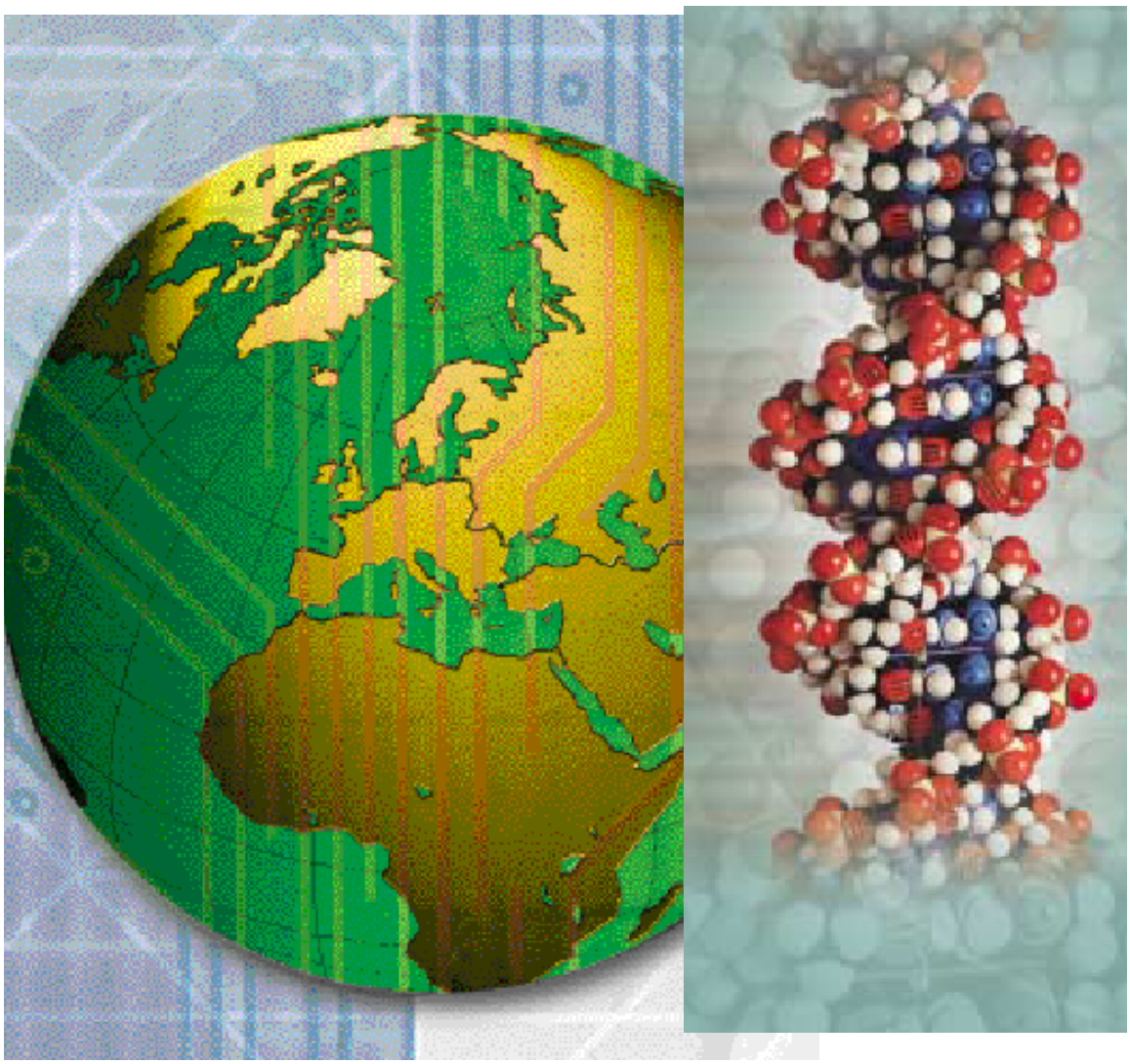
Life sciences, genomics and biotechnology for health

Computational Systems Biology (CSB) - Its future in Europe



8 March 2004

DG Research / F.4 Fundamental Genomics

Contact: Frederick Marcus or Bernard Mulligan, Research Directorate General, European Commission
mailing address: EUROPEAN COMMISSION (SDME 8/57), B-1049 Brussels, Belgium
e-mail: Frederick.Marcus@cec.eu.int Bernard.Mulligan@cec.eu.int
website: http://www.cordis.lu/lifescihealth/genomics/home.htm and http://www.cordis.lu/lifescihealth

Published by the EUROPEAN COMMISSION

Research Directorate-General

Workshop Report on

Computational Systems Biology (CSB) - Its future in Europe

for the
European Commission
Research Directorate General
Directorate F - Health Research

Based upon a Workshop
held in Brussels, Belgium on 10-11 September 2003

Lilia Alberghina, Steven Foord, Jaap Heringa, Ralf Herwig (chair), Stefan Hohmann, Norbert Hübner, Karl Kuchler, Frank Laplace, Nicolas Le Novère, Cedric Notredame, Mark Sansom (rapporteur), Uwe Sauer, William Saurin, Luis Serrano (chair), Age Smilde, Alfonso Valencia, Eero Vuorio, Hans Westerhoff (chair), Shoshana Wodak, Petra Wolff, Olaf Wolkenhauer

Editors: Frederick Marcus, Bernard Mulligan, Mark Sansom

8 MARCH 2004

TABLE OF CONTENTS

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 rapporteur and chairpersons. They and the other 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.

EXECUTIVE SUMMARY

The aim of this document is to define a research and policy agenda to promote the field of computational systems biology (CSB) in Europe, which would involve better structuring of European research and infrastructure support, and defining research areas. It is primarily aimed at European Commission research policy for building a European Research Area, and also can provide insights to organisations in Member states and researchers in the public and private sectors.

A group of European experts in computational systems biology attended a workshop organised by the services of the European Commission, entitled "Computational Systems Biology (CSB) - Its future in Europe." This document is its outcome.

CSB involves developing understanding of the interaction of components of biological systems, and the expression of this understanding in qualitative and quantitative terms - in particular, in terms amenable to electronic storage and communication. The most successful current implementations of CSB 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.

Eight areas for action are considered, short- and long-term, which can be implemented in the EU framework programme for research, by national funding organisations and by the researchers themselves:

1. Fragmented research in Europe
2. General Modelling requirements
3. Model organisms as data sources
4. Standardisation of *in vitro/ in vivo* experiments and their data
5. Standardisation of databases, software and modelling
6. Data required beyond present ‘omics (genomics, proteomics, metabolomics, etc.)
7. Training
8. Increasing excellence in experimental research projects via bioinformatics and CSB

1) Fragmented research in Europe : Research in Europe is conducted in multiple locations, with little coordination and inadequate funding. This may be due to the breadth of the field, and to a lack of available funding compared to the USA and Japan.

Substantial EU funding tied to structuring the field, in combination with national funding efforts, is necessary to foster computational systems biology in Europe. Preference should be given to strategic projects that provide both conceptual computational and experimental advances together, using relevant model systems.

2) General Modelling requirements: A major post-genomic challenge to advance from genomic sequence to a complete understanding of gene function and biological processes. Today the US and Asia are much more active than Europe in modelling and simulation of complex processes. A key priority is the development of Europe-wide initiatives to create and integrate relevant databases and analysis software, thus enabling systems-level interpretation of complex experimental data in functional genomics. The ideal situation seems to be one where the modelling development is slightly visionary beyond the current state of the art, but at the same time firmly anchored in experiments and their comprehensive data sets.

Research projects should focus on integrated modelling of several cellular processes leading to as complete an understanding as possible of the dynamic behaviour of a cell. 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 involve:

- *Coherent, high-quality, quantitative, heterogeneous and dynamic data sets as a basis for novel model constructions to advance from analytical to predictive modelling.*
- *Experimental functional analysis tools (in-situ proteomics, protein-protein interactions, metabolic fluxes, etc)*

3) Model organisms as data sources : At this stage of development of CSB, results from all biological systems and model organisms are relevant, but some model organisms have been more completely studied than others, and are more able to provide the full range of data needed for modelling.

Research projects should start with well characterised model systems at the single cellular level, while linking these to multi-cellular model organisms and man to develop aspects of health research : Potential single cell model systems to analyse include:

- *S. cerevisiae (yeast)*
- *B. subtilis*
- *E. coli*
- *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*
- *Various human cells - e.g. Neurons, Hepatocytes, Heart, etc.*

4) Standardisation of *in vitro / in vivo* experiments and their data : It is often the case that *in vitro / in vivo* data to be used in CSB modelling are inconsistent, inaccessible, unusable, incomplete, or unstructured. This often leads to CSB models being developed independently of data available, in the hope that data will eventually appear to provide parameters for the model, etc., or that parameters may be inferred in a "reasonable way" to make the model operate. Diversity in experimental regimes is vital, but nevertheless, data should be documented in a standard way, where possible. Within projects there should be clear standards.

Experiments should be designed taking account of standards for data collection, storage in databases, and analysis already defined and in place, and consistent with modelling that might take place. 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 CSB analysis, making full use of standard ontologies[1] and controlled vocabularies[2].

5) Standardisation of databases, software and modelling : A separate issue of standardisation relates to the computational software and modelling procedures themselves. As part of a number of worldwide projects, there are standard computer platforms being developed, such as the Systems Biology Workbench Initiative and SBML (Systems Biology Markup Language) (http://www.sbml.org), BioSpice https://community.biospice.org/ , the E-CELL project (www.e-cell.org), and the Virtual cell (http://www.nrcam.uchc.edu). Some of the current EU projects, such as EMI-CD and COMBIO also will develop standard analysis packages.

---

[1] "Ontologies" may be defined as systems for representing knowledge - in the present context, this particularly refers to the representation of models and hypotheses in computer-compatible terms. For example, the Gene Ontology (GO) project http://www.geneontology.org was set up as a collaborative effort to address the need for consistent descriptions of gene products in terms of their molecular function, biological process and location of action.

[2] Biological databases describe a wide spectrum of information. Their diversity makes efforts towards database integration difficult. For example, the BioBabel project http://www.ebi.ac.uk/biobabel/index.html aims to develop and implement standardised vocabularies and common ontologies to describe biological attributes in databases. This will allow users to do complex queries across databases in a simpler way.

Here 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 assumes 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. The ultimate goal are modular combinations of models and routine applications of 'standard' models in non-specialist (experimental) labs.

6) Data required beyond present 'omics (genomics, proteomics, metabolomics, etc.) : At present, genome, transcriptome, proteome, and metabolome studies dominate large-scale functional analyses. The missing link in contemporary functional analyses, however, is the capacity to observe the output of the true units of function. Such functional data may be, for example, exact cellular localization of proteins, their interaction in supramolecular structures, or reliable protein-protein interaction data. Capturing and modelling dynamic properties, with time course data and data on spatial distribution, is important. In addition, most data generated today give relative values. It is much more difficult, but necessary, to generate data with absolute levels (numbers of molecules per cell for instance).

Although all biological data should be collected with bioinformatics and CSB analysis in mind, there are key types of quantitative data becoming available which especially support CSB, and which require special attention for standardisation and analysis. These include:

- *Gene expression and transcription (microarrays)*
- *Protein-protein interaction (mass spectrometry, two-hybrid analysis)*
- *Genetic analysis and mutations (knock-outs)*
- *Comparative genomics (bioinformatics analysis)*
- *Metabolic flux analysis (C^{13}-labelling)*
- *In vivo imaging (e.g. time-lapse microscopy)*

7) Training : Significant levels of training are available in bioinformatics and computational biology at universities and national centres for each 'national research community', at the European level with Marie Curie fellowships, and also by private sector provision. However, there is often insufficient training for biologists in the use of CSB, especially since the field is rapidly developing. A framework for training in bioinformatics and CSB is needed.

European projects should emphasize the necessary role of training in bioinformatics and CSB.

