EU Projects Workshop Report on

Systems Biology

for the
European Commission, Research Directorate General
Directorate F - Health Research
Based upon a Workshop held in Brussels, Belgium on 8 December 2004

by

Ewan Birney, Andrea Ciliberto, Morten Colding-Jørgensen, Albert Goldbeter, Stefan Hohmann, Martin Kuiper, Hans Lehrach, Gisela Miczka, Erik Mosekilde, Hans Westerhoff, Olaf Wolkenhauer

Editors: Philippe Jehenson, Frederick Marcus (European Commission)

EXECUTIVE SUMMARY

Systems Biology (SB) is a rapidly developing scientific discipline that has already led to important advances in understanding biology and in developing medicines and their use for improving human health. SB is also an approach to explaining, predicting and controlling the complex cellular and physiological phenomena of living organisms in terms of the underlying chemical and physical processes and the involved feedback regulations on many different time and space scales. A systems approach requires the formation of EU-wide teams to mine and model the wide range and amounts of biological data coming from new technologies, both from high throughput and from detailed quantitative observations, and to direct the design of experiments. The approach is characterised by the use of mathematical modelling and modern simulation and data handling techniques to complement the present strongly empirical approach in the biological sciences. Furthermore, much of current research to combat complex diseases is characterised by very heterogeneous and fragmented efforts, which constitute a major drawback. Systematic and co-ordinated approaches that focus on the functional analysis of complex biological networks offer a tremendous potential for future medical and genomics research.

Leading participants in EU funded projects have identified key areas in systems biology for the near future:
- **GENERAL SYSTEMS BIOLOGY APPROACHES AND TOOLS**: SB within other projects; Areas for analysis; Types of Analysis; Methodologies; SB Data Integration Packages and various types of Modelling.
- **UNDERPINNING BIOINFORMATICS AND EXPERIMENTAL DATA SUPPORT**: Tools and Services: Bioinformatics, Databases, Software, Access, Services, Research and Infrastructures as the basis for systems biology (Model systems and Biobank Resources, Standards and Ontologies, Obtaining data within and beyond present 'omics', maintaining and expanding current databases).
- **TOPICAL AREAS FOR SYSTEMS BIOLOGY**: Developing systems biology (Cellular and sub-cellular systems; Efficient large-scale study of multiple interacting systems at the cellular and physiological levels; Physiology); Applying systems biology (Disease, Medicines and Treatment, Biotechnology, New Applications); Training.

JANUARY 2005

DISCLAIMER: This workshop was initiated and organised by staff of the Commission Services, who participated in this workshop and who assembled and edited this report with the assistance of the participants. The invited external experts provided both written and oral contributions to this report, and all the views expressed both individually and collectively in this report are those of the external experts, and may not in any circumstances be regarded as stating an official position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of the following information.

Contact: Frederick.Marcus@cec.eu.int Philippe.Jehenson@cec.eu.int See also: http://www.cordis.lu/lifescihealth/thematic.htm , http://www.cordis.lu/lifescihealth/genomics/home.htm , http://europa.eu.int , http://www.cordis.lu .

OVERVIEW OF SYSTEMS BIOLOGY

Why do systems biology (SB)? A systems approach to life sciences research is necessary for better qualitative and quantitative understanding of the functioning of biological systems in physiological and pathological conditions. This includes the exploitation of both biological information generated by high-throughput technologies as well as new types of data becoming available through developments in a broad range of experimental techniques such as mass spectroscopy, enhanced Raman spectroscopy, photoluminescence, space and time resolved laser interference, automated patch clamping, nuclear magnetic resonance spectroscopy, bioimaging, etc. The current research paradigm of one laboratory - one gene - one function - one disease has reached its limits, as diseases or important traits often arise from a combination of factors. Cataloguing and classification of molecules will ultimately not suffice to reason about the function of molecules or functioning of cells. Many important aspects of biology can only be researched by a combined experimental and computational approach to developing systems models that yield useful results. The implications of the findings in the life sciences will be immense and will allow a virtual description of cellular and physiological activities and functions. The ultimate goal is to understand the functioning of physiological and pathophysiological systems, and thus to rationalise model-driven drug development, and to investigate effects of treatments or support fruitful approaches in biotechnology. Knowledge gained from SB may also open up disease prevention and treatments avoiding extensive use of drugs.

Its role in studying health and disease: Prof. Denis Noble (Oxford) says: (in "*Modelling the heart: from genes to cells to the whole organ*") "successful physiological analysis requires an understanding of the functional interactions between the key components of cells, organs, and systems, as well as how these interactions change in disease states. This information resides neither in the genome[1] nor even in the individual proteins that genes code for. It lies at the level of (extensive) protein interactions within the context of subcellular, cellular, tissue, organ, and system structures. There is therefore no alternative to copying nature and computing these interactions to determine the logic of healthy and diseased states." This is even more the case if predictions have to be made which are quantitative, within the accuracy of the available data. After all, many processes in the body have to be kept within specific and possibly quite narrow limits to be compatible with health, or even with survival.

"The rapid growth in biological databases; models of cells, tissues, and organs; and the development of powerful computing hardware and algorithms have made it possible to explore functionality in a quantitative manner all the way from the level of genes to the physiological function of whole organs and regulatory systems. Systems physiology of the 21st century is set to become highly quantitative and, therefore, one of the most computer-intensive disciplines."

