumni website.** The Alumni Association homepage (http://www.embl.org/aboutus/alumni/) provides a virtual headquarters for EMBL alumni to find and contact former colleagues who consent to being listed. Services include: a members database with contact information on EMBL alumni, details on their current research activities and interests, information about funding sources and student exchanges; alumni news, bulletins and information on upcoming events; details on how to access methods or technologies offered by alumni or by EMBL's core facilities and services; and a jobs database where alumni can advertise open positions in their lab or find excellent students to fill them.

- **Local chapters.** Many alumni find that linking up with former colleagues in their area can be useful. EMBL alumni often face similar problems when moving to a new place, and a good chapter can act as a helpful local support structure for everything from learning the ropes of a national science system to finding good child care and schools. Chapters have now been established by volunteers in Austria, France, Greece, Italy, Spain and Portugal, Switzerland, Scandinavia and the UK. Regional meetings and activities have been held in Dresden (Germany), Vienna (Austria) and Barcelona (Spain). More are planned in the near future.
- **Fellowships.** The EMBL Alumni Association administers fellowships for young scientists to carry out research at one of the EMBL Units. Currently, these comprise a fellowship for a Swedish postdoctoral researcher (funded by the Swedish Foundation for Strategic Research, made possible through the efforts of former Director General and Alumni Association board member Lennart Philipson) and a Matti Saraste Memorial PhD Fellowship.
- **EIPP application pool.** This initiative provides EMBL alumni with an effective mechanism to find high-quality, motivated PhD students to work in their groups. EMBL alumni may establish contact with a pool of highly qualified students who applied to the EMBL programme but for lack of space were not admitted. The initiative has proven very successful and several students have obtained positions in the laboratories of EMBL alumni.

3.2. Future directions

Increasingly, the benefits of a strong network, both among alumni and between alumni and EMBL, are becoming clear. Considerable interest has been expressed among current staff to increase interactions with alumni. For example, the Laboratory's current pre- and postdoctoral communities have recognised the advantages of tapping into the experience of their former colleagues and are now establishing links to develop mentorship programmes to help with issues such as training, career development and more.

The possibilities for interaction among the extended EMBL community of current and former staff are very wide, and the potential for making a difference to individual scientists as well as to the European research community is enormous. As activities and initiatives grow, there will be an increasing need for coordination and support from within EMBL to help ensure that the alumni community and EMBL can make their ambitious plans a reality.

4. EIROforum

Over the past 50 years, Europe has established several world-leading, intergovernmental research organisations and facilities to address a wide range of fundamental questions in the natural sciences. These institutions combine excellent basic research programmes, infrastructures that serve diverse scientific communities and training centres for the European scientific community. EIROforum was created in 1998 by seven large intergovernmental research organisations responsible for research infrastructures that serve the scientific communities of their member states. EIROforum is a platform to promote research in Europe, increase visibility for these organisations and influence European research policy with respect to infrastructure. The partner organisations are: CERN, EMBL, EFDA-JET, ESA, ESO, ESRF and ILL.

A science policy paper ("Towards a Europe of Knowledge and Innovation") was published in early 2005. This paper outlines the position of EIROforum and presents its member organisations as an excellent potential source of advice on certain aspects of European research policy. Regular events and meetings are organised between the partner organisations and members of the European Commission, the European Parliament and other bodies responsible for European research policy. Even though the seven organisations are not always

of the same opinion, it has turned out to be much more effective to approach the Commission as a group rather than individually. EIROforum has become important as a means of providing a coordinated voice on relevant aspects of European science policy and we expect that participation in EIROforum will continue to be important for EMBL in the future. It will also assist EMBL in establishing good bilateral links with European community institutions. This will be especially important for finding a long-term solution for supporting EMBL's service infrastructures.

EMBL actively participates in EIROforum working groups on outreach and education, human resources, instrumentation, GRID computing and EU matters. The working group on outreach and education is participating in the largest EU-funded Science and Society initiative – NUCLEUS. EIROforum contributes two activities to NUCLEUS: "Science on Stage", which is coordinated by ESA; and the journal *Science in School*, which will be launched in 2006 and coordinated by EMBL.