8) Increasing excellence in experimental research projects via bioinformatics and CSB: With the increasing flood of experimental data, it is a recognized problem that often the data is not collected, stored and analysed in a way as to make the best use of the project results. To be successful, each life sciences research project should complement the experimental programme with clear objectives for experimental data storage and analysis, and where appropriate (e.g. high throughput experiments), a theoretical and consequent computational component and an experimental validation component[3].

Organisers and evaluators of larger genomics projects should recognise that a significant fraction, 10-50% of resources, should be devoted to bioinformatics, data analysis and CSB integration, depending on the nature of the project. Even smaller projects need significant data analysis resources.

[3] See discussion in "**Bioinformatics - Structures for the future**", a workshop report. ftp://ftp.cordis.lu/pub/lifescihealth/docs/bioinf_workshoprpt_2003_06_30_final.pdf

WORKSHOP SYNTHESIS - EXTENDED DISCUSSION

Workshop Background and Introduction to CSB

A group of European experts in computational systems biology attended a workshop on "**Computational Systems Biology (CSB) - Its future in Europe**" in order to provide a summary of the background, problem areas, current situation, and to fromulate recommendations for action by the Commission, policy makers, organisations in Member states and researchers themselves. As a result of submitted papers, presentations and discussions, the workshop participants emphasised messages in several key areas.

The essence of CSB is both understanding of the interaction of components of biological systems, and expression of this understanding in qualitative and quantitative terms. Traditional bioinformatics provides analysis and linkages between single entities, such as genes to RNA to protein to structure to function, and the relationships between and within organisms of these components, but in a linear and often non-quantitative way. The necessary CSB representations of complex subsystems/modules of cells/organisms must enable quantitative predictions of system behaviour. As an example, it is not enough to know the gene to protein to metabolic process pathways in a qualitative fashion, it is also necessary to measure rate and transport coefficients, and to understand intermediate pathways. The new flood of functional genomics data provided by array technology is a key input to this process. As they develop, well defined and verified models of biological systems will lead to new scientific understanding and discoveries, and will have important applications in medical/pharmacological research and applications.

What makes a good CSB project? The most successful current implementations of computational systems biology rely on iterative cycles of computerised (*in silico*) model construction/refinement and predictions, linked to wet-lab (*in vitro*) and living specimen (*in vivo*) experimental design, experimentation, and data taking and storage in a way that the computer databases and software can make the best use of them. Existing data, in particular genomic and postgenomic data, are instrumental for initial model construction. Quantitative and comparable data sets are necessary for model validation.

Fragmented research in Europe

Status: During the Expression of Interest exercise (http://eoi.cordis.lu/search_form.cfm) of the EU, it was concluded that there was significant fragmentation of European research in this area, and this was supported by workshop members. However, in many ways, it is due to the breadth of the field, and to a lack of available funding compared to the USA and Japan. The CSB workshop has helped to promote a common understanding among leaders of different scientific disciples. It is obvious that substantial EU funding, possibly in combination with national funding agencies, is necessary to foster computational systems biology in Europe.

Recommendations: In the near term, EU funding instruments such as STREPs (http://www.cordis.lu/fp6/instruments.htm) of a few million euros and teaching networks are necessary. STREPs are an immediate possibility for well-defined projects of small to medium sized interdisciplinary research teams. Such teams are in some sense the core for larger, high-quality network-based projects. Integrated Projects and Networks of Excellence including teaching networks are important to overcome fragmentation at all levels and to supply appropriately educated scientists. As a consequence of the broad applicability and the involved tools, one or two large projects and networks are insufficient to advance computational systems biology in Europe. *Substantial EU funding tied to structuring the field, in combination with national funding efforts, is necessary to foster computational systems biology in Europe. Preference should be given to strategic projects that provide both conceptual computational and experimental advances together, using relevant model*

systems. Computational systems biology is a nascent field, and preference should be given to strategic projects that provide both conceptual computational and experimental advances together, using relevant model systems. Selection criteria should include:

- Relevance of the model system
- experimental accessibility of the model system for hypothesis testing
- available European expertise
- possibility to establish European leadership.

General Modelling requirements

Status: There are a number of existing CSB projects around the world, and a number of new EU funded projects that are primarily CSB oriented or with a strong CSB component. (For a similar discussion of the situation in bioinformatics, see the June 2003 Bioinformatics Workshop report, available at ftp://ftp.cordis.lu/pub/lifescihealth/docs/bioinf_workshoprpt_2003_06_30_final.pdf). CSB projects range from detailed calculations of individual metabolic pathways to ambitious attempts to model the spatial and dynamic attributes of an entire cell, its component parts, interactions with other cells, and whole physiological systems. A major post-genomic challenge is moving from genomic sequence to a complete understanding of gene function and biological processes. Today the US and Asia are much more active in the area of modelling and simulation of complex processes than Europe. While there exists no European modelling platform, efforts have been made in the US with respect to the Systems Biology Workbench Initiative (http://www.sbml.org) at Caltech for several years. European scientists have contributed significantly to the development of SBML. The core language has been initialy developed by a team of four british computer scientists, and several other EU scientists have been involved in its definition right from the beginning. The goal of the USA DARPA-funded project BioSpice, partly incorporating SBML, is to create an open source framework and toolset for modeling dynamic cellular network functions https://community.biospice.org/. Furthermore, with the E-CELL project (www.e-cell.org) Japan gains a growing expertise in whole-cell in silico modelling. This is also a major aim of the Virtual Cell initiative in the US. The academic sector in these countries has been funded to a large extent in recent years, resulting in new initiatives (Alliance for Cellular Signalling, www.cellularsignaling.org), research groups and even institutes, for example the Institute of Systems Biology (www.systemsbiology.org). It is obvious that these research fields will gain major importance in the next couple of years. European research must face this challenge now and undertake efforts to close the gap. Other worldwide modelling projects and conferences include (cf. www.systembiology.net):

- ERATO Kitano (Japan/CalTech)
- Virtual cell (modelling software) (http://www.nrcam.uchc.edu)
- Bernard Pallson, UCSD
- International E. coli Alliance (Science August ‘02)
- ICSB 2000, 2001, 2002, 2003, 2004
- The Institute for Systems Biology (http://www.systemsbiology.org/)
- The Molecular Science Institute (http://www.molsci.org/)
- The Harvard Department of Systems Biology (http://sysbio.med.harvard.edu/).
- Center for Bioinformatics & Computational Biology (http://www.nigms.nih.gov/about_nigms/cbcb.html)

European research efforts especially include:

- Zhabotinsky,Turing patterns + Prigogine/Hess school
- Glycolytic oscillations (Duysens)
- Chemiosmotic ATP synthesis (Mitchell)
- Metabolic Control Analysis (Kacser, Heinrich, Groen)
- Phosphoneural net signal transduction (Hellingwerf)
- Silicon cell (Westerhof)

Some EU and European projects (see individual presentations in this report for details)

- EMI-CD-Platform for data integration and modelling of complex biological processes
- COMBIO - An integrative approach to cellular signalling and control processes: Bringing computational biology to the bench.
- FunGenES - Functional Genomics in Engineered ES cells
- QUASI - Dynamic operation of MAP kinase signalling pathways
- Industry: Wide interest biotech (Unilever, DSM, …) and Pharma (Bayer, GSK, Novo, AKZO)
- German hepatocyte project (www.systembiologie.de)
- EUSYSBIO/ESBIGH to promote systems biology

Recommendations: A general analysis of this range of projects may be made from discussion at the workshop, and from analysis of comments by reviewers and evaluators of projects, both accepted and rejected. Important advances have been made and results obtained on individual pathways, cycles and transport mechanisms (cf. Computational Cell Biology, Fall et al., http://www.compcell.appstate.edu). However, when projects become too ambitious, involving excessively complex biological systems as compared to the data available concerning the system, or a poorly coordinated research organisation or approach, a disappointing situation results. The ideal situation seems to be one where the modelling development is slightly visionary beyond the current state of the art, but at the same time firmly anchored in comprehensive data sets that are mean to be modelled. In order for this to happen, a combination of standards, data and approaches are required.

A key priority is the development of European-wide packages for creating and integrating relevant databases and analysis software to enable systems-level interpretation of complex experimental data in functional genomics. European biology databases require uniform standards to allow for transparent access to applications and heterogeneous distributed resources. Information necessary to go to the next stage of understanding of complex biological systems includes genome information, gene function, pathway, and interaction data. The focus should be on networking research institutions by generating a widely applicable and accessible communication and analysis layer as well as on database analysis and service development to enable the exploitation of a wide range of bimolecular information. These tools should lead to *in silico* simulations and predictions of gene function by integrated modelling of several complex cellular pathways. The consistency and the predictive power of the models should be evaluated in close collaboration with experimentalists.

Research projects should focus on integrated modelling of several cellular processes leading to as complete an understanding as possible of the dynamic behaviour of a cell. 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 involve:

- *Coherent, high-quality, quantitative, heterogeneous and dynamic data sets as a basis for novel model constructions to advance from analytical to predictive modelling.*
- *Experimental functional analysis tools (in-situ proteomics, protein-protein interactions, metabolic fluxes, etc)*

Model organisms as data sources

At this stage of development of CSB, results from all biological systems and model organisms are relevant, but some model organisms have been more completely studied than others, and are more able to provide the full range of data needed for modelling.

Research projects should start with well characterised model systems at the single cellular level, while linking these to multi-cellular model organisms and man to develop aspects of health research : Potential single cell model systems to analyse include:

- *S. cerevisiae (yeast) – model eukaryote, excellent exp. accessibility, presently the spearhead of European systems biology with a clear lead versus the US and Japan, existing European networks, strong industrial interest in biotech and in pharma as a model.*
- *B. subtilis – the gram-positive model microbe, excellent experimental accessibility, strong biotech industry interest, history of EC funding with excellent, established networks in place.*
- *E. coli – probably the best known microbe, excellent exp. accessibility, projects could be tied to the International E. coli Alliance (IECA) to position Europe within this world-wide program and to ensure that Europe has access to the conceptual advances made in this top-notch project.*
- *Filamentous fungi – strong European networks, history of EC funding, strong biotech interest. Disadvantage: limited experimental accessibility, additional levels of complexity, lack of a clear model case for higher cells.*

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*
- *Various human cells - e.g. Neurons, Hepatocytes, Heart, etc.*

Standardisation of *in vitro / in vivo* experiments and their data

Status: It is often the case that *in vitro / in vivo* data to be used in CSB modelling are inconsistent, inaccessible, unusable, incomplete, or unstructured. This often leads to CSB models being developed independently of data available, in the hope that data will eventually appear to provide parameters for the model, etc., or that parameters may be inferred in a "reasonable way" to make the model operate. Diversity in experimental regimes is vital, but nevertheless, data should be documented in a standard way, where possible. Within projects there should be clear standards.

Recommendations: *Experiments should be designed taking account of standards for data collection, storage in databases, and analysis already defined and in place, and consistent with modelling that might take place. 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 CSB analysis, making full use of standard ontologies and controlled vocabularies.* [cf. TEMBLOR project (http://www.ebi.ac.uk/integr8), involving the array data project DESPRAD (http://www.ebi.ac.uk/microarray/Projects/desprad), and BIOBABEL (http://www.ebi.ac.uk/biobabel) for standard ontologies and controlled vocabularies.] Since biological systems are often sensitive to the exact environmental conditions, their quantitative system responses are not directly comparable for quantitative modelling if, for example, different conditions (often subtle and unnoticed) were used to generate data sets in different labs. For the initial phase of model testing and hypothesis generation, it is thus of utmost importance to rely on consistent and standardized data sets. Beyond the use of one strain and defined physiological conditions, this includes standardisation of system perturbations (e. g. genetic or environmental modifications), well-defined and verified analytical methods, and consistent statistical data treatment. Appropriate control mechanisms to verify data comparability and reliability should be part of systems biology projects in research networks. This standardisation is crucial for the initial phase of computational systems biology to allow identification of faithful models and parameter sets. It should be understood though that once a suitable model is defined, it should be able to deal with non-standardized data. In fact, the identification of data in large heterogeneous sets that are quantitatively or structurally inconsistent with other data or the present model is a hallmark of systems biology. A close and continuous interaction between modelling requirements and experimental planning and operation leads to the best results.