What is SB? SB involves developing the understanding of a biological system through the mathematical/computational modelling of the interactions of components of the system, leading to the expression of this understanding in qualitative and quantitative terms - in particular, in terms amenable to electronic storage and communication. Examples of 'Biological Systems' are as old as modern biology, e.g. the Krebs (citric acid) cycle in metabolism. However, this system represents a typical example of what arduous gene-by-gene approaches can accumulate over time. These no doubt are systems, but not accomplished by an SB approach, nor do they constitute modern SB. Once this knowledge is framed in a dynamical simulation model it can, however, be used for SB. SB has long been developed in physiology and in modelling the effects of medicines. The key changes that make a modern approach to SB necessary and possible are the very recent developments of high-throughput technologies in all domains of biology as well as emerging technologies that allow generating new types of quantitative data at high precision and resolution. Furthermore, recent developments of complex systems theory have provided us with the mathematical concepts and tools we need to understand some of the complex dynamical phenomena we observe in the living world. Analyses of just a small fraction of the available data has led to the realisation that understanding of biology, health, disease and medicines requires an integrated approach to studying the processes involved. Even with our current state of knowledge, SB has already made impressive contributions to both fundamental understanding and to direct applications to health. As examples, the circadian rhythms such as 24 hour sleeping cycles can only be fully described with an SB approach. Carefully tailored models can already make useful predictions about disease processes and how medicines can be optimally applied, constituting the beginnings of personalised medicine, for example in optimising the timing during the day of medication for chemotherapy or for diabetes treatment.

Trends in SB: Theoretical (mathematical) and experimental biology have for many decades been separate disciplines. SB, on the other hand, emerges as a new discipline in which theoreticians and experimentalists

[1] We presently cannot decode this information from the genome sequence or from properties of individual proteins.

closely collaborate, ideally from the planning of an experimental study. There is a need for a continuous and iterative collaboration between modeller and experimentalist such that the modeller understands biological knowledge about the system and takes part in the definition of new experiments and that the experimentalist understands the principles of converting biological information into mathematical descriptions.
A dominant theme is how to build up complex systems "from genes to cells" by combining knowledge from different databases, different types of data, etc. This approach has been used with some success, but generally it creates more questions than answers. Scenarios emerging from such approaches are sensitive to the precise way the systems are built. Data are important, but the way these data are utilised in modelling is much more important. In academia it is of interest to develop complex models based on genetic information etc., and many "virtual cells" show interesting cell-like behaviour. However, industry has special problems with validation. Industry staking hundreds of millions of euros on a medicine or bioprocess based on a computational model needs that model to be as useful and comprehensive as possible. This is not achieved by better data alone. Rather, the way the model is built, the intimate interplay between its parts and even the software used to make the calculations are critical. This also means that tool development in SB is critical and needs to be both practical and visionary. SB profits from many different disciplines in the life sciences as well as bioinformatics, information technology, dynamic systems theory, etc. Systems is a science in its own right in that it aims at discovering unknown principles and 'laws' that occur in biological systems.
Approaches to modelling: There are two main modelling approaches in SB (also called Computational Biology (CB) or even CSB). Besides the construction of large-scale models, incorporating as many details as have been uncovered experimentally on a given pathway or signalling system and on a detailed "cartography" of various networks, another useful approach in SB relies on the construction of small-scale models of limited complexity, containing a reduced number of variables (2 to 20), and aiming at addressing specific questions. From these small-scale models, one can often derive conclusions of more general significance, e.g. concerning cellular rhythms, cell signalling and cell cycle dynamics, especially when including dynamic phenomena: multistability, oscillations, spatial and spatio-temporal patterns (e.g. in morphogenesis and cell to cell communication). Both approaches have merits and limitations, and they can converge by putting small-scale models (modules) into a common framework. Model standardisation is vital, but should not be used to suppress creative approaches. Models should also be closely tied with existing and/or new experimental data.
Who will benefit from this approach? There are many "customers", e.g. university researchers in genomics, university physiologists, medical researchers, pharmaceutical industry researchers, hospital clinicians, family and research doctors, patients, public health practitioners, the pharmaceutical industry, new and existing biotechnology enterprises, agrobiotechnology etc., and disciplines, e.g. cellular biology, physiology and medicine, which will benefit from SB, all of whom/which will ask different types of (scientific) questions. Expected benefits will be widespread, since a thorough understanding of complex biological processes will allow us to tackle many of our real world problems. Complex biological networks decide if we live or die, what we can eat, and whether the environment in which we live is sustainable. In providing solutions to many of these problems, systems biology might therefore be one of the key approaches of the 21st century.
What is the state of the art? SB is an emerging field in the life sciences, having a history in physiological modelling coupled with pharmacokinetics and pharmacodynamics. Impressive advances heve been made in modelling a number of individual processes in physiology, e.g. in modelling the dynamics of the heart. Increasingly however, modelling of molecular processes, involving most or all genes, gene products and metabolites is being used to understand complex disease processes. However, it is vital to appreciate where models have worked, and where not. The most successful current implementations of SB rely on iterative cycles of data analysis and computerised (in silico) model construction/refinement and predictions, linked to wet-lab (in vitro) and living specimen (in vivo) experimental design, experimentation, and data capture and storage in forms that can be represented and manipulated by computer software.

Conclusion: Systems Biology offers tremendous potential for future medical and genomics research, and should play a major role in future research programmes.

Key Areas for Systems Biology (SB)

GENERAL SYSTEMS BIOLOGY APPROACHES AND TOOLS

1) **SB within other projects:** Proposals that claim to have a systems approach should contain a strong analysis and modelling component. Systems biology related projects should assign an adequate fraction of resources (10-50%) for data modelling or integration, depending on the scope of the project. Even smaller projects need significant data integration and computer modelling resources, going beyond the basic data analysis needed for traditional functional genomics. This may encompass projects not only with medical/health goals but also projects in biotechnology such as genetic or physiological engineering of micro organisms or plants or process improvements.
2) **Areas for analysis**: Excellent reviews[2] describe SB and its anticipated development, e.g. covering *in silico* models of biological systems which provide a powerful tool for integrative analysis of physiological function. Using the computational models of the heart as examples, three types of integration are discussed: structural integration, functional integration and synthesis. SB covers (at least) three orthogonal types of integration and scale of analysis:
   a) structural integration implies integration across physical scales of biological organisation from protein molecule to whole organ (gene-protein-macromolecular complex-organelle-cell-network-tissue-organ-system-organism);
   b) functional integration of interacting physiological processes such as signalling, metabolism, excitation and contraction (regulatory-growth-metabolic-electrical-mechanical-transport);
   c) synthesis of experimental observation with physicochemical and mathematical principles (empirical data-ontologies-statistical modelling-systems analysis-predictive modelling-physico-chemical first principles-mathematical theory).