5. European Union

EMBL and the EU share in common the goals of improving the quality and the integration of European research efforts in the life sciences. This has led to a continuous increase in consultation between the two organisations in terms of exchange of information and advice. The EC has observer status on EMBL Council. As mentioned previously (Section C.1.4.7.), EMBL researchers are in enormous demand both as coordinators of, and participants in, EU-funded integrative activities such as the Networks of Excellence and Integrated Projects of Framework Programme 6. EMBL researchers are currently participating in a total of 58 EU-funded network projects and act as coordinators for 12 of these (see Annex 10). Considering that EMBL only has 84 research groups and teams, these numbers are a remarkable reflection of the importance EMBL attaches to helping establish a dynamic European Research Area. The funding of these initiatives provides substantial and very welcome support to the Laboratory, and enables us to pursue numerous projects that would otherwise be beyond our means. EMBL scientists also participate in EU-organised strategy meetings to discuss European science funding in specific fields and are frequently recruited as reviewers and evaluators of applications for EU funding.

In more strategic terms, EMBL is involved in the development of new initiatives for European Science. Together with EMBO as a founding member of the Initiative for Science in Europe, EMBL has been very active in organising and helping lead discussion of the European Research Council (ERC). We see the advent of European-scale funding evaluation through the ERC as being of great potential benefit for the scientific competitiveness of Europe. Acting either alone or as a member of EIROforum, EMBL has also provided advice to the EU and the European Strategy Forum on Research Infrastructures (ESFRI) on questions concerning the requirements for European-scale life science infrastructures and the experience we have of operating such activities in EMBL's Units. We have co-organised meetings with the EU on future requirements for life science data resources in Europe. The EBI exemplifies why infrastructures, if organised well, represent an enormous saving for Europe. Duplicating the EBI's services in each member state would lead to an enormous increase in costs and decrease in efficiency. It is our opinion that Europe should carefully consider which other infrastructures for life sciences should be organised in this way, and provide a mechanism by which EU member states and other interested countries can participate communally in meeting the construction and running costs of these infrastructures.

It is our view that a continuation of the constructive dialogue between EMBL and EU bodies that deal with research and infrastructure funding and planning is of value to our member states and their research communities, and it is therefore our intention to pursue this dialogue actively.

6. Communication and public affairs

As an institute that is largely publicly funded, EMBL has a responsibility to provide thorough, accurate and understandable information about its work to citizens and decision-makers in the member states. The best source of detailed information about EMBL research is its scientific publications. However, these are generally written for specialist audiences with a high degree of scientific expertise. Thus, we have also been developing a wide range of other types of information to make our work as accessible as possible to those who support the Laboratory. Most of these publications are produced by the Office of Information and Public Affairs (OIPA), which was founded in 1997 with the goal of improving both internal and external communication. Since 2002, EMBL-EBI has had a scientific outreach officer who handles public relations and outreach in close collaboration with OIPA. The other Outstations rely on OIPA for a growing number of Laboratory information and outreach functions, which are briefly described below.

6.1. Documentation and publications

OIPA produces the official EMBL Annual Report and Research Reports on a regular basis. On demand, OIPA has been producing brochures, exhibits and official documents for the Director General's Office, as well as teaching kits and other types of information that are used in many circumstances where targeted information is required. There is a growing need for more information of various kinds, particularly brochures regarding Laboratory services and projects. We are planning to introduce a uniform design and layout for all public documents that are published by EMBL. For a list of documents and brochures that were published in 2004/2005, please see Annex 20.

6.2. EMBL websites

With the incredible growth of the internet, it has become increasingly important to develop and provide content that can be accessed electronically. Our goal is to provide information that is concise, thorough, up-to-date and accessible to many different types of audience. The websites have now been given a uniform design and brought under one umbrella (www.embl.org). Each EMBL site has its own local, specific information website as part of this structure. Maintenance and updating of all general websites is handled by OIPA. The present content management software needs to be replaced in the near-to-medium term with a standard content management system that will be able to handle the increasing amount of information produced for all EMBL websites. The content management software will be selected and implemented in close collaboration between OIPA and IT Services in Heidelberg and the Outstations. The special, high-volume bioinformatics services that EMBL-EBI must provide reliably, to cope with an average of 2 million information requests a day, means that the EBI requires both dedicated personnel and a separate website.

The EMBL.org website also currently fulfils an important function as a major repository for internal information for EMBL staff. The fact that EMBL scientists are spread across five different sites and have needs that can vary widely will continue to raise demand for better and more flexible communication tools. In order to facilitate the exchange of data and to allow the creation of shared electronic workspace for cross-Unit project teams, a dedicated intranet portal is necessary. OIPA will collaborate with IT Services to build such a portal. A future intranet web service with a clear separation of the intranet from public web server content will allow better structuring and facilitate accessibility of content. At the same time, this will create the opportunity to be more flexible and allow individual content management by research groups or staff members. This, together with the new content management system and standardised user interfaces, will greatly enhance the collaborative character of the portal as well as the end-user's ability to use the information provided flexibly.