Standardisation of databases, software and modelling

Status: A separate issue of standardisation relates to the computations themselves. As part of a number of worldwide projects, there are standard computer platforms being developed, such as the Systems Biology Workbench Initiative and SBML (Systems Biology Markup Language) (www.sbml.org), the E-CELL project (www.e-cell.org), and the Virtual cell

(http://www.nrcam.uchc.edu). Some of the current EU projects, such as EMI-CD and COMBIO also will develop standard analysis packages.

Recommendations: *Here 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 assumes 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 consistent multi-platform, non-proprietary programming languages, such as SBML, with professional standards for software production and maintenance. The ultimate goal are modular combinations of models and routine applications of 'standard' models in non-specialist (experimental) labs.*

Data required beyond present 'omics (genomics, proteomics, metabolomics, etc.)

Status: At present, genome, transcriptome, proteome, and metabolome studies dominate large-scale functional analyses. The missing link in contemporary functional analyses, however, is the capacity to observe the output of the true units of function. Such functional data may be, for example, exact cellular localization of proteins, their interaction in supramolecular structures, or reliable protein-protein interaction data. While the definition of function is somewhat fluid and a matter of controversy, there was a broad consensus that simply collecting 'omics data is insufficient. Capturing and modelling dynamic properties, with time course data and data on spatial distribution, is important. In addition, most data generated today give relative values. It is much more difficult, but necessary, to generate data with absolute levels (numbers of molecules per cell for instance).

Recommendations: *Although all biological data should be collected with bioinformatics and CSB analysis in mind, there are key types of quantitative data becoming available which especially support CSB, and which require special attention for standardisation and analysis. These include:*

- *Gene expression and transcription (microarrays)*
- *Protein-protein interaction (mass spectrometry, two-hybrid analysis)*
- *Genetic analysis and mutations (knock-outs)*
- *Comparative genomics (bioinformatics analysis)*
- *Metabolic flux analysis (C^{13}-labelling)*
- *In vivo imaging (e.g. time-lapse microscopy)*

Data integration is consistently identified as a top priority. At the first level, consistent quality-controlled large-scale 'omics data sets must be made available via databases. This includes also consistent statistical data treatment for the data sets. At the next level, however, these data sets must be integrated into predictive models of some detail that identify inconsistencies, systematic experimental errors, and important connections between certain subsets of heterogeneous data; all of which become then priority targets for further experimentation.

Training

Status: Significant levels of training are available in bioinformatics and computational biology at universities and national centres for each 'national research community', at the European level with Marie Curie fellowships, and also by private sector provision. However, there is often insufficient training for biologists in the use of CSB, especially since the field is rapidly developing. A framework for training in bioinformatics and CSB is needed.

Recommendations: *European projects should emphasize the necessary role of training in bioinformatics and CSB.*

Increasing excellence in experimental research projects via bioinformatics and CSB

Status: With the increasing flood of experimental data, it is a recognized problem that often the data is not collected, stored and analysed in a way as to make the best use of the project results. To be successful, each life sciences research project should complement the experimental programme with clear objectives for experimental data storage and analysis, and where appropriate (e.g. high throughput experiments), a theoretical and consequent computational component and an experimental validation component.

Recommendations: *Organisers and evaluators of larger genomics projects should recognise that a significant fraction, 10-50% of resources, should be devoted to bioinformatics, data analysis and CSB integration, depending on the nature of the project. Even smaller projects need significant data analysis resources.*

KEY AREAS - DETAILED CONSIDERATIONS

Definition of CSB

Like Functional Genomics in its beginnings, also (Computational) Systems Biology as a field is interpreted differently by different people. It is often regarded as a further step of Bioinformatics and Functional Genomics. For instance, it is thought that Computational systems biology encompasses technologies that bring system and structure into the huge amount of functional genomics data. While Computational systems biology should make use of those data, their organisation is, however, the task of Bioinformatics. In addition, Computational systems biology is often defined as a discipline that strives for a complete understanding of whole cells and organisms. While it is difficult to comprehend what a "complete" understanding may encompass, a global understanding of how different subcellular systems interact and function in context is certainly of interest to Computational systems biology.

Possibly a useful description of Computational systems biology derives from its actual goals. Those are to understand the structure and function of biological systems that are composed of a certain number of interacting biomolecules, cells or even organisms. In other words, Computational systems biology strives at the understanding of the logic and the elucidation of the functional rules of modules or systems, rather than the individual parts of those.

Based on this definition, Computational systems biology is inherently multidisciplinary and requires the input from biomedical experimental research as well as from mathematics, computer sciences, physics and engineering. More specifically, Computational systems biology makes use of mathematical models (computer replicas of the system) that are based as much as possible on actual data in order to understand systems properties such as feedback loops, robustness, bistability and more. An important property of the mathematical models is that they can be used to predict properties of uncharacterised systems components, predict the results of experiments and help phrasing hypotheses, thereby assisting experimental planning, reducing the number of experiments and opening up for a number of possible applications such as in drug development, diagnosis, breeding and genetic engineering.

See also: www.systemsbiology.org

Why has Computational systems biology become a topic now?

The use of mathematical models in biological research is not at all new. However, for many years those models commonly had little if any footing on actual data and therefore lacked realistic use. In fact, they were regarded by experimentalists as a playground for mathematicians and being completely useless. What has changed in the last few years?

- The availability of global data, such as gene expression and proteomics data that provided information on most or all components of a module/system.
- The emerging (though still not general) interest of biologists to collaborate with mathematicians and researchers from other disciplines, which partly has been driven by relevant programmes from different funding agencies, including the EC.
- The realisation that Computational systems biology approaches can help advancing biomedical research and allow addressing research questions that cannot be targeted by experimentation alone.
- Persons that drive and shape the field and gave it the name Systems Biology, such as Leroy Hood, Hiroaki Kitano, Roger Brent, Hans Westerhoff and others.

What Computational systems biology can do and possibly deliver

Computational systems biology works with mathematical models that precisely replicate the structure and function of the relevant module/system under study. This means, to the best possible extent should the model be based on experimental data and it should be able to simulate as precisely as possible the actual operation of the system. The cooperation of experimentalists and mathematicians should result

in iterative improvement of the model and hence the understanding of the system. In other words, the model should in mathematical form contain all available experimental knowledge and connect individual data to a functional unit. As such, the model has predictive abilities, i.e. it opens for the possibility for in silico experimentation to test alterations to the system or new perturbations that have not previously been tested in experiments. Based on these requirements, Computational systems biology approaches can be applied for the following:

- Planning of experiments, thereby optimising the design of biological experiments, reduce their number and hence make research more cost-effective and targeted.
- Elucidating properties of biological modules/systems that cannot be understood on the basis of experimental data alone.
- Identifying components of the system hitherto unknown from experimental work.
- Help to understand the basis for diseases and diseases processes.
- Assist to identify the "weak spots" in systems, i.e. the possible targets for pharmacological intervention: drug target discovery and drug design.
- Help to determine, eventually on an individual basis, the best timing and mode of drug application to cure diseases.
- Help designing approaches of genetic engineering or breeding to optimise crops of biotechnologically relevant microorganisms.

Needs and actions

It can be expected that:

- Computational systems biology approaches become an integral part of biological/molecular biological research over the coming years.
- Computational systems biology will become important to fully exploit the potential of genomics and functional genomics.
- Computational systems biology will become highly important in drug target identification, drug design, assessment of side effects, drug approval and application to patients.
- Computational systems biology will become an important tool for breeding and genetic engineering of crops, farm animals and microorganisms.

For these reasons it is important to develop the area aggressively.

Europe could potentially be in a leading position since it has a tradition in Systems Biology approaches. But at this point Europe lags behind because Japan and the US have already invested heavily (publicly and privately) and have built relevant infrastructures (such as the ERATO Kitano project in Japan, the Institute for Systems Biology in the US, and others). A present problem in Europe is fragmentation of the area, which is apparent already when it comes to a definition of the field. For this reason it may be advisable to define the type of possible future Computational systems biology projects explicitly (see example below). Fragmentation is also apparent when it comes to funding: while some countries like Germany, Finland and the Netherlands have already put programmes in place, other countries lag behind.

What kind of action is needed in FP6?

- Success stories: well-defined Computational systems biology projects that testify the power of the approach.
- A visible larger project, either and NoE or an IP, see example call text below.
- Actions to coordinate at the EU level national initiatives and encourage such initiatives where they do not yet exist.
- Training, especially of researchers trained both in experimental and mathematical research. This can be achieved through Marie Curie Actions (Networks, EST…).

The following is an example call text for a NoE/IP in FP6:

Elucidation of how system properties arise in defined cellular modules. The objective is to enable researchers to study properties and dynamic operation of complex biological modules/systems. Projects should make use of existing data as well as experimental and computational approaches to understand the properties and operation of cellular modules/systems in model organism. Among deliverables should be tools for predictive in silico experimentation to use the full potential of genomics and post-genomics.

What could be done for FP7 and beyond?

- Call for several projects (IP) to achieve silicon replicas of larger modules (metabolism, signalling, trafficking, organelles, cell cycle, gene expression, replication, cytoskeleton) in model organisms.
- Extensive coordination and management of the approaches.
- A vision to strive for whole cell projects in FP7 and 8.
- Extensive support for training.
- Requirement for Computational systems biology components in all relevant biomedical projects.
- Thereby establishing a mid and long term vision and sustainable funding perspectives for the area.

What are some of the key needs in CSB?

Experimental standardisation

Standardisation of quantitative heterogeneous data sets pertains primarily to their direct comparability. Since biological systems are often sensitive to the exact environmental conditions, their quantitative system responses are not directly comparable for quantitative modelling if, for example, different conditions (often subtle and unnoticed) were used to generate data sets in different labs. For the initial phase of model testing and hypothesis generation, it is thus of utmost importance to rely on consistent and standardized data sets. Beyond the use of one strain and defined physiological conditions, this includes standardisation of system perturbations (e. g. genetic or environmental modifications), well-defined and verified analytical methods, and consistent statistical data treatment. Appropriate control mechanisms to verify data comparability and reliability should be part of computational systems biology projects in research networks. This standardisation is crucial for the initial phase of computational systems biology to allow identification of faithful models and parameter sets. It should be understood though that once a suitable model is defined, it can certainly deal with non-standardized data. In fact, the identification of data in large heterogeneous sets that are quantitatively or structurally inconsistent with other data or the present model is a hallmark of computational systems biology, and will become of great value.

Computational standardisation

A separate issue of standardisation relates to the computations themselves. Here standardisation is necessary at the level of unambiguous (and simple) representation of the networks and components modeled, model description (including reaction specification, measurement units, etc.), data storage and retrival, and the computer codes. The latter assumes a particular importance because models developed by different networks/groups represent typically modules of cellular operation that should be compatible with each other. Hence, development and use of consistent use of multi-platform, non-proprietary programming languages such as SBML is a priority. The ultimate goal are modular combinations of models and routine applications of ‘standard’ models in non-specialist (experimental)

labs. Hence, the development of open-source, professional software (and maintenance) should be encouraged.

Beyond present 'omics:

Functional data At present, compositional transcriptome, proteome, and metabolome studies dominate large-scale functional analyses. The missing link in contemporary functional analyses, however, is the capacity to observe the output of the true units of function [2]. Such functional data may be, for example, exact cellular localization of proteins, their interaction in supramolecular structures, or reliable protein-protein interaction data. While the definition of function is somewhat fluid and a matter of controversy, there was a broad consensus that simply collecting 'omics data is insufficient.

In linking genes and proteins to higher-level biological functions, the molecular fluxes through fully assembled biochemical networks determine the systemic phenotype in metabolic research [3]. The capacity to quantitatively observe this whole network operation by methods of metabolic flux analysis based on 13C-labeling experiments, thus provides a global perspective of the integrated, system-wide regulation at the transcriptional, translational, and metabolic level. Such quantitative functional information is highly important for computational systems biology.

