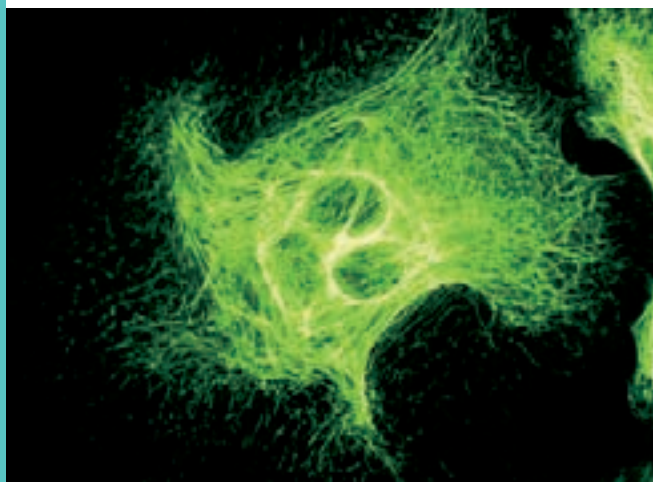


Functional genomics has emerged as a new and rapidly evolving scientific discipline dedicated to studying the functions of genes. With the complete sequencing of the genomes of an increasing number of organisms (bacteria, fungi, plants and animals), culminating in 2001 with the full sequence of the human genome, research priority has shifted from the identification of genes to the elucidation of their functions. The ultimate goal of functional genomics is the improved understanding of cell organisation at different levels, from individual genes to groups of biomolecules and complete genomes.

Integrated Approaches for Functional Genomics

An ESF scientific programme



High throughput technologies are a hallmark of functional genomics experimentation, with their capacity for collecting data on a genome-wide scale. DNA microarrays ('DNA chips'), where many thousands of genes are spotted onto an area no bigger than a microscope slide, allow researchers to sample thousands of genes in parallel for expression analysis in health and disease. Similarly, developments in proteomics, the protein complement of a cell or organism, are yielding information on the expression of proteins in different developmental stages, cell types or disease states, as well as on the complex protein interactions which underlie cellular functions. Knocking out individual genes in cells and organisms, and observation of resulting phenotypes, provide a direct insight into physiological function. Structural genomics, the determination of protein structure by physical methods such as X-ray crystallography or by computational prediction and modelling programmes, reveals functional roles by comparison with proteins of known structure and function.

These types of experiment create massive amounts of information which require efficient data management processes. Bioinformatics, the application of computers to biological data analysis, has become an integral part of almost all projects within genome research. Data acquisition, data analysis and databases to store and search for data are the main areas where informatics supports researchers.

By making available resources to foster collaborations between laboratories and for regular meetings and training, this programme aims to enhance the development of functional genomics in Europe.



The European Science Foundation acts as a catalyst for the development of science by bringing together leading scientists and funding agencies to debate, plan and implement pan-European initiatives.

Scientific background

The publication of the human genome sequence in February 2001 was hailed as a major step forward in human knowledge and the fight against disease.

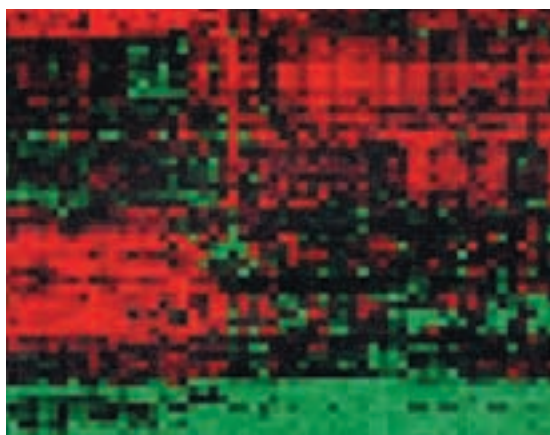
Simultaneously, the genomes of an ever-increasing number of organisms have been sequenced, with plants, bacteria, fungi, and animals all yielding their sequence information as a result of major international collaborations and cooperation. However, despite all the attention such revelations receive, the availability of the DNA sequences is only the first step in a complex process of discovery. It is the understanding of gene function and the subsequent application of this new information that will ultimately yield benefits. Consequently, a shift of emphasis is now occurring from genome mapping and sequencing to determination and

interpretation of genome function. This is the challenge for the post-genomic period which we are now entering and is the area known as *functional genomics*.