6.3. Press and media relations

EMBL created a press office and hired a full-time press officer in 2004. The goal has been to increase the visibility of EMBL in the media, thereby informing a wider audience about our activities and accomplishments. The press officer centrally coordinates all EMBL media relations, with support from the EMBL-EBI scientific outreach officer, and acts as a knowledgeable interface between scientists and media professionals. The Laboratory publishes more than 30 press releases per year, usually intended for the mainstream media, concerning scientific work, announcements of major grants, projects and events, and other newsworthy items. Many of these stories have received wide coverage in the national and international European press. EMBL-EBI has also focused on improving coverage in the scientific press, in order to reach its user base more effectively. We expect that this activity will continue at the same level over the next five years. An overview of the press releases that were published in 2004–2005 and their coverage by the general media is provided in Annex 21.

6.4. Areas for further development

- We will continue to reflect upon and improve EMBL's documentation. Integration of the contents of the regular print publications with electronic information will help to facilitate the collection and storage of information. We will carefully select the appropriate formats for each publication depending on content and target audience.
- An EMBL design guideline will be introduced to present EMBL to the outside world in a more uniform and easily recognisable way. We have started this process by implementing the new websites in a more uniform design and by redesigning some of the print publications.
- EMBL increasingly needs brochures to spread focused information about specific activities, particularly for bioinformatics, structural biology and other service-oriented Units and facilities, but also for training programmes, etc.
- We expect that the trend established under Framework Programme 6 – where EU-funded projects are expected to disseminate information about projects and results widely – will continue and probably increase. OIPA will provide support and advice to project administrators who have little experience in this area.
- We need to have a greater institutional presence at international conferences and political events related to research, using such occasions to advertise our services, further our missions and tighten our network of alumni. OIPA, together with representatives from each EMBL site, has initiated a coordinated approach to exhibiting at such conferences.
- We are continually seeking ways to improve internal communication within and between the Laboratory sites. The development of an intranet portal will have the largest impact for the scientists but we also need to improve the dissemination of non-scientific information throughout EMBL. The newsletter itself is not sufficient for internal communication; therefore, we are developing a web-based portal (today@EMBL) that will be updated daily with news, events, announcements, general information, interactive functions, calendars, etc. This will initially be developed as part of the introduction of a content management system for the websites; however, eventually, it will be a component of a fully-functional intranet that will be developed in collaboration with IT Services (see Section C). This is one example of the need to improve the coordination of information between the EMBL sites; there are others – for example, in media relations.
- Framework Programme 7 is likely to see a much-enlarged budget for Science and Society activities, which EMBL could participate in, certainly in the framework of the Science and Society Programme and ELLS, and where appropriate in coordination with EMBO and EIROforum. We are developing a strategic plan that addresses the proper level of such activities and how they can contribute to the missions of the Laboratory.

- OIPA will continue to contribute to communication training for EMBL staff and pre- and postdoctoral fellows.

7. Science and Society

During the last quarter of the twentieth century, a dramatic increase in interest in molecular biology and a change in its perception by the public took place. With the advent of the new biotechnologies and their application to food production, pharmaceuticals and biomedicine during the 1970s, 1980s and 1990s, knowledge and discussion of molecular biology spread far beyond academic boundaries. In the process, the situation of biology changed profoundly as new possibilities were created that had socio-economic, regulatory and even moral significance. New forms of knowledge about the workings of biology, ranging from media-led debates to highly sophisticated interdisciplinary think-tank expertise, developed and spread throughout society.

Faced with such growing public involvement and exposure, the number of scientists who became engaged in promoting a better understanding of science among the public, discussing and debating how science can best serve society, increased. EMBL and EMBO, as flagship scientific organisations in Europe, have been at the forefront of this movement and have actively incorporated these important concerns within their activities. As a leading European research laboratory, EMBL recognises its obligation to expose its scientists to the evolving social concerns and ethical debates relating to applications that result from progress made in the life sciences. EMBL sees the importance of engaging scientists to reflect seriously and participate effectively in science–society interactions.