Data integration was consistently identified as a top priority. At the first level, consistent (same strain/conditions) and quality-controlled large-scale 'omics data sets must be made available via databases. This includes also consistent statistical data treatment for the heterologous data sets that goes beyond the current ad hoc practice. In collaboration with experimentalists, this is the realm of bioinformatics. At the next level, however, these heterologous data sets must be integrated into predictive models of some detail that allow to identify inconsistencies, systematic experimental errors, and important connections between certain subsets of heterogeneous data; all of which become then priority targets for further experimentation.

Computation vs. experimentation

In sharp contrast to functional genomics, computational systems biology does not follow a large-scale data collection and analysis scheme. Computation and experimentation are simultaneously occurring and integrated components of computational systems biology research. Models may be build from publicly available data to indicate – with lower confidence of course – the most important next experiments for a given experimental subsystem. Initially quantitative, inspired guess experimentation may be the major effort of most projects, but eventually model-derived hypotheses will become increasingly important for experimental design. In the intermediate and long run, computational systems biology will significantly reduce novel experimentation because computations identify pivotal missing components for quantitative understanding of the fully assembled system or module.

Fragmented research in Europe:

What is required from the EC? Fragementation includes research and funding in different countries but also know-how and approaches in different scientific fields. How can a fruitful environment be created?

The CSB workshop has helped to promote a common understanding among leaders of different scientific disciples. It is obvious that substantial EC funding, possibly in combination with national funding agencies, is necessary to foster computational systems biology in Europe. At short term, STREPs, NoEs, and teaching networks are necessary. STREPs are an immediate possibility for well-defined projects of small, interdisciplinary research teams. Such teams are in some sense the core for larger, high-quality network-based projects. NoEs and Teaching networks are important to overcome fragmentation at all levels and to supply appropriately educated scientists, respectively. At an intermediate scale, larger research networks are important (IPs). Appropriate calls for IPs in computational systems biology make only sense, however, if a reasonable volume is made available. As a consequence of the broad applicability and the involved tools, one or two IPs are insufficient to advance computational systems biology in Europe.

How to select for high-leverage projects?

Computational systems biology is a nascent field, hence cannot be expected to yield applied benefits such as novel drug targets etc. immediately. Instead preference should be given to strategic projects that provide conceptual computational and experimental advances, using relevant model systems. Selection criteria should include:

- Relevance of the model system
- experimental accessibility of the model system for hypothesis testing
- available European expertise
- possibility to establish European leadership.

Potential model systems

- S. cerevisiae – model eukaryote, excellent exp. accessibility, presently the spearhead of European systems biology with a clear lead versus the US and Japan, existing European networks, strong industrial interest in biotech and in pharma as a model.
- B. subtilis – the gram-positive model microbe, excellent exp. accessibility, strong biotech industry interest, history of EC funding with excellent, established networks in place.
- E. coli – probably the best known microbe, excellent exp. accessibility, projects should be tied to the International E. coli Alliance (IECA) [4] to position Europe within this world-wide program and to ensure that Europe has access to the conceptual advances made in this top-notch project.
- Filamentous fungi – strong European networks, history of EC funding, strong biotech interest. Disadvantage: limited experimental accessibility, additional levels of complexity, lack of a clear model case for higher cells.
- Neurons – was discussed as an example of a higher cell type with a competitive situation for Europe, highly interesting but low exp. accessibility.
- Hepatocytes – given the strong funding in Germany, EC projects may aim at connecting this nucleus to other top European groups in the field.

Immediate needs

- Coherent, high quality data sets as a basis for model construction
- quantitative dynamic data sets for time-dependent changes
- absolute concentrations of proteins (and their modification) and mRNAs
- new experimental tools for functional analysis (in situ proteomics, reliable protein-protein interactions, metabolic fluxes etc)
- heterogeneous data integration
- software and model standards
- new modelling concepts
- advance from analytical to predictive modelling

Computational Systems Biology in biomedical research

CSB is hot, BUT there are problems.....

- Insufficient cross-talk of biology, mathematic & engineering
- Many CSB activities - fragmented area
- Lack of integration and coordination of current CSB activities
- Only few national funding programs exist in Europe
- Little participation of Europe in other CSB activities (USA, Asia)
- Student training not widely available at Universities
- CSB potential very high, but exploitation difficult (SMEs...)

The immediate needs of CSB in Europe.....

- Competitive CSB demands FP6, 7... (ERC?) and national support

- Connection of existing national funding programs - ERA-NET
- Database and communication platform for European scientists
- Identify researchers in member states as CSB contact points
- Human ressources - Ph.Ds, post-doctoral levels & junior groups
- Identify CSB topics and appropriate model systems

Standard protocols for data aquisition and common tools

How can we address problems and needs of CSB..

- Human ressources to enable competitive CSB development
- Limit models systems (bacteria, yeast, mammalian systems)
- Establish national programs in member states - academic
- Involve SMEs and Big Pharma (research & funding)
- Networking and coordination of ongoing CSB activities
- CSB funding should use bottom-up and top-down approaches

Training of human ressources - a key for European CSB.....

A bright future for European CSB.....

- Concentrate and network CSB research activities
- CSB needs to think "European"
- Critical mass in funding and scientists - think "Big"
- Establish Europe as competitive key player in global CSB

Summary thoughts....

- Workshop helped convey importance of CSB to EC - FP7
- EC funding (think BIG) might induce national activities where non-existing (i.e. via ERA-Net).
- Multidisciplinarity is definition of CSB - proposal review by EC!
- Training & Mobility (CSB Ph.D. in MC, support hi-le conferences)
- National CSB centers as training sites - Projects in CSB
- Realistic - Choose model systems where experimental tools are availailable and thus data generation is possible!
- NoE for coordination of European activities

CSB anchored in Model Organism Data

Publication of the preliminary nucleotide sequence of the human genome at the turn of the millennium was one of the milestones in modern biology. Yet this information package of 3 000 000 000 nucleotides marks only the beginning for modern "postgenomic" research on molecular genetics and life sciences in general. A characteristic feature of such research is the generation of increasing amounts of raw data requiring advanced informatics services and tools to process this information into biological knowledge. In addition to the draft nucleotide sequence of the human and mouse genomes, those of several other eukaryotic model organisms, e.g. *Caenorhabditis elegans*, *Drosophila melanogaster*, *Arabidopsis thaliana, Saccharomyces cerevisiae, Saccharomyces pombe* and those of more than 100 bacteria and archaea are currently known. This offers unique possibilities for comparative biological analyses. Analysis, organisation and mathematical modelling of large amounts of data present a major challenge to modern bioinformatics. Examples of research fields utilising high throughput methodology and thereby producing huge data sets include structural biology, molecular modelling and simulations and genetics of multigenic traits and diseases.

Multidisciplinarity and integration are characteristic features of postgenomic research. Genes, gene products, their regulatory networks and interactions with environment must be analysed as

components of higher order structures, metabolic pathways or entire cells and organisms. This type of an integrative and holistic approach has been termed **systems biology**. Research defined as computational systems biology is characteristically multidisciplinary and dependent on **bioinformatics**, the computer-assisted analysis of biological data. Close collaboration of biologists, biochemists, physiologists, chemists and physicists with computational biologists and mathematicians is needed for the characterisation and modelling of the complex interactions of genes, proteins and metabolic processes.

A range of ethical, political and economical constraints limit the generation, processing and application of biological information about populations and individuals. These challenges are not only faced by researchers, but are also important to decision makers and laymen. Storage and use of human genetic information and manipulation of genomes pose ethical questions and challenges, necessitating research on the ethical, social and cultural dimensions of bioinformatics and computational systems biology.

THE SCOPE OF AN INTEGRATED CSB PROGRAMME

In order to understand the complex biological systems, knowledge of the molecular characteristics of individual components or phenomena is not enough. A holistic view and integrative, multidisciplinary approach is needed to study the complex interactions between components and networks.

Examples of research fields that the Research Programme on Computational systems biology and Bioinformatics will cover:

Structural biology
Functional genomics and proteomics
Molecular genetics
Bioinformatics, biomathematics, and computational biology
Ethical, social and cultural aspects

Why Computational systems biology?

Culminating in complete genome sequences and genomics, molecular biology, biochemistry and biophysics have led to appreciable understanding of the macromolecules of living cells and to an impressive number of tools. The tools enable one to obtain much more such information when needed. However, obtaining *all* information about *all* molecules in *all* organisms remains too costly, and may limit the seeing of the forest for the ever-increasing number of trees. What appears needed is a focus on the original and true issues, such as the understanding of how living organisms function, of how they sometimes dysfunction (such as in disease), and how their function can be improved both in therapeutical and in biotechnological settings.

What seems to limit the understanding of function now is the phenomenon that much of the function of living organism comes about in the complex interactions of the macromolecules. It is the (lack of) understandings of these complex interactions, i.e. the computational systems biology more than the molecular biology, that is now limiting progress.

An operational definition of computational systems biology

There are various definitions of Computational systems biology [cf. www.systembiology.net]. Yet, it is not a vague discipline. Computational systems biology is neither the Biology of Systems (which is Physiology), nor the physical-chemistry and mathematics of their components (modern molecular biology), it is the in between. It focuses on the new properties, important for biological function, that arise in the interaction of the components of Biological systems, i.e. that are not present in those components in isolation.

The in-between can be at various levels of biological organization. One that is particularly acute, thanks to the explosive advances in genomics, addresses the level between macromolecules and the simplest form of autonomous life, i.e. single living cells, such as microorganisms and tumor cells. At the level between organisms and ecology, there is another example of Computational systems biology. We shall focus on the former example.

Deliverables of Computational systems biology

For good reasons, research agencies require scientific activities to lead to results that are important for society. Without guaranteeing delivery within two years, we here mention a number of deliverables:

- Discovery of new scientific principles that govern at the system level and not at the molecular level (: high quality science leading to Nobel prizes, Fields medals)
- Insight in the pathology of multifactorial diseases (e.g. cancer, type-II diabetes, rheumatoid arthritis, heart failure, infectious diseases) and in the diverse pathologies of unifactorial diseases against the backdrop of polymorphisms.
- Multifactorial, subtle therapies for various diseases.
- New drugs deriving from network-based drug design
- New drugs and strategies to combat multidrug and antibiotic resistance
- Computer models (and -replica) of patients helping to manage their disease and to pretest and optimize therapies.
- Much reduced frequency of animal experimentation through substitution by computer models
- Increased insight of the public in genomics research through layman-accessible computer simulations (and 'games') of living organisms, of research issues and of therapy and biotechnology
- Support of companies *vis-à-vis* regulatory agencies such as the FDA which will soon require computer model validation in addition to experimental validation of drug descriptions
- Insight in how functional systems can work that are evoluationarily stable; inspiration for man made society and ethics discussions

What is needed for Computational systems biology?

Computation

New behavior of systems relative to their components arises through the nonlinear interactions of the latter. Such nonlinearity cannot be understood through the standard intuition, but requires assistance by computations.

Experimentation

The behavior of nonlinear systems depends on their operating point and on the magnitudes of their parameter values. These need to be determined quantitatively and sufficiently accurately, which has not been the priority of molecular biology until now. Therefore a new line of experimentation is needed, part of which should be directed towards experimentation inside living cells.

Conceptual advances

A computer replica of a living organism has the tendency of being equally unintelligible as the original. Therefore new conceptual tools are needed to facilitate the understanding of biological complexity. Already existing examples of such tools include metabolic and hierarchical control analysis, modularization concepts (such as elementary modes), stability analysis, but more will need to be developed.

Their integration

The above lines of Computational systems biology will need to develop in an integrated manner in a procedure that also includes discovery, hypothesis, validation and falsification.