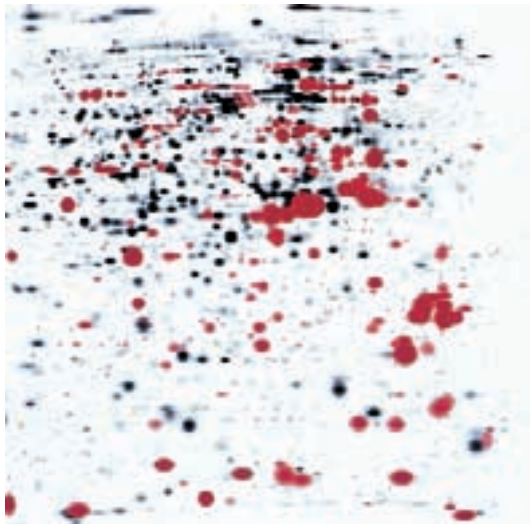
Functional genomics employs major innovative technologies for genome-wide analysis supported by information technology. These activities depend both on experimental and computational methods. While high throughput experimental technologies generate data on gene expression, protein structure and protein interactions, powerful informatics systems are required for the efficient management of experimental data, integration of information that is distributed in heterogeneous sources and establishment of 'computer assisted experimental strategies'.

Developments in large-scale, high-throughput technologies and robotics now allow researchers to simultaneously profile vast numbers of different genes or proteins in parallel. These technologies, which include DNA and protein microarrays (chips), large scale 2-dimensional gel electrophoresis and mass spectrometry (proteomics), are further keys to unravelling gene function. It is possible to observe the expression of many genes or proteins under a single situation, such as a developmental stage, in a particular cell type or in a certain disease state. Profiles can also be compared against a changing status, such as a disease or development. By building up these pictures and comparing them with data held in many laboratories, functional information on particular genes and proteins can be deduced.

What creates the complexity of an organism? The primary DNA sequence certainly holds an important part of the basic information, but the assembly and fine-tuning depends on how different proteins are modified and interact in manifold permutations. The complexity of the genome is more than matched by



DNA array data showing up and down regulation of genes in a cell.
© S. Brunak



that of the proteins it encodes; the proteome probably exceeds the genome by an order of magnitude in size, given alternative splicings of many genes and the post-translational changes by which proteins are modified. Molecular interactions are critical to most biological processes. Protein-protein interactions are being analysed in order to generate macromolecular interaction maps, e.g. knowledge of the yeast genome has enabled the systematic analysis of interactions for all 6,000 proteins.

Protein structure and function are inseparably connected. It follows that being able to determine or predict interactions and to model structures will illuminate protein functions. Major programmes of structure determination for all representatives of protein folds and families, by high throughput protein expression, crystallisation and X-ray crystallography, are under way in various centres.

Powerful informatics systems have been developed to manage the large quantities of experimental data and integrate this information. The massive amount of data generated on gene expression, protein structure, protein interactions and structure has been closely mirrored and facilitated by computer programs designed to organise and access it. Many *in silico* methods and programs have been