To this end, EMBL launched a Science and Society Programme in 1998 and this has become an integral part of scientific life in the Laboratory. A variety of activities and events organised at the different EMBL sites 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. Some of the EMBL Science and Society activities are targeted exclusively to the EMBL research community, whereas others are directed towards a broader audience and are open to the general public. The EMBL Science and Society Programme has initiated highly successful collaborations, primarily with EMBO, but also with other leading life science and cultural institutes, jointly organising thematic symposia and yearly interdisciplinary conferences. In 2001, EMBL, the German Cancer Research Centre (DKFZ) and the University of Heidelberg launched a public seminar series, "Heidelberg Forum – Biosciences and Society", aimed at informing and engaging local audiences. The EMBL Science and Society Programme has edited and produced several special issues dedicated to Science and Society published by high-quality journals, promoting its communication and reaching broad audiences. These are freely accessible to the public via the internet (http://www.embl.org/aboutus/sciencesociety/publications.html).

The life sciences have enormous potential for further development and practical application. However, a popular consensus needs to be developed as to how to assess and deal with the diverse repercussions of such development. More than ever, in the years ahead, there will be a need for interdisciplinary dialogue to inspire synthetic insights and a common worldview. The new ways in which science is now being applied in the pursuit of knowledge and economic growth must be carefully adjusted to public interests and value systems across Europe. It is the common responsibility of all, scientists as well as non-scientists, to engage in an ongoing process of shaping a shared understanding of science. The EMBL Science and Society Programme will continue to work towards this important goal.

G. ENDOWMENT

Basic molecular biology research often takes unpredictable turns that create opportunities for new and exceptional projects. In some cases, these new projects will not have been considered during the preparation of the EMBL Programme owing to the length of the funding cycle. However, such projects might present a unique opportunity for EMBL that should not be missed if funding can be raised. Setting up a foundation to manage donations and allocate them to such projects is one method of overcoming these funding bottlenecks.

The EMBL Endowment Foundation (EMBLEF) was approved by Council in November 2002. EMBLEF was to be established as a charitable trust and to be managed independently from EMBL. Since its establishment, EMBLEF has attracted several distinguished scientists and influential scientific leaders from several countries as members of its Board of Trustees and Advisory Board: Prof. Kari I. Kivirikko, Prof. Daniel Louvard, Dr Jean-Francois Conscience, Sir Ken Murray and Prof. Christiane Nuesslein-Volhard have agreed to join the Board of Trustees, and Sir Paul Nurse and Prof. Fotis Kafatos have been appointed to the Advisory Board. The Director General initially appoints the members of both Boards.

EMBLEF was established in summer 2003. Since then, the essential infrastructure of the foundation has been set up. Through interaction with EMBL scientists, the development of networks and contacts, and cooperation with the Office of Information and Public Affairs to increase EMBL's media coverage, the groundwork is complete to begin fundraising for defined projects. The first large fundraising initiative will be for advanced training activities at EMBL, specifically aimed towards the realisation of the ATC project. Future areas for fundraising include disease-related research projects and scientific opportunities with high potential for new discovery.

H. ADMINISTRATION

1. Implementation of human resources management solutions

The size of EMBL and its administrative complexity, with 19 member states, four host countries and the staff turnover system, imposes a requirement for efficient financial and personnel management systems. As a result of the financial difficulties of the Laboratory in the 1990s, much of the administrative infrastructure became outdated. In 2003, the EMBL Administration began to correct these deficits and implemented SAP financial and purchase modules. The introduction of SAP Human Resources Management solutions is a necessary next step to streamline the workflow processes within the whole of the administration, and to provide an integrated system with improved control environment and a high level of transparency.

The considerable increase in staff numbers since 2000, due mainly to growth of the EBI and Monterotondo Outstations, has had a major impact on the work volume of the EMBL personnel section. Moreover, this work volume is larger than in most other organisations owing to the limited tenure policy and high staff turnover. This necessitates the use of integrated computerised support in order to manage the daily work efficiently. Furthermore, such integrated computerised support will have the additional benefit of increasing the level of service offered to staff, management and external bodies, as well as provide high-quality, up-to-date, human resource (HR) information and analyses.

The transition to SAP payroll is scheduled to take place in 2007 and clearly takes priority over the implementation of other SAP-HR modules. The SAP licence package purchased by EMBL will contain a range of HR support tools such as SAP Travel Management, E-Recruiting, Performance and Organisational Management. To ensure that these tools adequately provide the support and output to meet the administrative needs of EMBL, a comprehensive review is required, not only to assess currently existing administrative work processes, but also to fully assess the information needs of the whole of the EMBL community, including the EMBL Council and external bodies. This review process was started during 2005. The technical customisation process of these additional tools is foreseen to commence in 2007.