Model Systems

Computational systems biology requires the integration of much and precise information about a system, which is difficult and expensive to obtain, and which requires the collaboration of much man power from many, diverse disciplines. Because of this breadth, focus of much of the activity on a limited number of models systems is required. The model systems should be selected on the basis of: (i) experimental accessibility, (ii) possibility to obtain the information needed by the computation system biology, (iii) relevance, (iv) the existence of scientifically exciting computational systems biology issues in the, (v) the possibility that Europe can contribute substantially, possible in a leadership role.

Computational systems biology is a discipline in development. Although appreciable roots of Computational systems biology lie in Europe, North America and Japan have taken strong positions

already. Therefore, we should distinguish model systems in which Europe might take the lead, and model systems where Europe may be an equal partner to the Japanese and the Americans.

Where Europe may lead:

- *Lactococcus lactis* (model prokaryote; simple model system; substantial Dutch, Danish and French initiatives); thorough industrial (biotech) interest
- *S. cerevisiae* (best-known eulkaryote; first sequenced eukaryote, largely thanks to a European effort; much of the system biology of this organism reside in Europe); thorough industrial (biotech) interest
- The hepatocyte (model mammalian cell; large German initiative http://www.bmbf.de/pub/systems_biology.pdf ; tremendous interest pharmaceutical industry)
- The neuron (Fundamental interest on signal transduction and integrations. Huge medical applications. Existing network on model system DopaNet http://www.dopanet.org)
- Metabolic and Hierarchical Control Analysis (conceptual method; historically led by European groups)
- Silicon cell (computer replica of parts of living cell; here in the sense of *precise* replica Europe leads)
- Where Europe should participate in world-wide programmes
- *Escherichia coli* (best known model prokaryote; existing world-wide International Alliance for *E. coli* Alliance; IEcA)
- Virtual cell: modeling tools for molecular cell systems biology connected to the SBML initiative

Computational approaches

Europe can certainly make a leadership contribution in Computational Systems Biology. Recent American work has excelled in applying exiting engineering views to cell biology. They have lacked specificity. They did not always address reality.

Because of its tradition of collaboration between research groups, Europe may well lead in the challenge of making computational systems biology contact to the reality of molecular cell biology. The following types of programs could be characteristic:

The silicon cell: precise replica of living cells. For the creation of these through collaboration is required between computional scientists and experimnetall biologist.

Interactive Computational systems biology: modeling activity in direct contact with experimentation; computation aided experimental design; experimentation based model optimization

Concepts for computational systems biology: theoretical tools phrased in terms of molecular cell biology have always been developed most in Europe. Further developments of these for the new computational systems biology are a good bonanza for Europe (examples: Control Analysis, non equilibrium thermodynamics)

What requires EU support?

- Coordination
- Postdoc grants
- Training grants
- Network of excellence:
- Support for coordination of the national activities: Such support should not only comprise support for visits, but also support for salaries of transdisciplinary and transnational scientists, i.e. scientists that venture to a laboratory in a different EU country with a different aspect of Computational systems biology (e.g. an experimental cell biologist going for one or two years to a lab in a different country to models his system of interest; postdoc or sabbatical, or mobility). In addition there should be a network of excellence coordinating the best national CSB initiatives.
- Substantial support for training activities.

CSB AS AN EMERGING FIELD

Describing what an emerging field isn't is always easier than to describe what it is.

Computational Systems Biology (CSB) is not Bio-informatics which is mainly concerned with information retrieval or with information extraction. By information retrieval I mean finding biological information in huge amount of data by building systems (e.g. blast, fasta or SRS) or databases (e.g. the embl nucleotide data bank, swissprot or unigene). By information extraction I am referring to those programs which elucidate new information about unidentified biological features (e.g. genscan or clustalw).

A CSB study must have a dynamic dimension: data gathered must not only describe temporal co-occurrences of measures for various biological phenomenon or objects, they must somehow describe the trajectory of a biological system through time.

CSB goals.

CSB aims at building models representing systems or subsystem of living organisms (e.g. a cell, a couple of cells interacting, a molecular subsystem accomplishing a given function, or a metabolic pathway). It is expected that those models can describe and predict the dynamic behaviour of such systems.

The predictivity of the models that are desired in CSB must not be achieved by sacrificing the understanding of those systems. Though it can be legitimate to use "black boxes" in the process of building these models, a complete explicitation must be striven for.

A reasonable objective of CSB research programs could be to model living systems or organisms the way they are and also the way they could be. Question like "what happens if a signaling pathway is changed for another?" could be answered by CSB models. In principle, independent of the ethical issues involved, CSB should be able to aid in the design of artificial organisms meeting some arbitrary specification. This can be seen as a test of the understanding of the modeled systems.

Some difficulties.

The relations between experimentalists and theoreticians are not clear in biology, both communities have different agendas and time frames. To overcome this difficulty one could not incite laboratories of both kinds to collaborate. One could also incite researchers of both kinds to work in the same laboratories or the laboratories to recruit researchers of both kinds.

CSB aims at describing the mechanisms that explain dynamic phenotypes of living organisms. However, such phenotypes must be considered only if they have some functional meaning for a given organism. The paradox resides in the fact that it is much easier to define what a function is at a high conceptual level (e.g. reproduction, nutrition...) than at the level of genes or proteins: is the function of the lactose repressor "to bind the operator in absence of inducer", or "to block the synthesis of proteins of the lac operon proteins in absence of substrate"? It is clear that the second proposition makes more sense than the first one, but it is also a far less objectivable proposition. The first proposition says "what" the lac repressor can be used for, the second one explain the role it plays in a (rather simple) biological system. We are very far from having a rational and objectively defined nomenclature of biological functions, and indeed, one can even wonder whether it should be the case.

CSB needs

CSB will bear upon rather important computer program developments. One can hope that these developments will be made with highly professional standards but we must keep in mind that most of the program developments in academic research laboratories are performed by graduate students or by post-docs that are mainly interested in obtaining biological results and not in creating robust software. This is probably a field where SME could bring some service to the academic community.

Data standardisation is another difficult point. It is clear that CSB research programs will have to integrate very important volumes of data and one can expect that they will come from many different experimental benches. It will be important to define standards, but we must keep in mind that premature definition of standards can lead to some important observations being missed. DNA sequence is rather well standardized, however, at the opposite end of the spectrum, protein-protein interactions data are far from being homogeneous. An intermediary possibility between data standardisation and total anarchy could be to define Standard Operating Procedures (SOPs) and to attach to each piece of data the reference to the SOP used to acquire it.

Expected industrial repercussions.

We can expect that the repercussions of CSB will be felt in several industries.

In the pharmaceutical industry, CSB should lead to better and earlier predictions of possible adverse effects of drugs. In the agro-food industry, a better understanding of the biological systems used in production should lead to increased control of the bio processes involved. The chemical industry should expect the design of new bio processes for the synthesis of known or even new compounds.

We must bear in mind that these are long-term repercussions, and that any short-term expectation may be met with deception.

Strategies and future efforts to install CSB funding

Although there are significant overlap of CSB with bioinformatics and biological projects it should have an own significant European funding in the form of a NoE or IP. The network should co-ordinate European CSB research, student training and drive force the integration of different research groups. Besides these efforts, CSB should be a part of appropriate experimental projects. To push the field, there must be some good projects that proof the necessity for CSB research. This will help stressing the importance of CSB for research and proof that it has practical implications for research and knowledge gain. A third part of funding effort should be available for the development of theoretical algorithms, new modelling tools and theoretical methodology.

Definition of CSB

Systems Biology is hard to define. It stands between different sciences and has an impact to these sciences vice versa. Thus, it is mainly defined by these interactions. It was discussed whether it is necessary to give a common definition. It might be an issue for the NoE to define CSB, its needs and goals.

Standardisation

It was emphasized by most of the speakers that standardisation is one of the main issues of CSB. We need standardisation in terms of experimental methods and analysis tools. Experimental design methods must be introduced from the beginning of a project. We should have a few experimental systems with standardised conditions for the test and performance of modelling strategies. The methods to apply depend on the questions of interest and determine the level of understanding of the system.

Data integration

The interaction of the different levels of information is crucial for successful CSB strategies.
There is a fundamental need for integrating data from proteomics, metabolomics, transcriptomics, and for the coupling of this information. Here, there is a significant overlap with bioinformatics methods and co-operations should be enforced. Furthermore, linking of databases on enzyme function, metabolic pathways, transcriptional regulation is needed. For example, we will have to represent the physical interaction between biochemical entities, map them onto their tissue and cellular localisation. We need tools for querying, displaying and analysing automatically the structure of interactions on various levels of granularity. One of the main bottlenecks is the curation, annotation and update of the databases. Static databases should be linked to forward modelling systems in order to parse the information into dynamic models. We need to create different type of views on the same network of molecular entities. A higher level of integration addresses the integration of model methodology itself by new mathematical methods.

Education
There is a fundamental need for education in CSB. CSB is the technology of the future but there is still insufficient cross talk and still not enough public recommendation (in comparison for example to bioinformatics). CSB is a fragmented area with a lack of integration and co-ordination within EU. There are a few national funding initiatives but there is a need for EU-wide initiatives and co-ordination. We need student training and academic program.

What can we expect from CSB ?

From a practical point of view (drug development) it was pointed out that CSB has little impact yet at all. Interesting biomarkers are screened without understanding of the model system, rather in a “trial and error” way. The question was raised, how CSB will tackle with genetic variation that is one the main issues in drug development. The situation at the moment is that CSB has high potentials that must be brought into practical results.

RESULTS OF QUESTIONNAIRE

A questionnaire was circulated to the participants and observers, to try to get a uniform view across projects, programmes, and proposed ideas. The table of replies is shown in Appendix A4 (formatted into two parts for the 11 replies). This questionnaire provides an overview of the various approaches, and also provides a cross comparison of various approaches.

It was found that this provided a very useful format for cross comparing topics and approaches, and even served as the basis for one of the contributed papers. The questionnaire topics were as follows:

- Title of project or topic
- Short description
- Choose the Research Area
- Level of complexity
- Criteria for choice
- How important is it?
- Do we have enough data and understanding to solve the problem?
- Is the problem tractable computationally?
- Goals of Research
  - Explain
  - predict
  - control
- Choose right software tools for problems complexity level
- For multiple interacting systems, hierarchy of tools and interfaces appropriate to problem, data, solution
- Database requirements
- Resources

The table was found to provide a cross comparison from various projects that lead to important insights. One of the most important areas comes to the key question in CSB approaches to problems: **Do we have enough data and understanding to solve the problem? Discuss the level of data input available: Genome, proteins, pp interaction, expression data?** The following answers provide a wide range of replies, but give a good summary of the current state of the art and the problems that are being faced. The most important point illustrated is that it is essential to have good collaboration between modellers and experimentalists to provide the right data in the right form.

- ✓ Many data are available in terms of genetics, biochemistry and more recently of global omics data. Nevertheless many of the data, as good as they are, are not tailored for systems biology approach (different biological systems, qualitative and not quantitative, noisy ...). Comparison of different hierarchical levels of data (transcriptional vs. protein vs. protein modification/interaction, vs. metabolic marker) is completely lacking in current data sets.
- ✓ Correlated data sets will be generated during the project, which in this type of interface density do not currently exist, which is highly advantageous. Rigorous adherence to data formats will be mandatory, and participants will be schooled to the standards required. Data types: Gene expression, proteomics (in particular posttranslational protein modifications, localisation and turnover rates), gene methylation, ion fluxes, and other parameters to be decided during project preparation. Use of data from other projects will be considered.
- ✓ Any problems we have to solve will only become tractable if we are able to organise the many pieces of information at our disposal. Breakthroughs normally occur when old problems are addressed with new insights/technologies.
- ✓ Two main problems: Firstly, data is extremely heterogeneous, i.e. produced on different levels of cellular information, produced with different biotechnical methods, in different labs, etc. Thus, data is often poorly correlated. A fundamental need therefore is the development of data

integration methods. Secondly, data is (in most cases) not well-designed to solve the problem. There is a fundamental need to understand that data production and data analysis are two elements that have interactions and feed-back. Methods of experimental design must be introduced at the initial phase of the experiments.