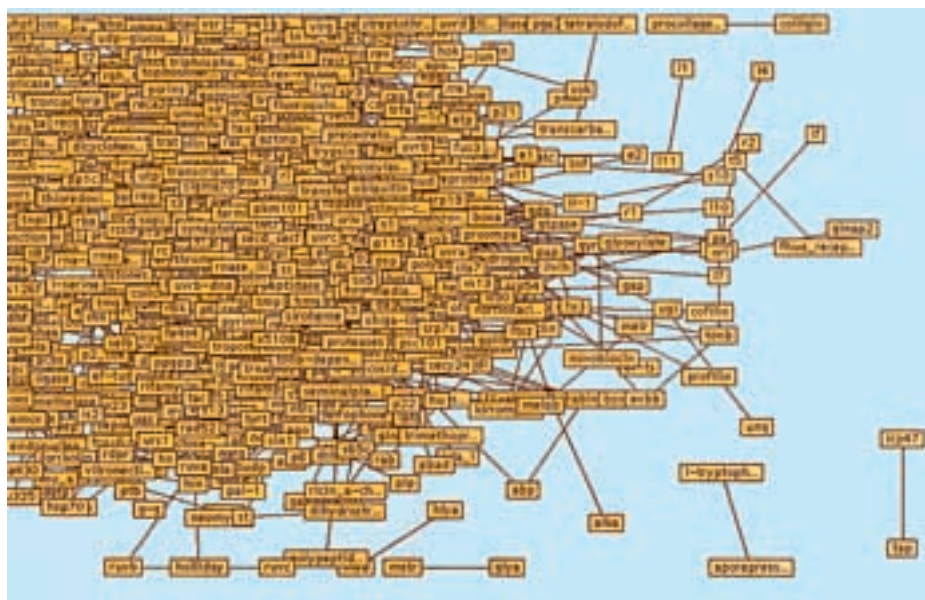
developed to support genomics activities and are routinely used in structure and function prediction. Functional genomics data open up new possibilities and challenges for bioinformaticians involved in the development of methods to address questions related to genome comparison, networks of protein interactions in cells and networks of coordinated expression in the transcriptome or proteome. As data accumulate, modelling and simulation of biological processes, from isolated biochemical pathways to the complete cell factory, are becoming possible and may well become the central focus of bioinformatics; relevant models and systems will need to be developed.

Despite all the progress being made, there are a number of perceived deficiencies and bottlenecks which currently affect functional genomics research. One is the lack of a structure or organisation which would enable scientists working in the same and different fields to exchange information about their technologies. Such communication is particularly crucial in view of the speed with which these fields are moving and the multidisciplinary nature of many of the technologies. Another major need which affects many of the methods is for appropriate benchmarking and standardisation, to enable valid exchange of experimental data and comparison of results and reduce the considerable 'noise' currently existing in the presentation of experimental information. A third problem is a lack of software tools for data handling and of a widely accepted suite of analysis methods integrated into a coherent system for prediction of protein structure and function. Fourthly, there is an increasing problem of insufficient or erroneous functional assignment in sequence databases; part of the error propagation is caused by the lack of traceability of results. This programme seeks to put into place the necessary organisation to enable these problems to be tackled at a European level.

2D electrophoresis gel showing about 10,000 different proteins expressed in a single cell. This is only a fraction of the total number of interacting proteins in the cell.
© J. Degett

The vast amount of textual information about biological systems stored as plain text abstracts of publications can only be handled with new bioinformatics platforms. Each box contains a possible protein name or chemical compound, the lines indicate different types of interactions.

© C. Blaschke - A. Valencia



The ESF programme

The objectives of the programme are to facilitate exchange of information, integrate scientific approaches and facilitate collaboration within Europe. Programme activities and resources are interdisciplinary, in order to disseminate the range of expertise available in Europe and result in the rational use of available resources. The programme will act as an incubator from which new collaborative projects will emerge that can later be proposed for public funding at either national or European level. Programme activities will include **workshops, training courses** and **conferences**. The Programme will also offer a number of **exchange grants**.

The programme is coordinated by a Steering Committee composed of one representative from each participating country.

Aims and objectives

Functional genomics is such a broadly encompassing term, that a structure was required for the programme, that included the major technologies involved and provided a means of linking these together and addressing common issues. The programme identified a set of ‘verti-

cal’ and ‘horizontal’ scientific areas that will be used to structure the organisation of cooperation at the European level. Vertical areas correspond to specific technological or scientific fields, while horizontal areas correspond to requirements common to several vertical areas. The vertical organisation is intended to foster collaborations in the development of parallel techniques, including individual technologies of expression profiling, proteomics and mutation analysis, in order to facilitate exchange and communication within particular fields. At the same time, the horizontal organisation is required to integrate functional and expression data from different sources with databases and literature information through bioinformatics tools of universal application.