   Other dimensions exist, for example the entire reproductive and developmental processes of organisms, evolution and diversity at the population level, cross organism interaction (disease interactions and transmission, food chain transmission, development of bacterial resistance, mutants of HIV, new influenza variants, etc.).
3) **Types of Analysis:** Just as there are several dimensions in analysis, so are there several dimensions in types of output desired, for example the different levels in the paradigm of doing science: Explain, Discover, Predict, Control:
   a) Explain:
      i) Simulations to show that data are internally self-consistent and sufficient for testing hypotheses
      ii) Functional genomics analysis of expression data to study systems under normal conditions
      iii) Test hypothesis of how pathways operate
      iv) Use model-assisted data-mining to increase understanding
      v) Elucidation of properties and (conserved) rules according to which biological systems operate
   b) Discover
      i) Study models of systems in silico in order to discover laws and principles for the behaviour of biological systems
      ii) Use analytical mathematics to prove those laws and principles
      iii) Study experimental systems so as to assert the relevance of the laws and principles
   c) Predict:
      i) Design wet-lab experiments to explore certain hypotheses
      ii) Perturb computational systems in various ways to see how the system would react to different stimuli, either minor via concentration changes, or major, e.g. by knock-out of a whole gene
      iii) Simulate and experimentally test the effects of new stimuli (drugs) or other perturbations on a physiological or cellular system, predict the response of patients to different types of treatments.
   d) Control:
      i) Use models to design new stimuli (drugs) or other perturbations to achieve desired effects in physiological or cellular systems.
      ii) Control the generation of new data to be able to decide between models, and/or to improve parameter estimates, based on automated analysis of alternative models.
      iii) Develop a range of industrial applications for controlled biological systems.

---

[2] *Integrative biological modelling in silico* by Andrew D. McCulloch and Gary Huber. http://www.novartisfound.org.uk/catalog/247abs.htm#McCulloch. See also **Computational Cell Biology**, C.P. Fall et al, ed., ©Springer-Verlag 2002 http://www.compcell.appstate.edu and Computational Systems Biology workshop report, March 2004, ftp://ftp.cordis.lu/pub/lifescihealth/docs/csbworkshop_2004_03_en.pdf .

4) **Methodologies**: Many of the successful models are oriented to answering particular questions. Up to now, no model incorporates all data, solves everything, or is available to answer all questions. Many successful models in fact have the following characteristics:
   a) They are strongly coupled with experimental work occurring in conjunction with the modelling.
   b) They tend to ask one question and often focus on the time dependent behaviour of one parameter, even though the modelling may be much more complex.
   c) Many successful models in the literature solve between two and ten differential equations, and often involve arbitrarily set parameters. This does however involve enormous simplifications, since real biological networks are often much more complex. The availability of both much larger data sets generated as parts of functional genomics projects and – perhaps more significantly – far more accurate measurements of specific processes, as well as new modelling techniques, could very well change this situation, and allow modelling to proceed on a more realistic scale.
   d) There are very successful modelling alternatives to use of differential equations, for example Boolean networks which are not as detailed, but which can answer many useful questions.
   e) Multiple level modelling is possible and successful, but the modelling at each level is strongly focussed on the particular problem to be solved.
   f) Successful modelling depends on having a wide range of biological data available in a consistent and quantitative form, although only a small fraction of it may be used in a particular model.
   g) More and more modelling is done using standard platforms for programming such as SBML, powered by standard databases such as REACTOME and KEGG. However, the data and the choice of modelling are often influenced by a coupled experimental programme.
   h) Modelling also needs to invoke complex systems theory. Biological systems operate far from thermal equilibrium, and display self-sustained oscillations, pulses and burst at all different levels of organization. The development of bifurcation theory has for the first time provided us with the mathematical tools and concepts needed to deal with such problems. Yet, many mathematical problems in this field still remain unsolved. We do not understand for instance, how a simple model of a bursting and spiking cell can change from one type of dynamics to the other. The interaction of many oscillating subsystems (cells) with slightly different parameters is also a matter of great importance.
   i) For SB consortia, it is absolutely necessary to work under standardised conditions, especially if more than one team or project is employing the same model organism. Working with the same standard operation procedures (SOPs) will ensure comparable results; experimental conditions as well as data analysis and presentation have to be well documented since incomplete information is useless. Consortia should focus on well defined biological questions and tackle those in one or a limited number of different (model) organisms to ensure coherent results of wide impact. Since SB is an interdisciplinary research area modelling efforts should include experimental work (wet lab) and methodological development.
   j) Models should make some attempt to connect with functional genomics, in the sense of ultimately contributing to understand the entire living organism
5) **SB Analysis Packages and various types of Modelling**: There are clear needs for mathematical modelling, concepts, principles and developing new methodologies for system identification, parameter estimation, spatio-temporal modelling. This also includes the combination of dynamic pathway models, formal analysis of the role of feedback in biochemical networks, their robustness and sensitivity as well as hybrid approaches to model the coordination cell function. Standardisation is necessary at the level of the networks and components modelled, model description (including reaction specification, measurement units, etc.), data storage and retrieval, and the computer codes. The last takes a particular importance because models developed by different networks/groups represent modules of cellular operation that should be compatible with each other. Hence, a priority is the development and use of multi-platform, non-proprietary programming languages, such as SBML, with professional standards for software production and maintenance, and for web accessible live models (such as in the silicon cell). The ultimate goals are modular combinations of models and routine applications of ‘standard’ models in non-specialist (experimental) labs.