- ✓ All data types mentioned are available or will be generated in the course of the programme. They will contribute to achieve the goals of the funding initiative System of Life - Systems Biology
- ✓ Data only available for input and output but only little for intermediate steps and barely quantitative data or data at single cell/single pathway level. This is addressed in the project.
- ✓ We don't and this is a huge problem. Systems Biology of neuronal cells is impaired because we do not know where exactly are the proteins, in which amount, how they interact etc. A strong emphasis should be put on the funding of large scale data mining in model systems. Follow the example of the Alliance For Cellular Signaling.
- ✓ There are generally a lot but dispersed data available, and we lack accurate quantitative data from one experimental system: strain, cultivation, …… Most of the presented data are at the compositional level, e. g. transcriptome and proteome. Thus, we lack additionally quantitative metabolic data on metabolite levels and in vivo reaction rates for a global understanding of metabolic control.
- ✓ No. Discuss what is needed and how to obtain it. What is necessary to build models, find biomarkers etc?
- ✓ Indeed genomics, but different from what is common practice now. It is not so useful to determine everything at one genomic level (e.g. all mRNAs in a cell); it is much more useful to determine promoter activity, mRNA, protein, enzyme activity, metabolites, flux etc corresponding to one function; I call this vertical genomics). Interaction data are highly important, but above all data need to be more quantitative and precise and reliable. Much more kinetic data are needed. And data need to be generated that are necessary for the calculations; CSB driven experimental research is necessary.
- ✓ There are not enough data! We are lacking accurate quantitative measurements of protein concentrations measured for a large number of proteins (say 20) over a reasonable period of time (with sufficient time points). Ideally I would like to have the same experiment on at least two levels - transcriptome and proteome.

WORKSHOP COMPONENTS AND CONTRIBUTIONS

LIST OF PARTICIPANTS

1. Lilia Alberghina, Università Milano-Bicocca, Italy
2. Steven M. Foord, Director Target Bioinformatics Analysis, GSK, Stevenage, United Kingdom
3. Jaap Heringa, Bioinformatics Unit, FEW/W&I Vrije Universitei, Amsterdam, Netherlands
4. Ralf Herwig, Group Leader Bioinformatics, MPI-MG, Berlin, Germany
5. Stefan Hohmann, Department of Cell and Molecular Biology, Goteborg University, Sweden
6. Norbert Hübner, Max Delbruck Centrum, Berlin, Germany
7. Karl Kuchler, University and BioCenter of Vienna, Austria
8. Frank Laplace, BMBF, Germany
9. Nicolas Le Novère, Récepteurs et Cognition, Institut Pasteur, Paris,France (EMBL-EBI - 1 Oct 03)
10. Cedric Notredame, Structural & Genomic Information Lab., Marseille, France
11. Mark Sansom, Laboratory of Molecular Biophysics,University of Oxford, United Kingdom
12. Uwe Sauer, Institute of Biotechnology, ETH Hönggerberg, Zürich, Switzerland
13. William Saurin, Genomining, Montrouge, France
14. Luis Serrano, European Molecular Biology Organisation, Heidelberg, Germany
15. Age Smilde, TNO Nutrition and Food Research (& University of Amsterdam), Zeist, Netherlands
16. Alfonso Valencia, Protein Design Group, CNBC, Madrid, Spain
17. Eero Vuorio, Academy of Finland, Finland
18. Hans V. Westerhoff, Vrije Universiteit Amsterdam, Netherlands
19. Shoshana Wodak, U.L.B. , Bruxelles, Belgium
20. Petra Wolff, FZJ-Julich, Germany
21. Olaf Wolkenhauer, University of Rostock, Germany

OBSERVERS

1. Ivan Arisi, Lay Line Genomics S.p.A., Roma, Italy
2. Michael Cahill, PROTEOSYS AG, Mainz, Germany
3. Stuart Govan, IEE, Stevenage, Hertfordshire, UK
4. Pekka Ihalmo, Department of Bacteriology and Immunology, University of Helsinki, Finland

Commission Staff

DG Research

Directorate E. - Mark CANTLEY

Unit F.4 - Indridi BENEDIKTSSON; Josefina ENFEDAQUE MORENILLA; Miklos GYORFFI; Manuel HALLEN; Sirje KIVI; Christina Kyriakopoulou, Jacques REMACLE; Henriette VAN EIJL

Workshop Organisers - Frederick MARCUS; Bernard MULLIGAN

DG Information Society

Unit C.4 - Sofie NORAGER, Ilias IAKOVIDIS

WELCOME SPEECH

Manuel Hallen
Directorate F - Health
Directorate-General for Research, European Commission,
Head of Unit: Unit F.4 "Fundamental Genomics"

WHY THIS WORKSHOP?

TO DEVELOP A STRATEGIC VISION OF CSB RESEARCH FOR RESEARCHERS, FUNDING BODIES, AND POLICY MAKERS, BY:

DETERMINING KEY ISSUES FOR THE NEXT FEW YEARS FOR

- Research Topics and New Directions
- Research Policy Makers (EU, National, Regional, International)
- Overcoming fragmentation in research structures and results

AND BY DISCUSSING AND ELABORATING

- Current problems and solutions
- Research policies of the various research and innovation participants
- Specific near term recommendations
- Long term changes
- Future areas of discussions

PROBLEMS OF FRAGMENTATION

The European 6th Framework Programme (FP6) (2002-2006) is aimed at overcoming the fragmentation of European research at all levels. It is a key element of the European Research Area (ERA) initiative.

One area which is a key element in life sciences research, but which is currently fragmented in Europe, is Computational Systems Biology (CSB)

CSB can become an essential element in life sciences by:

- unifying and understanding the results of biological research
- developing common research resources and tools for CSB researchers and laboratory experimentalists
- supporting CSB modelling and research as a goal in itself
- providing a pathway from basic research to development of products, such as new medicines / therapies.

FP6 Priority 1: Life sciences, genomics, biotechnology for health

One of seven major thematic priorities of FP6

The objective is to help Europe generate new knowledge by focusing on genomics and using sequence data and other results to translate it into applications that enhance human health.

Fundamental and applied research will be supported, with an emphasis on integrated, multidisciplinary, and co-ordinated efforts that

- address the present fragmentation of European research and
- increase the competitiveness of the European biotechnology industry.

Major areas of research include:

- Fundamental Genomics
- Applied Genomics and Biotechnology
- Genomic approaches to health and disease
  - Cancer
- HIV/AIDS, malaria and tuberculosis
  - Article 169 European and Developing Countries Clinical Trials Partnership" (EDCTP), concentrating on TB, AIDS and malaria.

Fundamental knowledge and basic tools for Functional Genomics in all organisms (Unit F.4 activities)

KEY RESEARCH AREAS:

TOOLS:

- Gene expression and proteomics
- Structural genomics
- Comparative genomics and population genetics
- Bioinformatics

CSB is not listed as such, but is accessible under bioinformatics

USING ABOVE TOOLS:

- Multidisciplinary functional genomics approach to basic biological processes

EXAMPLES OF FP5 PROJECTS

- Pilot Integrating Projects to overcome fragmentation
  - GENOMEUTWIN: European twins to identify genes involved in disease
  - EUMORPHIA: Study human disease through mouse genomics
  - SPINE: Structural Proteomics in Europe
- COGENE: Co-ordination of Genome Research in Europe
- TEMBLOR bioinformatics projects: Integr8, Desprad, Intact, EMSD

EXAMPLES OF FP6 PROJECTS (IN NEGOTIATION)

CSB:

- COMBIO: An integrative approach to cellular signalling and control processes: Bringing computational biology to the bench
- EMI-CD: European modelling initiative - combating complex diseases
- EUSYSBIO: An SSA on Systems Biology

BIOINFORMATICS

- BIOSAPIENS: A European Network for Integrated Genome Annotation
- ATD: The Alternate Transcript Diversity Project

PLUS PROJECTS WITH CSB / BIOINFORMATICS COMPONENT (e.g. FUNGENES, QUASI)

FP6 TOPICS FOR 2nd CALL IN PROGRESS (deadline November)

- Bioinformatics and Genomics Grid (NoE)
- Integrated Software Platform to tackle genomic sequence-structure-function relationships (IP)
- STREPS available for systems biology topics
- Systems Biology (CA)

FOR MORE INFORMATION

See CORDIS and also FAQ / Training document within our website
ftp://ftp.cordis.lu/pub/lifescihealth/docs/faq_training_2003_07_29.pdf
http://www.cordis.lu/lifescihealth/genomics/home.htm

- LSH-2003-1.1.0-2 CA proposals should focus on structuring European research ... in systems biology...
  - The purpose is to promote and support the networking and co-ordination of research and innovation activities at national, regional and European level.
  - The scale of activities is up to 2-3 million Euros, and duration is typically 2-3 year (more if strongly justified)

WHY WORKSHOPS?

These workshops instigated by the European Commission allows us and the participants to address issues at the European level.
Several previous workshops have been very successful in this respect:

- Mouse genomics
- Diabetes
- Rare diseases
- Cardo-vascular disease
- Bioinformatics

and future workshops are planned:

- ❖ population genetics
- ❖ structural genomics

WHY HAVE WE INVITED YOU?

You are among the top experts in computaional systems biology and/or bioinformatics in Europe, from a range of institutions and managing important facilities.
Many of you are involved in Framework Programme research.
With this small group, we can initiate a full and open discussion.
We hope you can work together to develop a common understanding of the way forward for Europe.

AREAS FOR DISCUSSION AT THIS WORKSHOP

Current and already planned activities in CSB
Future requirements
Structuring European CSB research in the future

CONCLUSION

THANK YOU VERY MUCH FOR COMING TODAY
BEST WISHES FOR YOUR PRESENTATIONS AND DISCUSSIONS!

SUMMARIES OF PRESENTATIONS

SESSION 1 PRESENTATIONS - Hans Westerhoff

Ralf Herwig
EMI-CD-Platform for data integration and modelling of complex biological processes
Collection of data analysis and modelling methods
Genomics yield dozens of targets; now we need strategies to identify the feasible targets
Some modelling is needed, but also more structure of the networks

Luis Serrano: COMBIO - An integrative approach to cellular signalling and control processes: Bringing computational biology to the bench
Needed:
Guidelines for biologist; why and when to use which CSB method
Database of the real 4-D cell
Visualization tools
Sutble modulations
Experimental tools: redesigning 30 % of a cell
Standardisation of software and tools
Reach the experimentalist

Petra Wolff and Frank Laplace (Hepatocon + EUSYSBIO)
German:
Lengthy prediscussion to reach hepatocyte
Modelling platform (3 integrated subconsortia)
2.5 experimental consortia
Close integration
Hepatocon ultimately
EUSYSBIO: SSA
8 workpackages; further Eur SB on all fronts; from government organization through teaching, governments, publishing to science
Preparing Eur SB society? NoE? 'Generation'

Norbert Hübner
Bioinformatics in FUNGENES, an Integrated Project
Concrete aspects needed
Both Genetical genomics and physiological genomics
Embryonic stem cells: novel targets, embryonic development
Focuses on expression networks
Close interaction between experimental and modelling is needed

Stefan Hohmann
CSB in QUASI, a STREP
QUASI: Dynamic operation is important
What is needed for CSB:
Platforms for collaboration biologists-mathematicians
Among experimentalists: sense for quantative data
Standards for data acquisition and for modelling
Suitable experimental model system
Training of people with mltidisciplinary background
Pair students

good existing EU Computational SB approaches?
Partial differential equations, fractal space; one does not know which one to chose, standardisation of tools and software, spatial database, sample base, experimental strategy base, SOP (Luis, Karl, Stephan)

existing programs; WHY INSUFFICIENT?
Partial differential equations, fractal space; one does not know which one to chose, too little, does not appeal to biologist, do not always deal with real problems for biologist; Platforms for collaboration biologists-mathematicians, Standards for data acquisition and for modelling, Suitable experimental model system, Training of people with multidisciplinary background

how integrate Computational - Experimental?
Teaching, pair students, Marie Curie schools, Training of people with multidisciplinary background

how to collaborate?
All view this as important issue: suitable research questions (not too mathematical, not too biological?)