2. Heidelberg refurbishment

The main laboratory in Heidelberg was constructed more than 25 years ago and has never undergone a thorough renovation. Refurbishment had become absolutely essential and the first phase of the project, which is to be financed by a long-term loan, was approved by Council in November 2004. The loan and interest will be repaid over the period 2007–2011 from a supplementary grant from Council. The entire renovation will be carried out over the period 2005–2010 as the funds required for the different phases of the operation become available. The refurbishment will lead to improved and safer facilities for the scientists, as well as reduced operational and maintenance costs as a result of improvement in the building plant infrastructure.

The first phase of the refurbishment programme, costing some €9.8 million, is already underway. Work on extremely urgent elements was started in 2004 and, with Council approval, is planned to continue until the end of 2006. This phase consists of the most urgent work, relating to areas in which the ability of the Laboratory to function or the safety of personnel were threatened, and comprises:

- work on heating and hot water systems where modernisation is necessary to facilitate measures against Legionnaires disease and to prevent breakdowns;
- renewal of plumbing and ventilation systems where fire safety had become an issue and where the age of machinery and deterioration in the ventilation and water systems were threatening to halt the work of the Laboratory in some areas;

- replacement of obsolete cooling systems, which had been introduced piecemeal and were threatening loss of capacity in the areas of microscopy and computer infrastructure;
- renewal of electrical systems, distribution and transformer equipment to meet current demands safely;
- renewal of lift machinery and re-sealing the terrace area to prevent leaks into the accommodation below;
- converting the Old Animal House back into usable accommodation;
- upgrading waste management facilities to avoid fire and health risks.

Full details of the work and associated costs were provided to EMBL Council in EMBL/Fin.Com./2004/40.

The second phase of the refurbishment programme is planned over the period 2007–2010, and comprises:

- further work on heating systems, together with some conversion work in the radioactive area;
- finalising replacement of water pipes;
- extension of cooling assemblies to less-urgent areas;
- renovation of internal and external lighting systems and conversion of main and sub distributors in the main building;
- renovation of insulation and floor coverings, and expansion of common and storage areas and of technical infrastructure areas.
- expansion and renewal of the canteen and cafeteria

Details of the costs and nature of work planned at this stage were provided in EMBL/Fin.Com./2004/20. Council asked EMBL Administration to provide a detailed plan for funding these items together with the draft indicative scheme 2007–2011.

After the implementation of the second phase, a regular programme of maintenance and renewal for all EMBL Units will be instituted, designed to avoid the need for emergency capital expenditure requests in the future. This will be implemented after completion of the second phase of the Heidelberg refurbishment (i.e. from 2011).

3. Introducing long-term care insurance at EMBL

Long-term care insurance covers part of the costs of home care or institutional care if one requires frequent or substantial help with normal day-to-day activities on a long-term basis. This need for help might be caused by physical or mental illness or disablement as a result of accident, illness or ageing. In 1995, Germany added long-term care insurance as a compulsory component to its health insurance schemes. Since then, many other European countries, as well as several international organisations, have introduced similar schemes.

EMBL Administration recognises the importance of this provision and has committed to including a long-term care component into its own social security scheme. However, the limited tenure policy and young population of EMBL necessitates a scheme that should be transportable worldwide after leaving the organisation and should include family members at a very reasonable cost.

The initial proposal to introduce a compulsory scheme with a private insurance company met substantial resistance when presented to the EMBL community, mainly because of its cost implications for staff departing from EMBL employment. The requirement to continue the insurance after leaving the organisation was not perceived as a necessity by the majority of staff, as some will be able to rejoin a national scheme or find one at their new place of work.

EMBL Administration is now developing a voluntary scheme whereby members of personnel and their families are eligible for a subsidy towards a long-term care insurance scheme of their choice. The subsidy amounts to 50% of the real cost premium, in line with contributions offered by other international organisations, with an upper limit that is increased for women to compensate partly for the typically higher premium costs charged by private insurers to women because of their higher life expectancy.

The cost implications for EMBL of a voluntary scheme can be expected to be substantially lower than its compulsory counterpart. Its disadvantages are that some members of personnel may be refused insurance cover following a medical examination, or may have a waiting period before insurance claims can be made.