The vertical areas are:

V1. Analysis of phenotypic changes resulting from mutagenesis and gene disruption

Mutation and gene disruption are powerful tools for investigating gene function and for identifying novel genes without prior knowledge of homology in other systems. Screening for phenotypes of interest leads to discovery of new genes, some of the mutations becoming

models for human genetic diseases. Model organisms include the mouse (*Mus musculus*), thalecress (*Arabidopsis thaliana*), yeast (*Saccharomyces cerevisiae*), fruit fly (*Drosophila melanogaster*), nematode (*C. elegans*) and bacteria such as *Pseudomonas spp.* A significant priority is to develop novel screens for recovering mutants, in parallel with expression profiling and proteomics to further define new phenotypes. The programme will bring together researchers working with different organisms to share ideas and technologies.

V2. DNA arrays and chips in expression profiling and mutation detection

DNA microarray or 'chip' technology, in which very large numbers of cDNA species or oligonucleotides are gridded for hybridisation to target DNA molecules, is advancing rapidly. Parallel expression analysis of many thousands of genes can be achieved. In expression profiling, DNA microarrays provide a global, high throughput approach to discovering which genes are expressed at a detectable level, where are they expressed, and which are over- or under-expressed in a given developmental stage, following environmental alteration or in disease. The use of different array systems has created a particular need for comparison of results and benchmarking between different laboratories.

The other key application of DNA array technology is in screening for polymorphisms and mutations, a major aim being the large-scale detection of DNA sequence variation amongst individuals. Characterisation of single nucleotide polymorphisms (SNPs) is important for the discovery of variations that affect biological function. This programme will aim to promote European integration by setting up a network of relevant laboratories, with establishment of appropriate standards for assays and

output files, facilitating comparison of information produced in different centres.

V3. Proteomics: protein identification, characterisation, expression and interactions

The proteome is the entire complement of proteins expressed in a cell or an organism. Proteomics includes the identification of proteins in biological tissues, the characterisation of their physicochemical properties (complete sequence, post-translational modifications), and the description of their behaviour (function, expression level). Proteomics involves many different methodologies, databases and bioinformatics tools, and is closely linked to developments in genome sequencing projects, microarray technology and expression profiling. While these provide complementary information, it is unusual for them all to be concentrated in a single laboratory. The programme will encourage linkages between proteomics groups and those with expertise in other technologies and bioinformatics through workshops and the interactive website.

V4. Structural genomics: protein structure determination, classification, modelling and docking

Structural genomics is the assignment of 3-dimensional structures to proteomes and the investigation of their biological implications. Protein structure is an important indicator of function, particularly where the structure of a new protein is homologous to one already known. Two levels of assignment are employed in structural genomics, one being experimental large-scale determination of protein structures using X-ray crystallography, and the other, computational structure prediction through detection of homologies with proteins of known structure. With numerous genome sequences already

available, it is becoming increasingly likely that a family to which a new protein belongs is represented already in the databases. Collaborative developments in structural genomics to be included in the programme are closely linked to those in bioinformatics tools and databases.

V5. *In silico* methods for the description of cellular systems by data and literature mining, predictions and simulations

Many computing methods and programs have been developed to support genomics activities and are routinely used in sequencing projects for sequence assembly, identification of coding and regulatory regions, homology searches, protein domain identification, and structure and function prediction. The improvement of these methods must be supported to achieve better accuracy and increased throughput and to include annotation information derived from experimental sources or directly from the literature. As data accumulate, modelling and simulation of biological processes, from isolated biochemical pathways to the complete cell factory, are becoming possible and may well become the central focus of bioinformatics. For this, relevant models and systems will need to be developed.