what is lacking from the description: 'for in silico prediction of gene function and for the simulation of complex regulatory networks'
To find new targets; to make new biological discoveries; to solve enigmas

how much of Biology and Medical Sciences should be Systems Biology?
Not all; considerable; growing

SESSION 2 PRESENTATIONS- Ralf Herwig

1. Speakers and schedule
Lilia Alberghina, University Mailand, Italy, “Cellular networks: new tools and approaches”
Olaf Wolkenhauer, University of Rostock, Germany, “Simulating what cannot be simulated”
Mark Sansom, Laboratory of Molecular Biophysics, University of Oxford, UK, “Molecular simulations of membrane proteins: towards a virtual membrane”
Uwe Sauer, Institute of Biotechnology, ETH Zürich, Switzerland, “Computational and experimental approaches in systems-oriented metabolic research”
Age Smilde, TNO Nutrition and food research & University of Amsterdam, Netherlands, “Systems biology as a platform for biomarker discovery”
Karl Kuchler, University & BioCenter Vienna, Austria, “Training and education for CSB”
Alfonso Valencia, Protein design group, CNBC Madrid, Spain, “Possibilities and limitations of the genomic information: biodegradation networks as a case study”
Shoshana Wodak, ULB, Brussels, Belgium, “Bioinformatics requirements for Systems Biology”
Michael Cahill, Proteosys AG, Mainz, Germany, “Embryonic stem cell differentiation: possible elements of a co-ordination action”
Discussion

The purpose of this section was to work out current and future requirements for CSB. What are the actual and future needs in respect to storage and retrieval of biological information (algorithms, computational means), what are the effects of existing and new technologies and implications for databases, standardisation procedures and software. What are the requirements for bioinformatic and CSB tools for potential medical applications.

2. Summaries of the talks

Lilia Alberghina, University Mailand, Italy, “Cellular networks: new tools and approaches”
Lilia Alberghina described research on the control of cell cycle regulation. She pointed out the need for new tools and approaches, in particular new techniques that allow the measurement of as many molecular properties as possible in real time. Ideally, these new tools allow an iterative process between simulations and experiments. On both sides (experiment and theory) there is a fundamental need for standardisation procedures.

Olaf Wolkenhauer, University of Rostock, Germany, “Simulating what cannot be simulated”
Olaf Wolkenhauer described the practice of modelling with noisy data. He stressed the importance of normalisation methods in array analysis. He formulated the need for new technologies for CSB that explain and describe the dynamical aspects of biological systems.

Mark Sansom, Laboratory of Molecular Biophysics, University of Oxford, UK, “Molecular simulations of membrane proteins: towards a virtual membrane”
Mark Sansom showed results on molecular dynamics simulations of biomolecules. He presented concepts for a virtual outer membrane. He formulated the need to scale-up simulations of 100Katoms to millions of atoms. He stressed the fact that simulations should be extended from the cellular level to more complex systems, for example whole organs.

Uwe Sauer, Institute of Biotechnology, ETH Zürich, Switzerland, “Computational and experimental approaches in systems-oriented metabolic research”

Uwe Sauer talked about the interaction of components in metabolic processes. He made clear how wet-lab experiments are used to test hypotheses (e.g. metabolic reactions). He pointed out the importance of the information of function. He defined the unit of function as the molecular flux.

Age Smilde, TNO Nutrition and food research & University of Amsterdam, Netherlands, "Systems biology as a platform for biomarker discovery"
Age Smilde gave an overview of the structure and organisation of TNE and the implementation of a CSB component. A major issue there is the integration of proteomics, metabolomics and transcriptomics data and the coupling of this information. He introduced the underlying concept of "biomarkers" that is useful for medical applications.

Karl Kuchler, University & BioCenter Vienna, Austria, "Training and education for CSB"
Karl Kuchler talked about the importance of CSB as the technology of future. He pointed out that SB is a fragmented area with a lack of integration and co-ordination within EU. There are a few national funding institutions but little participation of EU in other CSB activities. He demonstrated the integrative role of ERA-NET in the connection of these national funding initiatives. He emphasized that we need databases and communication platforms. Training of human resources must be intensified as well as academic programs (e.g. EUSYSBIO, ESBIGH).

Alfonso Valencia, Protein design group, CNBC Madrid, Spain, "Possibilities and limitations of the genomic information: biodegradation networks as a case study"
Alfonso Valencia pointed on the difficulty that analytical methods take time and have to cope with the development of new techniques. He exemplified this with the REGIAdb project. A main issue is the update of databases. He also introduced a new project BioSapiens on comparative sequence analysis.

Shoshana Wodak, ULB, Brussels, Belgium, "Bioinformatics requirements for Systems Biology"
Shoshana Wodak gave an overview of the aMAZE database. She emphasized the aspect of biochemical pathways. A major challenge of the post-genomic era is the linking of databases, for example on enzyme function, metabolic pathways, transcriptional regulation. There is a need to represent the physical interaction between biochemical entities and to map them onto their tissue and cellular localisation. Furthermore, we need tools for querying, displaying and analysing automatically the structure of interactions on various level of granularity. The main bottlenecks are curation, annotation and update of databases. A further need is to develop interfaces from these databases to simulation systems in order to automatically populate dynamic systems.

Michael Cahill, Proteosys AG, Mainz, Germany, "Embryonic stem cell differentiation: possible elements of a co-ordination action"

Michael Cahill pointed out the role of ES cells as a model system for SB and explained strategies for future EU-wide SB activities.

3. Discussion

The topics of the discussion can be summarised in six different points.

3.1 Strategies and future efforts to install CSB funding

Although there are significant overlap of CSB with bioinformatics and biological projects it should have an own significant European funding in the form of a NoE or IP. The network should co-ordinate European CSB research, student training and drive force the integration of different research groups. Besides these efforts, CSB should be a part of appropriate experimental projects. To push the field, there must be some good projects that proof the necessity for CSB research. This will help stressing

the importance of CSB for research and proof that it has practical implications for research and knowledge gain. A third part of funding effort should be available for the development of theoretical algorithms, new modelling tools and theoretical methodology.

3.2 Definition of CSB

Systems Biology is hard to define. It stands between different sciences and has an impact to these sciences vice versa. Thus, it is mainly defined by these interactions. It was discussed whether it is necessary to give a common definition. It might be an issue for the NoE to define CSB, its needs and goals.

3.3 Standardisation

It was emphasized by most of the speakers that standardisation is one of the main issues of CSB. We need standardisation in terms of experimental methods and analysis tools. Experimental design methods must be introduced from the beginning of a project. We should have a few experimental systems with standardised conditions for the test and performance of modelling strategies. The methods to apply depend on the questions of interest and determine the level of understanding of the system.

3.4 Data integration

The interaction of the different levels of information is crucial for successful CSB strategies.
There is a fundamental need for integrating data from proteomics, metabolomics, transcriptomics, and for the coupling of this information. Here, there is a significant overlap with bioinformatics methods and co-operations should be enforced. Furthermore, linking of databases on enzyme function, metabolic pathways, transcriptional regulation is needed. For example, we will have to represent the physical interaction between biochemical entities, map them onto their tissue and cellular localisation. We need tools for querying, displaying and analysing automatically the structure of interactions on various levels of granularity. One of the main bottlenecks is the curation, annotation and update of the databases. Static databases should be linked to forward modelling systems in order to parse the information into dynamic models. We need to create different type of views on the same network of molecular entities. A higher level of integration addresses the integration of model methodology itself by new mathematical methods.

3.5 Education

There is a fundamental need for education in CSB. CSB is the technology of the future but there is still insufficient cross talk and still not enough public recommendation (in comparison for example to bioinformatics). CSB is a fragmented area with a lack of integration and co-ordination within EU. There are a few national funding initiatives but there is a need for EU-wide initiatives and co-ordination. We need student training and academic program.

3.6 What can we expect from CSB ?

From a practical point of view (drug development) it was pointed out that CSB has little impact yet at all. Interesting biomarkers are screened without understanding of the model system, rather in a “trial and error” way. The question was raised, how CSB will tackle with genetic variation that is one the main issues in drug development. The situation at the moment is that CSB has high potentials that must be brought into practical results.

SESSION 3 PRESENTATIONS - Luis Serrano

My session was the third one and therefore many of the things mentioned there, have already been included in the previous sessions. I will summarize in a concise manner the points raised in this session not addressed previously.

1) Eerio Vuorio. He raised the problems of organizing transnational projects and the difficulties of getting outside the national borders to develop new initiatives. There was a comment on EMBL focusing in Systems Biology as the future main research topic.
2) Big Pharma. They raise dthe issue of the small impact of systems biology in drug development, although all of them have a department in this area. However, he also admitted that it takes several years for a scientific field to mature enough so as to be of use for them. A worrying issue is the overlapping between massive data gathering from Pharma and smaller scale initiative sin basic labs. Although there is some openness in some cases, not all the information reaches the public.
3) Heringa. He gave a good summary of different research initiatives in Systems Biology in Holland and particularly in his institute. It raises the point of what is better, several small and focused initiatives or large consortiums like the Geman one.
4) Westerhof. Hans gave a very nice lecture, showing that although we are far away from been able to use systems biology in a routine manner, is it possible already to use it for practical well defined purposes.
5) Le Novere. He made a strong emphasis in the need of having professional software engineers helping developing the software needed for Systems Biology. He described an existing effort about the Systems Biology of a model set of neurons. He pointed out that CSB is about kinetics, a dimension often unreachable by wet-biology alone.
6) NotreDame. Cedrid emphasized that although modelling and data gathering are crucial for Systems Biology we should not forget about Databases, and also he made a good point about establishing objective criteria to asses simulation software, as well as biocomputing functional prediction.

GENERAL SUMMARY OF PRESENTATIONS - Jaap Heringa

Ralph Herwig:
Data integration is difficult. Nature paper 2002
Luis Serrano:
Need new in vitro tools to simulate and test metabolic pathways – currently we can stick up to 7 genes into a yeast chromosome, but not more.
COMBIO project (with Wodak, Valencia, BIOBASE GmbH, etc.)
Database and visualization
Mathematical modelling of network dynamics
Would like database with cell compartment info through time: where are compounds through time
Comment Foord: they have prot interaction data where sometimes the proteins are not even in the same organelle! So need this info to filter interaction data.
With all the new data sets, there is a major problem (understood by everyone) with data acquisition, upgrading and annotation leading to error propagation.
How many parameters do we need to measure and how many can we guess? If the model is good can we interpolate more parameters? This has to be found out
Petra Wolf:
Lots of omics with generated technology platforms.
Focus on human hepatocytes (form about 80% of liver mass) – long term goal: the virtual hepatocyte.
Aims:
Generation of quantitative data
Analysis of functional relations
Establishing standards
BMBF funding in 3 areas:
Cell biology, modelling and tool box

2nd part of talk:
EUSYSBIO Fp6 project – multi team project – Hans Wff is in the list.
Is all about teaching and marketing, and coming out of the discussion also in vivo experimentation.
Comment Serrano: Why hepatocytes? This is a complex cll. Is E. coli not better?
Generally: How to select model system. What do you want to understand?
Another identified problem: how do we get national funding across EU borders?
Norbert Huebner
Bioinformatics in FUNGENES, an IP
Genomics of transcription regulation. – cell differentiation and lineage commitment for identifying novel therapeutic targets.
ES cells:
Mouse embryonic stem cells – entoderm, mesoderm, endoderm
Integration of approaches and omics data.
ES cell pluripotency can be kept if differentiation is delayed.
Accurate models of binding specificities of TFs, global RNA expression profiles for specific cell types
Comparative sequence analysis – we need the new genomes
Future tools:
Intergenic region arrays for all model organisms
Tf characterization tools:
Libraries of expression constructs for tagged TFs
No data on competing binding sites for the same TF is included in the proposal.
Comment: how definable is the ES system? Cancer cells can give completely different results in vivo or in vitro. Can we model enough to understand.
Stephan Hohmann
Transcriptome: always the entire transcriptome or a (well defined) part of it.