UNDERPINNING BIOINFORMATICS AND EXPERIMENTAL DATA SUPPORT

6) **Tools and Services: Bioinformatics, Databases, Software, Access, Services, Research and Infrastructures as the basis for systems biology:** It is recognized and emphasized that the full range of bioinformatics tools is an absolutely vital underpinning of all systems biology work, and bioinformatics analysis is not only essential in its own right, it also in many cases merges smoothly with SB, depending on the nature of the problem. However, this report tends to concentrate on particular needs for developing SB capabilities and using a SB approach, and tends to place a special emphasis on those aspects of bioinformatics and other relevant research and infrastructure needed for SB. Dedicated infrastructures and supporting research to generate the data and resources are needed in the areas of bioinformatics, biobanks, high throughput, structure, imaging and clinical facilities.

   a) **Model systems and Biobank Resources:** Research projects should start with well characterised model systems with corresponding biobank resources, at the single cellular level, while linking these to multi-cellular model organisms and human and relevant biobank and tissue resources to develop aspects of health research. Potential single cell model systems to analyse include: *S. cerevisiae* and *Sz. pombe* (yeasts); *B. subtilis*; *E. coli*; *L. lactis*, filamentous fungi. Multicellular model organisms could include any of the standard model organisms, depending on data available, plus the human cells relevant to particular health aspects; for example: mouse, rat, zebrafish, worm, *Arabidopsis*, mosquito, fly and various human cells incl. neurons, hepatocytes, heart. Organisms are sometimes chosen as having special contributions to either fundamental knowledge or to specific health or biotechnological applications, but are often highly relevant to both. The analysis of human materials (e.g. tumour samples from cancer patients) could help to provide an early focus of the analysis of specific disease processes.

   b) **Standards and Ontologies**: Standards needed include those for data collection from experiments, storage in databases, 'modelbases' and analysis, consistent with modelling requirements, including dynamic data where possible. This process is already underway with a wide range of data at the bioinformatics level, and needs to be extended to make the data useful for SB analysis, making full use of standard ontologies and controlled vocabularies and developing new ones where appropriate. Standards need to evolve with advances in SB. There is a need to maximise transfer capabilities.

   c) **Obtaining data within and beyond present 'omics' (genomics, etc.) , maintaining and expanding current databases:** Although experimental biological data should be collected with bioinformatics and SB analysis in mind, there are key types of quantitative data becoming available which especially support SB, and which require special attention for standardisation and analysis. Current bioinformatics databases also need to be supported in order to make them compatible with the new SB information needs and expanded to take account of new information in these areas, including key areas for bioinformatics-based research and databases. In addition to extensive high throughput data, there is also a need for accurate quantitative measurements in key areas to provide constraints (or very few constraints) to actual dynamic models.

      The key data areas include:

      i) Gene expression , transcription, post-transcriptional control, regulatory RNAs
      ii) Protein-protein, protein-nucleic acid and protein-metabolite interactions
      iii) Protein modification
      iv) Kinetics and non-equilibrium thermodynamics
      v) Genetic analysis and mutations
      vi) Comparative genomics
      vii) Metabolic flux analysis
      viii) Control Analysis
      ix) In vivo imaging
      x) Haplotypes
      xi) Protein homology
      xii) Protein identification
      xiii) Protein structure
      xiv) Toponomics
      xv) Immunology
      xvi) Molecular concentrations/states (e.g., phosphorylation read-outs etc)

TOPICAL AREAS FOR SYSTEMS BIOLOGY

Developing systems biology

7) **Cellular and sub-cellular systems:** In addition to incorporating a bioinformatics and Computational Systems Biology (CSB) approach in most experimental systems biology projects, a number of dedicated projects could be established to build up European SB capabilities at the cellular and sub-cellular level. There should be no proposals that aim for either models or data, SB implies data driven modelling to design new data production experiments [no models without data and no data without models].
   a) Research projects should focus on (i) modelling at least one process comprehensively, which is already a challenge, and where appropriate, by using levels of complexity to treat interactions, (ii) integrated modelling of several cellular processes leading to as complete an understanding as possible of the dynamic behaviour of a cell or a tissue. Several projects may be required to develop modules (metabolism, signalling, trafficking, organelles, cell cycle, gene expression, replication, cytoskeleton) in model organisms. This modelling should involve realistic analysis of experimental data, including a wide range of data for transcriptomics, proteomics and functional genomics, and interactions with cellular pathways including signal transduction, regulatory cascades, metabolic pathways etc. It should further involve generation and analysis of:
      i) Coherent, high-quality, quantitative, heterogeneous and dynamic experimental data sets as a basis for novel model constructions to advance from descriptive to predictive modelling.
      ii) The results from experimental functional analysis tools (in-situ proteomics, protein-protein interactions, metabolic fluxes, etc.)
      iii) Normal and diseased (perturbed) states, physiology and pathophysiology and their mechanisms and progression

8) **Efficient large-scale study of multiple interacting systems at the cellular and physiological levels**: Most systems biology work is currently limited to specific levels of modelling and data integration. There is a need for improved technologies and methods for large-scale modelling, and current projects provide the basis for this advance.
   a) The applications will be wide reaching, e.g. to develop the cellular level modelling to couple to the type of physiological modelling occurring in the Network of Excellence BioSim, which directly aims at the drug design process. Understanding the hierarchical relationships will clearly allow a greatly improved description of function.
   b) Several international efforts are being launched, the metabolome, regulome, transcriptome, etc., to address functions at higher levels by mapping the nodes and networks involved in cellular and biological functions. These efforts will allow European laboratories to structure their efforts and contribute significantly to these international initiatives.
   c) This research will demand a multidisciplinary approach linked to experimental work and data collection from high throughput and emerging high precision technologies, including array technologies, proteomics, molecular biology, bioimaging and genetics of model organisms. The EU project MolTools is an example of fostering novel enabling wet-lab technologies. A strong experimental component closely linked to the bioinformatics and computational systems biology component will be required for analysing and integrating the data collected. This multidisciplinarity can only be accommodated at the European level.