The horizontal areas are:

H1. Standardisation, benchmarking and comparison of different experimental systems

Standardisation and benchmarking in functional genomics are required at both the biological and informatics levels. It is important that researchers be able to combine their efforts by relating experimental results produced in one laboratory with those made in others. However, this is only feasible if the experimental procedures, the source materials used and the analysis

algorithms are comparable and well defined. Currently there is a serious problem of heterogeneity of experimental approaches and of the data generated and a detrimental lack of traceability of information. Hence information tools are needed which will register the source and the quality of data and increase homogeneity, to ensure that data from different laboratories match each other, use the same nomenclature and are expressed in the same way.

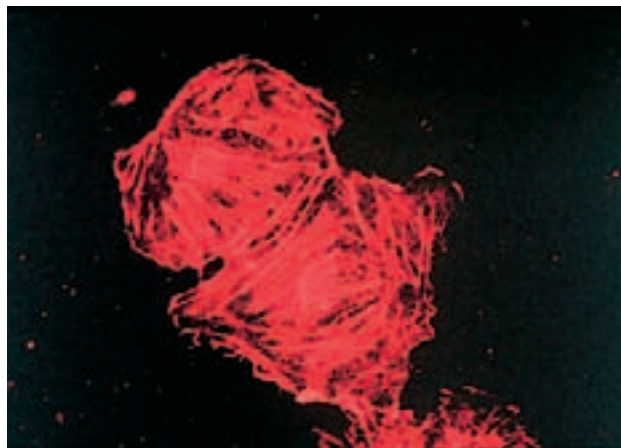
H2. Data management: data-bases, interfaces and ontologies

Bioinformatics has become an integral part of almost all projects within genome research. Only a few sophisticated tools have been developed so far for the complex data arising from expression profiling studies. Several European databases are under development, the main one being at the European Bioinformatics Institute (Cambridge) and another at the Max Planck Institute for Molecular Genetics (Berlin). Furthermore, no major database exists to deal with the large amounts of experimental data from different functional genomics projects although a number of specialised databases for certain expression data have been developed. Despite several attempts around the world to build databases specifically for the management of biomolecular interactions, no standard has yet emerged. The programme will aim to support efforts to achieve a common subset of ontologies, classes and structures that are best able to store and represent the experiments and resulting data in the areas involved.

H3. Combination and integration of functional genomics data to derive new biological knowledge

In order to go beyond the information acquired by individual functional genomics technologies, new biological knowledge will be derived by the

combination and integration of the different types of data produced in laboratories, extracted from the scientific literature and available in the databases. The first requirement for the experimentalist is to link his own results to external sources in order to better interpret and exploit them. Linking local data to external sources of information requires either direct references, such as cross reference in sequence databases, or computer supported methods, such as programs for homology searches, literature extraction or genome comparison. The programme proposes the development of integrative platforms (concepts and software tools) able to expand and increase knowledge accumulated in the individual fields.



Actin is the most abundant protein in most animal cells.
© J.E. Celis

Funding

ESF scientific programmes are principally financed by the Foundation's Member Organisations on an *à la carte* basis. *Integrated Approaches for Functional Genomics* is supported by 30 Member Organisations in 20 countries:

Österreichische Akademie der Wissenschaften, Austria; Fonds zur Förderung der wissenschaftlichen Forschung, Austria; Fonds National de la Recherche Scientifique, Belgium; Fonds voor Wetenschappelijk Onderzoek – Vlaanderen, Belgium; Akademie věd České republiky, Czech Republic; Grantová agentura České republiky, Czech Republic; Statens Naturvidenskabelige Forskningsråd, Denmark; Statens Sundhedsvidenskabelige Forskningsråd, Denmark; Estonian Science Foundation, Estonia; Suomen Akatemia/Finlands Akademi, Finland; Centre National de la Recherche Scientifique, France; Institut National de la Santé et de la Recherche Médicale, France; Deutsche Forschungs-