9) **Physiology:** A key goal of health research is to simulate both the healthy state of human systems and the role of a disease mechanism and defence mechanisms, including the immune system. All these programmes will require resources and produce outputs that could be common to all of them, including databases, software tools, cellular and physiological models. The goal is understanding the functioning of physiological and pathological systems, involving the following:
   i) Sequence, structure, function, etc. bioinformatics databases, tools, research to support SB
   ii) Comprehensive model of a single cell (see e.g. German SB hepatocyte program)
   iii) Comprehensive model of physiology systems linking to human cells
   iv) Neuroinformatics
   v) Modelling of evolving state of organisms and people, involving developmental biology, ageing, mutations, circadian rhythms, etc.

Applying systems biology

10) **Disease, Medicines and Treatment**: Closely following the understanding of physiology is the need to model the effects on pathophysiology and diseases. In particular, SB will support the development and application of the medicines and various forms of treatment to deal with them. The fields of pharmacokinetics and pharmacodynamics have long been established, but need to be taken to a much higher level of sophistication and simulation ability. In the cases of complex diseases such as the various types of cancer, it is essential to develop new strategies, based on the application of high throughput techniques from functional genomics to acquire information on most or all genes and gene products involved in the disease process, as well as in the response of the entire organism to any possible treatment. Analysing the very large quantity of information combining clinical, experimental and computational inputs requires the use of whole genome modelling to be able to translate information into predictions of the effects of different therapeutic schedules and drugs that are adapted to different specific genetic and physiological backgrounds. SB could also potentially be applied to develop novel treatments and disease prevention regimes that reduce or circumvent the use of drugs. Areas include:

a) Rationalised drug development and modelling; network-based drug design

b) Improved modelling of effects, optimum dose and optimum timing of use of medicines, including personalized patient and group data where possible and appropriate.

c) Modelling of side effects of medicines, by modelling a wide range of systems in addition to the system targeted by the drug

d) Modelling of altered nutrition and other life style changes to treat or prevent pathological phenotypes

e) SB of complex, multifactorial diseases, e.g. varieties of cancer

f) Modelling of immune system, evolution of viral and bacterial resistance to medicines, interaction between organisms (internal ecology)

g) Dynamical diseases (physiological disorders involving an abrupt switch to altered modes of dynamic behaviour)

h) Public health relevant systems

11) **Biotechnology**: SB will be developed and applied in different areas such as the engineering or breeding of industrially important micro organisms and plants. Today, 25% of all medicines are plant derived, and the spectrum of medicinal feedstocks and efficiency of production can still be greatly enhanced. Expertise developed in these areas may be beneficial for SB in medicine and vice versa and relevant coordination will be important. SB also holds great potential to foster sustainable development, by accelerating and rationalising the production of plant-derived biofuels and chemical feedstocks in preparation of the inevitable depletion of fossil carbon.

12) **New Applications**: New applications of SB are being developed, such as within Synthetic Biology and systems design. SB will operate here at the interface with nanotechnology. In FP6, the 'new and emerging science and technology' (NEST) programme has been used as an instrument to develop such emerging technologies.

Training

13) **Training**: Training in bioinformatics and SB should be supported. This could be achieved by thematic calls within the Marie Curie programme, or other actions teaming up with European organisations (such as FEBS, EMBO, ESF).
   a) Existing M.Sc. and Doctoral courses and collaborations between universities should be extended
   b) Best practices and course materials should be produced and exchanged in a more systematic fashion
   c) Special training infrastructures and networks could be established.

CONTRIBUTED PAPERS / PROJECT SUMMARIES
(in order of presentation at workshop)

DIAMONDS - Dedicated Integration And Modelling Of Novel Data and prior knowledge to enable Systems biology

Martin Kuiper, Plant Systems Biology, RUG-VIB, Gent, Belgium

The DIAMONDS project aims to demonstrate the power of a Systems Biology approach to study the regulatory network structure of the most fundamental biological process in eukaryotes: the cell cycle. An integrative approach will be applied to build a basic model of the cell cycle, in four different species including *S. cerevisiae* (budding yeast), *S. pombe* (fission yeast), *A. thaliana* (weed, model plant) and human cells. To do this, a Consortium is assembled of leaders in the fields of cell cycle biology, functional genomics technologies, database design and development, data analysis and integration technology, as well as modelling and simulation approaches. The project combines a number of complementary data sets, toward an advanced mining and modelling environment designed to assist the biologist in building and amending hypotheses, and to help the investigator when designing new experiments to challenge these hypotheses. By doing this simultaneously in widely different organisms we will ensure that the tools are generally applicable across species. By bringing together in the design phase biologists, bioinformaticians, bio-mathematicians and (commercial) software developers we will ensure that a user-friendly, intuitive data analysis environment is created. The main data streams generated *de novo* within the project concern transcript profiling and proteomics data (Y2H and TAP). These data will be complemented with information extracted through comparative genomics, and prior knowledge coming from literature mining (text mining tools). The project will bring together a number of existing technologies to build a knowledge warehouse in a relational database designed to contain cell cycle regulatory network information, accessible through an intuitive user platform (GUI) with embedded modelling tools. This platform will enable both top-down and bottom-up hypothesis-driven research, and will serve as a basis to develop more rigorous dynamical models for cell cycle variants.