gemeinschaft, Germany; Magyar Tudományos Akadémia, Hungary; Hungarian Scientific Research Fund, Hungary; Enterprise Ireland, Ireland; Health Research Board, Ireland; Consiglio Nazionale delle Ricerche, Italy; Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Netherlands; Norges Forskningsråd, Norway; Polska Akademia Nauk, Poland; Slovenian Academy of Arts and Sciences, Slovenia; Consejo Superior de Investigaciones Científicas, Spain; Oficina de Ciencia y Tecnología, Spain; Vetenskapsrådet, Sweden; Schweizerischer Nationalfonds zur Förderung der wissenschaftlichen Forschung / Fonds National Suisse de la Recherche Scientifique, Switzerland; Scientific and Technical Research Council of Turkey, Turkey; Biotechnology and Biological Sciences Research Council, United Kingdom; Medical Research Council, United Kingdom; Natural Environment Research Council, United Kingdom.

Integrated Approaches for Functional Genomics Steering Committee

Mike Taussig (Chair)

The Babraham Institute
Babraham Hall
Babraham
Cambridge CB2 4AT
United Kingdom
Tel: +44 1223 496557
Fax: +44 1223 496045
Email: mike.taussig@bbsrc.ac.uk

Søren Brunak

*Center for Biological
Sequence Analysis*
Dept of Biotechnology
The Technical University
of Denmark
Building 208
2800 Lyngby
Denmark
Tel: +45 45 25 24 77
Fax: +45 45 93 15 85
Email: brunak@cbs.dtu.dk

Piotr Ceglowski

Institute of Biochemistry
and Biophysics
PAN
Pawinskiego 5a
02-106 Warsaw
Poland
Tel: +48 22 658 47 23
Fax: +48 22 658 46 36
Email: piotr106@ibb.waw.pl

Antoine de Daruvar

LION Bioscience AG
60, rue Lagrange
33000 Bordeaux
France
Tel/Fax: +33 5 57 57 12 47
Email:
daruvar@lionbioscience.com

Christoph Dehio

Division of Molecular
Microbiology
Biozentrum of the University
of Basel
Klingelbergstrasse 70
4056 Basel
Switzerland
Tel: +41 61 267 2410
Fax: +41 61 267 2118
Email:
Christoph.Dehio@unibas.ch

Joachim W. Engels

Institute of Organic Chemistry
Johann Wolfgang Goethe
University
Marie-Curie-Strasse 11
60439 Frankfurt/Main
Germany
Tel: +49 69 798 29150
Fax: +49 69 798 29148
Email: joachim.engels@
chemie.uni-frankfurt.de

Michel Georges

Service de Génétique
Factorielle et Moléculaire
Faculté de Médecine Vétérinaire
Université de Liège
Boulevard de Colonster 20 B
43/3
4000 Sart Tilman (Liège 1°)
Belgium
Tel: +32 4 366 41 50
Fax: +32 4 366 41 22
Email:
Michel.Georges@ulg.ac.be

Manuela Helmer Citterich

Dept of Biology
University of Rome
Tor Vergata
Via della Ricerca Scientifica
00133 Rome
Italy
Tel: +39 06 72 59 43 14
Fax: +39 06 20 23 500
Email: citterich@uniroma2.it

Des Higgins
Department of Biochemistry
University College Cork
Co. Cork
Ireland
Tel: +353 21 4904 134
Fax: +353 21 427 4034
Email: d.higgins@ucc.ie

Radovan Komel

Medical Centre for
Molecular Biology
Institute of Biochemistry
Medical Faculty
University of Ljubljana
Vrazov trg 2
1000 Ljubljana
Slovenia
Tel: +386 1 543 7640
Fax: +386 1 543 7641
Email: komel@ibmi.mf.uni-lj.si

Martin Kuiper

Department of Plant Genetics
Flanders Interuniversity
Institute for Biotechnology (VIB)
University of Gent
K.L. Ledeganckstraat 35
9000 Gent
Belgium
Tel: +32 9 264 5182
Fax: +32 9 264 5349
Email: makui@gengenp.rug.ac.be

Ulf Landegren

Rudbeck Laboratory
Dept of Genetics and Pathology
Dag Hammarskjölds väg 20
75185 Uppsala
Sweden
Tel: +46 18 471 4910
Fax: +46 18 471 4808
Email:
ulf.landegren@genpat.uu.se

Jean Louis Mandel

IGBMC – CU de Strasbourg
1 rue Laurent Fries – BP 163
67404 Illkirch Cedex
France
Tel: +33 3 88 65 32 44
Fax: +33 3 88 65 32 46
Email: mandeljl@igbmc.u-strasbg.fr

Andres Metspalu

Dept of Biotechnology
Institute of Molecular and Cell
Biology, Estonian Biocentre
University of Tartu
23 Riia Street
51010 Tartu
Estonia
Tel: +3727 375 029
Fax: +3727 420 286
Email: andres@ebc.ee

Meral Özgüc

Dept of Medical Biology &
TÜBITAK DNA/Cell Bank
Faculty of Medicine
Hacettepe University
06100 Sıhhiye Ankara
Turkey
Tel: +90 312 305 14 83
Fax: +90 312 311 07 77
Email:
mozguc@genetic.gen.hun.edu.tr

Václav Paces

Institute of Molecular Genetics
Academy of Sciences of the
Czech Republic
Flemingovo nám 2
166 37 Praha 6
Czech Republic
Tel: +420 2 2431 0234
Fax: +420 2 2431 0955
Email: vpaces@img.cas.cz

Hans Prydz

Biotechnologiserenteret
University of Oslo
Gaustadalleen 21
0349 Oslo
Norway
Tel: +47 2284 0531/32
Fax: +47 2284 0501
Email: hans.prydz@biotek.uio.no

Mart Saarma

Institute of Biotechnology
University of Helsinki
Viikinkaari 9 – P.O. Box 56
00014 Helsinki
Finland
Tel: +358 9 1915 9359
Fax: +358 9 1915 9663
Email: mart.saarma@helsinki.fi

Helmut Schwab

Institute for Biotechnology
Graz University of Technology
Petersgasse 12/I
8010 Graz
Austria
Tel: +43 316 873 8418 or 8419
Fax: +49 316 873 8434
Email: helmut.schwab@TUGraz.at

Alfonso Valencia

Protein Design Group
U. Autonoma de Madrid
CNB - CSIC
28049 Madrid
Spain
Tel: +34 91 585 45 70
Fax: +34 91 585 45 06
Email: valencia@cnb.uam.es

Gert Jan B. Van Ommen

Leiden University
Medical Center
Dept of Human Genetics
Wassenaarseweg 72
PO Box 9503
2333 Leiden
Netherlands
Tel: +31 71 527 62 93
Fax: +31 71 527 60 75
Email: gjvo@lumc.nl

Andras Varadi

Biological Research Center
Institute of Enzymology
Hungarian Academy of Sciences
Karolina út 29-31
1113 Budapest
Hungary
Tel: +361 466 9498
Fax: +361 466 9498
Email: varadi@enzim.hu

Martina Hildebrandt

ESF Senior Scientific Secretary
Life & Environmental Sciences

Marianne Minkowski

ESF Senior Scientific Secretary
Biomedical Sciences

Joanne Dalton Goetz

ESF Administrative Assistant
European Science Foundation
1 quai Lezay-Marnésia
67080 Strasbourg Cedex
France
www.esf.org
Tel: +33 (0)3 88 76 71 22
Fax: +33 (0)3 88 37 05 32
Email: jdalton@esf.org

External Programme Administrator:

Annette Martin

The Babraham Institute
Babraham Hall
Babraham
Cambridge CB2 4AT
United Kingdom
Tel: +44 1223 496246
Fax: +44 1223 496045
Email: annette.martin@bbsrc.ac.uk

Cover picture: Vimentin filaments in a cell showing a complex structure pattern. Vimentin is just one of many structural proteins. © J.E. Celis

November 2001

© European Science Foundation

For the latest information on this programme consult the following web sites: www.esf.org/genomics and www.functionalgenomics.org.uk