COSBICS - Computational Systems Biology of Cell Signalling

Olaf Wolkenhauer, University of Rostock, Germany

Cancer can be considered a failure of communication at molecular level. The area of cell signalling investigates the communication among cells and the transmission of information from receptors to the activation of genes. These processes are conceptualised through biochemical reaction networks (pathways). COSBICS takes a systems biology approach by treating these processes as inherently dynamic phenomena.

COSBICS is to focus on the dynamic mathematical modelling of two cancer related signalling systems: the Ras/Raf/MEK/ERK pathway and the JAK-STAT pathway. COSBICS is to investigate with these pathways the heart of the intracellular communication network that governs cell growth, differentiation and survival.

While the biology of this project is centred around these two particular systems, the theoretical side of this project concentrates on the development of generic methodologies that are applicable to pathways in general. These include investigations into the role of feedback loops and protein translocation on the dynamic properties of the system under consideration. The project investigates data-driven approaches to extract model parameter values from quantitative time series data and concepts for the design of experiments.

More details on this project are available from www.sbi.uni-rostock.de

EMI-CD - European Modelling Initiative Combating Complex Diseases

Hans Lehrach, Max Planck Institute for Molecular Genetics, Berlin, Germany

Primary targets of the Sixth Framework programme are activities for the development of advanced genomics and bioinformatics tools, their applications for health and for the combat of multigenic complex diseases. The understanding of biological processes relevant for these diseases and the modelling of these processes in silico is an essential step in the development of new drugs, medical diagnostics and therapies. Analytically, the modelling of disease processes is an integrative and multi disciplinary approach that has to cope with data from diverse functional genomics platforms such as gene expression and protein expression data, functional sequence data, physiological data, environmental factors and many others. The aim of EMI-CD is to provide an adequate modelling and data integration platform that is able to manage the complexity and heterogeneity of these underlying data sources, and to generate hypotheses and models for disease processes based on in silico predictions. Modules of EMI-CD are:

1. Database integration. The platform will have an interface to the SRS annotation system (LION Bioscience Ltd.) and will incorporate inference of knowledge on given pathways, in particular the Reactome database developed in the EBI-ENSEMBL group.

2. Experimental data integration. A critical issue is the integration of experimental data. (MicroDiscovery GmbH Berlin). Data from different techniques will be integrated, correlated and fed into the modelling platform.

3. Data analysis. Advanced approaches for data analysis rely on mathematical models of the regulation process. The starting point here will be a pathway core that represents prior knowledge on a particular system. Combinatorial search algorithms are then used for the computation of core expansions in the light of their level of fitness to given experimental data. This expansion methodology is contained in the software developed by the Computer Science group of the Tel-Aviv University and will be connected to the system.

4. Modelling and simulation. At the Max-Planck Institute for Molecular Genetics we developed and implemented the forward modelling system PyBioS for simulation of complex biological systems. The system allows the incorporation of metabolic networks and signalling pathways as well as the import of gene regulatory networks. In the course of the project we will introduce a library of standardized kinetic models and further analysis tools.

COMBIO - An integrative approach to cellular signalling and control processes: Bringing computational biology to the bench.

Andrea Ciliberto, Technical University of Budapest, Hungary

It is increasingly being recognised that progress of modern day biology will require the understanding and harnessing the network of interactions between genes and proteins and the functional systems that they generate. Given the complexity of even the most primitive living organism, and our still very limited knowledge, it is unreasonable to expect that we might, in the near or even medium term, reach such understanding the level of an entire cell. However, significant progress towards a system-level understanding should be achievable by applying an integrated approach to the analysis of a set of well defined and biologically important cellular processes.

It is not our goal here to come out with a new software package, or to simulate a whole cell. Rather our project aims to bring computer models and simulations to the experimental community. To do so we will focus on two systems that involve different aspects of Biological systems: Networks and Self-organization and we will apply different simulation approximations to both of them. This will enable us both to identify the modelling and simulation strategies that are better suited for a particular experimental problem.

The present 3-year project combines a unique group of experimentalists, bioinformaticians and modellers in order to gain detailed understanding of two key processes: the P53-MDM2 regulatory network and the self organisation process whereby chromatin controls microtubule nucleation and organization. A major objective will be to benchmark the ability of current modelling and simulation methods to generate useful hypothesis for experimentalists and to provide new insights into biological processes of realistic complexity. The expected result will be a set of guidelines, specifying which and how simulation methods should be used, given the problem at hand. These guidelines will also indicate how best simulations and experimental procedures might be combined to answer key questions about biological function. This should make a fundamental contribution to fill the gap between molecular biology and cell physiology, and provide ways for elucidating the mechanisms of action of pharmacological compounds.

QUASI - Quantifying signal transduction

Stefan Hohmann, Göteborg University, Sweden

The present understanding of cellular signal transduction is restricted, at the best, to the wiring schemes of signalling pathways. Little is known about the details of their dynamic operation and the importance of quantitative, spatial and time-dependent parameters for signalling output. Those are, however, crucially important for drug discovery and application. QUASI is a multidisciplinary project with the goal to obtain a coherent and detailed picture of the dynamic operation of a model signalling transduction network. The signalling pathways contain the evolutionarily conserved MAP kinase cascade module, which is of central importance for signalling in human cells and implicated in human diseases such as cancer and inflammatory disorders. A better understanding of the dynamic operation of these pathways offers new opportunities for drug discovery and for efficient individualised treatment based on the genetic setup of the patient (pharmacogenomics).
To achieve the goals of QUASI, quantitative data of high definition on signal transduction activation and deactivation are being obtained using frontline experimental approaches encompassing global gene expression, proteomics, bioimaging and chemical genetics. Software-implemented mathematical models of signalling dynamics have already been constructed from pre-existing data during the course of the first year. Predictions from the model will be used as a basis for experimentation to further enhance models in a recursive manner. An interactive dynamic visual interface is being constructed to allow the experimenter to explore the effects of virtual manipulations of the system. This tool will enhance the intuitive understanding of intracellular signalling and this interface could be developed into a general tool for education as well as prediction of drug effects.
QUASI clearly sees possible benefits and synergies from collaborating with other projects such as COSBICS and EMI-CD (and possibly others) and has taken contact with these projects to develop common activities. Related to this, several different project proposals (CA, RTN, STREP) have recently been submitted.

EUSYSBIO: The Take off of European Systems Biology (SB)

Gisela Miczka, Projectmanagement Organization Jülich, Germany

The specific support action EUSYSBIO is an activity within Priority 1 of FP6. The goal of EUSYSBIO is to identify and bundle existing strengths and competences in European Systems Biology. The project started in November 2003 and runs until October 2005.

Nine participants (one as associated partner) out of five countries work on nine work packages. Work packages are designed to identify existing strengths and weaknesses in the field of SB in Europe, Asia and America. As a second step ongoing SB activities shall be coordinated. EUSYSBIO will thus lay the foundation for the successful start of European SB within the FP6 and FP7. In addition, the SSA focuses on training of young scientists, on the international networking of activities and players as well as establishing a platform for contacts between science, industry and public.

EUSYSBIO will form the nucleus of further European activities and prepare the start of a powerful pan-European research initiative in SB. EUSYSBIO has already catalyzed three scientific activities, i.e. the 5th International Conference for Systems Biology (October 2004, Heidelberg), the first Advanced Course on Systems Biology (March 2005, Gosau) and a book defining Systems Biology (to be published in 2005).

As an important result of the SSA a critical mass of national SB programmes has been assembled to apply for a Coordination action in the framework of the ERA-NET programme. The proposal for the ERA-NET called ERASYSBIO is in preparation. Submission is planned in March 2005.

BIOSAPIENS (1) - A European Network for Integrated Genome Annotation

Ewan Birney, EMBL Outstation, European Bioinformatics Institute, Hinxton, United Kingdom

The objective of the BIOSAPIENS Network of Excellence is to provide an infrastructure to support a large scale, concerted effort to annotate genome data by laboratories distributed around Europe. This will use both informatics tools and input from experimentalists. Experimental validation of a statistically significant subset of the predictions will be an integral part of the process, leading to an iterative improvement in methods. The Network will bring together many of the best laboratories to create a European Virtual Institute for Genome Annotation, divided into nodes, each focussed on one aspect of genome annotation. Through integration the institute will help to improve bioinformatics research in Europe, by providing a focus for annotation and by the organisation of European meetings and workshops to encourage cooperation, rather than duplication of effort. It will also be pro-active in forging closer integration between the experimentalists and bioinformaticians, through a directed programme of genome analysis, focused on specific biological problems.

The annotations generated by the Institute will be available in the public domain and easily accessible through a single portal on the web. This will be achieved through a distributed annotation system (DAS), which will evolve to take advantage of new developments in the GRID. The BIOSAPIENS Network of Excellence will increase European competitiveness, especially for Small and Medium Enterprises (SME's), by new discoveries, increased integration, expert training and improved tools and services.

The Institute will establish a permanent European School of Bioinformatics, to train bioinformaticians and to encourage best practise in the exploitation of genome annotation data for biologists throughout Europe. In summary the Institute will further a European Research Area for Bioinformatics, enhancing Europe's role in the academic and industrial exploitation of genomics.

BIOSAPIENS (2) - Reactome, A systems biology ready database

Ewan Birney, EMBL Outstation, European Bioinformatics Institute, Hinxton, United Kingdom

Systems biology focuses on developing useful models of biological systems. One common biological system to study are molecular pathways, i.e., a set of molecular components and their interactions inside a cell. These range from small molecule metabolism, e.g., glycolysis, through the housekeeping functions in the cell, e.g. DNA repair, to cell-specific signalling, e.g. Insulin receptor signalling. There are two main levels in understanding these pathways. The first is the qualitative content of which molecular entities (i.e., which proteins and small molecules) make up the pathway and what is the topology of the interactions between them. The second is the quantitative dynamic modelling of the pathway providing a sensible model of how the pathway evolves from a starting state. This second level is necessarily dependent on the first, and it is the first level, i.e., the qualitative network, which is the focus on pathways databases.

Currently worldwide there is a paucity of pathway information at both levels. The most established pathway database, KEGG, has a number of limitations, including an emphasis on small molecule metabolism to the detriment of signalling pathways and an emphasis on bacterial and plant systems. It also does not track the precise molecular species involved in a reaction (e.g., the phosphorylation state) and has limited tracking of the primary literature. In contrast, Reactome (www.reactome.org) is a focused pathway resource on mainly human pathways with its scope to cover all pathways, from signalling through to metabolism. Reactome can export Systems Biology Markup Language (SBML), providing a natural bridge from this qualitative input through to the more quantitative modelling systems. Reactome currently has just shy of 1,000 proteins participating in around 1,500 reactions, including DNA repair, Cell Cycle, Apoptosis, Haemostasis and Carbon metabolism. All of Reactome is backed by primary literature.

Reactome is useful not only for systems modelling but also for the interpretation of large scale datasets. One early feature implemented is the ability to “paint” the large pathway network of Reactome with numerical values from high throughput sources, e.g., expression datasets or evolutionary data. The Figure shown on the reactome webpage shows one such overlay of the human/mouse dn/ds ratio (a measure of the amount of negative selection on the gene) on the Reactome pathway map.

BIOSIM (1) - Biosimulation - A New Tool in Drug Development

Erik Mosekilde, Technical University of Denmark, Denmark

International competition in the pharmaceutical industry is increasingly becoming a competition with respect to the ability to understand complex biological processes and exploit the rapidly growing amount of biological information. The methods that are currently applied in the development of new medicines suffer from the lack of effective means to evaluate, combine, and accumulate biological knowledge. Essential improvements must involve the use of computational models that can provide a dynamic and more quantitative description of the relevant biological, pathological and pharmacokinetic processes.

BioSim is a Network of Excellence that can restructure and strengthen the area of biosimulation by focusing on the development of professional, physiologically-based drugs at significantly lower costs. The modelling approach is strongly recommended by the American Food and Drug Administration that already uses mathematical models in its evaluation of applications for drug approval. Academic institutions in Europe have significant expertise in biological modelling, and several groups are individually at the research front in their specific areas. At the present, however, the research is strongly fragmented, and the industry itself has relatively few qualified experts in the field. The Network will provide a new forum for collaboration across disciplinary boundaries as well as between industry, regulatory authorities, and academia.

BIOSIM (2) - Biosimulation - A New Tool in Drug Development

Computational Systems Biology of the mammalian circadian clock: From molecular mechanism to disorders of the sleep-wake cycle

Albert Goldbeter, Unité de Chronobiologie théorique, Université Libre de Bruxelles, Belgium

Cellular rhythms have represented for long a field of choice for modelling studies in Systems Biology. The reason why there are so many rhythms in biological systems is primarily due to the multiple modes of regulation which control the temporal organization of cellular processes. Thus, regulation of ion channels through the dependence of their conductance on the membrane potential underlies the excitable and rhythmic properties of nerve and cardiac cells, regulation of enzymes underlies metabolic oscillations such as glycolytic oscillations in yeast, receptor regulation is at the core of cyclic AMP oscillations in Dictyostelium amoebae, regulation of intracellular transport processes underlies cytosolic Calcium oscillations in a variety of cell types, while regulation of gene expression is involved in the generation of circadian rhythms. Computational models closely related to experimental observations have become an indispensable tool for the study of all these oscillatory phenomena.

The model for the mammalian circadian clock, based on molecular regulatory mechanisms, can account for some physiological disorders observed at the level of an organism. The model for the mammalian clock thus provides a useful tool for probing the origin of physiological disorders in terms of changes in network properties, resulting from gene mutations.

BIOSIM teams will collaborate in the experimental and theoretical investigation of the effects of circadian rhythms on anticancer drug delivery. The aim of this project will be to use Computational Systems Biology to optimize the temporal patterns of anticancer drug delivery. This issue more generally pertains to Chronopharmacology, which studies the effect of the time of administration of a drug on the physiological efficiency of the pharmacological treatment. In carrying out this investigation we will focus on the recently uncovered links between circadian rhythms and cell proliferation.

BIOSIM (3) - Biosimulation – A New Tool in Drug Development

Morten Colding-Jørgensen, Novo Nordisk A/S, Denmark

In Novo Nordisk, biosimulation has been used with increasing success over a decade. The most successful type of modelling has been a mixture of top-down and bottom-up modelling. The idea is to tailor-fit the modelling to the problem in question. This requires limited modelling efforts but extensive insight in the mechanisms behind the studied behaviour. With this method, we have been able to reveal unknown physiological and pathological mechanisms and drug targets.

At present, we are mostly using what we call "virtual experiments". These are simple models designed to mimic a given experiment type. The results from both are analyzed and new real and/or virtual experiments are designed and performed. The final outcome is a clear picture of the mechanisms and variability of the study.

We have also begun to use "virtual patients", where model parameters are varied, so they mimic different patient types. This helps us in dose finding, possible side effects, demographic differences etc.

A future application is progression of diseases. By changing model parameters with time, it is possible to mimic the dynamics of the disease from start to cure (or death). This enables us to study the different paths the disease can take and to find treatments that can influence these paths and help the patients.

BIOSIM (4) - Biosimulation - A New Tool in Drug Development

Hans Westerhoff, Free University of Amsterdam, The Netherlands

See http://www.systembiology.net and http://www.bio.vu.nl/hwconf

Traditionally, mathematical modelling of intracellular phenomena has remained limited to one particular level of intracellular organization. BIOSIM will explicitly connect intracellular levels, connecting metabolic pathways, with gene-expression analysis and with signal transduction. It will also connect phenomena occurring at different scales in terms of length, time and chemical dimensions, yet together contributing to cell function. BIOSIM will provide a forum where groups working at any of the individual levels will not only meet and discuss but also actively integrate their models.

In the light (and times) of functional genomics we will have increasing and eventually almost complete access to all components of the living cell. The molecular information that is collected is vast, much too much to be connected, certainly not by the human mind, but not even by computer models. It is of great importance therefore to realize that the great majority of the information, though perhaps of great scientific interest, is irrelevant to understand biological function at the level of the cell or organism. The complete crystallographic structures of all cellular proteins, for instance, are largely irrelevant once the much fewer number of kinetic characteristics of those proteins are known: It is of great importance selectively to use all the information that is essential to understand the living cell and the organisms made of such cells. An important way to do this is to make so-called silicon pathways, i.e. computer replicas of pathways inside living cells, based on the experimental kinetic data. Connecting such silicon pathways will then lead to silicon cells. The latter can be used for systems biology *in silico*, i.e. to discover principles and laws of biological systems, to do network-based drug design, to validate drug leads, and to search for possible side effects of drugs, e.g. in the setting of Food and Drug Administration.

Examples were given in which (i) two new drug targets against trypanosomes have been found, (ii) the differential importance of kinases and phosphatases for cell differentiation and growth has been understood in terms of new laws for signal transduction, and (iii) the possible basis for some genes being (anti)oncogenes was computed using the silicon cell approach (cf. www.siliconcell.net). Obviously no single group can provide all the necessary data and computation. BIOSIM will connect its component groups such that the minimum critical size is reached.