SESSION 2
Lilia Alberghina
Cellular networks
Paul Nurse (Nature 2003) The intracellular communication involves switches, amplifiers, feed-back loops, timers, etc.
Sic1 protein docks into Cdk1. There is Sic1 and phosphorylated Sic1. Sic1 has much higher affinity for binding to Cdk1 than Sic1 201P (phosphorylated).
Olaf Wolkenhauer
Simulation.
Systems dynamics give rise to molecular function (paradigm shift).
Signal or system oriented approaches.
In systems biology you can NOT just take data from other sources. You have to be there and know e.g. under what conditions the data were measured.
Feedback regulation.
Models so far are poor at spatial description (e.g. differential equations).
They learned most from discussion about models and failed models. Models shape experiments.
Uwe Sauer – ETH
We need models to predict system behaviour.
From experiment->model->in silico representation-> prediction->experiment (round)
Example phophofructokinase ATP <-> ADP, A[D/T]P is needed at hundreds of points in the metabolic pathways. What is the function? Relates to the whole network.
Challenges:
Higher-throughput
Resolution of dynamics
Higher cell types
Integration with genome-wide compositional data
Advance from analytical to predictive computer models
Standardisation of methods/strains/data
Age Smilde – TNO
Involved in 2 genomics projects: CMSB and Kluyver Centre
Models: Differential equations at the one hand and data-driven models on the other hand. There must be some modelling system in between.
Karl Kuchler – Vienna
Finland, Germany and Netherlands have national funding agencies as the only EU countries.
EUSYSBIO and ESBIGH
Do SB teaching, European SB course every year, a practical course every other year.
Alphonso Valencia
REGIAdb: EC project (>30 labs) on TFs in Arabidopsis
It has been difficult to make a database while collecting the data. The data structures change all the time and this is a nightmare for the databases.
Data acquisition
Data management (databases)
Complex data (integrated DBs)
Ontologies
Biosapiens
Sequence comparison and inference are limited.
Comparing network structures
Text mining (from papers) is not finished but is still an emerging field.
Microbial diversity is a lot greater than higher species diversity.
The biodegradation network is big and includes:
? Different species (multi-species network)
? Different activities
? Different conditions

? Large part unknown

Shoshana Wodak
Metabolic networks: thousands of molecules interacting
Databases need tools to analyse the information.
Tools for querying, displaying & analyzing automatically the structure of networks of interactions.
aMAZE: representing biological function as networks of molecules & interactions. Started at EBI, now at ULB (Brussels)
A rich Object Oriented Model. A graph representation of metabolome, regulation networks, and signal transduction pathways.

Kohn maps are used a lot nowadays to represent signal transduction. But this representation does not allow analysis. You can only look at the icon representation.
Michael Cahill - Embryonic stem cell differentiation: Possible elements of a co-ordination Action
Stem cells. Company: ProteoSys.
Predict that conserved embryonic systems correspond to very conserved mechanism.
Cladistics.
SESSION 3

Cedric Notredame
DB TARGET project -- identifies candidates for structure determination
Model evaluation and certification in special institutes (groups)
For November:
NoE, IP, STREP, Co-ordinated action.

GENERAL DISCUSSION - KEY POINTS

Function – We discuss it a lot – we should define function early. We must be much more careful .

Models – Huge data sets should be organised, what kind of models do we want? Other models are just predictive – pure black box.

Sys Bio goes beyond huge data sets. Elucidates functional models – perspectives – model systems – we need model systems Standards – they already exist in many areas. We say what we need. Training

The issue of standardisation was perhaps a bit overrepresented at the workshop. It is certainly important, but different points that pertain to this subject were mixed in the discussion, as one might expect when people from different backgrounds discuss. There was a broad consensus that large-scale data sets are immensely important and will continue to keep us busy in bioinfo and sys biol. It should be emphasized though, that systems biology approaches can also work with small but quantitative data sets. It is not always necessary (or even advised) to do loads of chip experiments. The next rounds of proposals will undoubtedly contain a mixture of projects that use the term systems biology will in fact they do large scale data analysis. While this is very important all by itself, it is not systems biology unless tied to predictive models of some biological detail and experimentation for hypothesis-testing.

Projects: CSB should be standard part of projects

Journals: Launching new journal in a vibrant field is essential.

Topics in Sys Bio: in cellular and super-cellular aspects. Link key initiative streps and trans-national kinds (a) integrated efforts Expt, bioinfo, CSB, (b) model organisms (c) both specific and generic (included as nodes). What is output? 1) standard, validated datasets, as basis for analysis. 2) Evaluate different simulation methods. 3) Common software. 4) Training. We must move forward, but keep with biologists. Must produce science, Best is to get good results with good focus. Small projects are not enough. We need larger projects, e.g. co-ordinating networks, and training. Keep focus, science, small and large.

National programs: We should identify and integrate programs – communication – expt. and comp – Built on excitement.

This workshop: For the Commission, such events provide much more feedback on the future research than any typical larger Conference or Meeting. There is a consensus in the scientific world that the Systems Biology offers unique opportunities to reach ambitious goals both in the basic science and in the design of a new generation of drugs. As this field requires large financial resources, multi-disciplinary expertises and complex trans-national research programs, it is essential to correctly choose and thus limit the biological systems to be investigated. In thus respect the European Commission may play a crucial role, as a multi-state institution but at the same time actively involved in building a trans-national research policy.

Support for CSB: CSB needs more support, co-ord and money. We need European platform to develop field. One instrument – ERANET and Forum of Research Managers – Important goals – common announcement and funding measure in 2005. In Germany, Oct, start planning CSB on bacteria, announcement in 2005 – Indeed in discussion.

Formalisms: We need formalisms for data modelling and techniques – biostats very important.

Industry: We should be interested in applications – need to think about industry.

CSB and International Collaboration and Competition: The development of Systems Biology as a recognised field of research in life science is actually a race. In this respect, Japan and United States, which launched their first large-scale projects in 1998, acquired a clear advantage. Moreover, the setup of serious projects in Systems Biology necessitates several years of planification and development.

Therefore, it is of utmost importance that European Union enters the competition immediately, if it means to be an influencial player in the future.

The role of CSB versus and in Systems Biology Projects: Computational Systems Biology is a mandatory part of any serious Systems Biology project, that otherwise generate only databases, without new knowledge and understanding of life, and little biomedical prospects. Systems Biology concepts are directly related to engineering, and could be viewed as applications of cybernetics to biology. As such, many tools are already in use within the field of Computational Systems Biology. However, most often they implement a single algorithmic approach (logical computation, stochastic - Monte-Carlo - approaches, ordinary differential equations, partial differential equations etc.). In addition, those tools are often unable to directly extract the numerical values from databases. Their use for large-scale reconstruction of living systems is therefore a problem. Cooperation between theoreticians, software engineers and experimental biologists should lead to the development of more powerful simulation platforms, modular and user-friendly. CSB is about dynamic interactions. Need to spread good practice.

Systems Biology - a large field: It will become important as part of many Biology projects. This tends to make it seem vague, but it is not. Once a field has become useful it tends to connect with other fields and seems to be vaguer. Compare Cancer Research, and Bioinformatics. One may think that these fields are vague, but they are not; they are wide; they define themselves by what they do, and this makes it harder to have them defined by the human mind, like the odd numbers can be defined. What is required for such a broad new discipline that runs the risk of suffering from European fragmentation, is a Network of Excellence according to the original formulation of the instrument, which will make the existing and new SB initiatives synergize, and allow what now remains artificially national to become transnational-within EU. The NOE may seem a bit vague because of this, but there will be tremendous synergy. The components of the NoE will not be vague; they will be as specific as the German hepatocyte programme, and the Dutch L. lactis growth rate program, and a number of FP6 STEPS and IPs. There is a worry about project evaluation committees looking at the SB grants. IF these will consist of the mainstream molecular biologists, then we shall end up with MolBiol grants disguised as SB grants. Solution: get non competing UDS and Japanese System Biologists in to judge, because most European System Biologists will want to submit themselves)

Commission Priorities – The Commission should recognise importance of CSB. We should think big in this area. Multi-disciplinarity is key to CSB, proposals should have this. We also need CSB PH.D. , midex expt. and computers – e.g. Marie-Curie and high level conferences and national training sites – model systems with expert tools, for knockout tools .

Pharmaceutical Industry: Big Pharma does a lot of CSB, but no impact yet. We can get data by integrating, we could do much better. Just do it.

Visualisation tools: To do CSB, we need new experimental tools to visualise.

CONTRIBUTED PAPERS

SESSION 1 - THE FOUNDATIONS - Current and already planned activities

Hans V. Westerhoff
System Biologist, European Union

Chairman session 1; Introduction

Recent challenging readings on Systems Biology

Lazebnik Y.(2003) Cancer Cell. 2:179-82. Can a biologist fix a radio?--Or, what I learned while studying apoptosis.

Henry, C.M. (2003) Chem & Engin. News 81, 45-55. Cover story: SYSTEMS BIOLOGY. Integrative approach in which scientists study pathways and networks will touch all areas of biology, including drug discovery http://pubs.acs.org/cen/coverstory/8120/8120biology.html

The essence of System Biology:

From interactions to life; repeatedly
Systems Biology internationally; examples
Leroy Hood: ISB (Seattle)
Al Gilman Alliance Cell Signalling
E-cell (Japan)
ERATO Kitano (Japan/CalTech)
Virtual cell (modelling software)
Bernard Pallson, UCSD
Roger Brent
International E. coli Alliance (Science August '02)
ICSB 2000, 2001, 2002, 2003, 2004
Cf. www.systembiology.net

Where is Europe?

European Systems Biology
(avant la lettre); examples
Zhabotinsky,Turing patterns + Prigogine/Hess school
Glycolytic oscillations (Duysens)
Chemiosmotic ATP synthesis (Mitchell)
Metabolic Control Analysis (Kacser, Heinrich, Groen)
Phosphoneural net signal transduction (Hellingwerf)
Silicon cell
Many more …….
…….

Where is Europe?

It is gearing up…….
ICSB2000, 2001, 2002, 2003, 2004
Silicon cell
Industry: Wide interest biotech (Unilever, DSM, …) and Pharma (Bayer, GSK, Novo, AKZO, …) [but not married to Europe]
German hepatocyte
Specific Support Action EUSYSBIO/ESBIGH
Cf. www.systembiology.net

European Union: FP6 ideas
Developing bioinformatic tools and resources for data storage, mining and processing; Developing computational biology approaches for in silico prediction of gene function and for the simulation of complex regulatory networks
Enormous potential for European Systems Biology
Best scientific foundations
Proven history of accomplishments through collaboration of excellence
Many national programs
Catalytic funding of synergy will be highly successful

Where should Europe be going?
Session 1: Foundations
what are good biology knowledge bases ?
good existing EU Computational SB approaches?
limitations to present computational approaches?
existing programs; WHY INSUFFICIENT?
how integrate Computational - Experimental?
how to collaborate?
what is lacking from the description: 'for in silico prediction of gene function and for the simulation of complex regulatory networks'
how much of of Biology and Medical Sciences should be Systems Biology?
Inventory of Europe taking off (part one) as introduction to formulating the great leap forward

CoSyB
'developing computational biology approaches for in silico prediction of gene function and for simulation of complex regulatory events.'
Network of excellence creating a single European Computation Systems Biology work environment (virtual centre of excellence) directed at creating synergy between excellent European consortia that engage in Systems Biology research.
CoSyB

The NoE centre should consist of the consortia plus extra coordination and work force placed between them. At the centre (i) computation and bioinformatics methods should be optimized for systems biology, (ii) an active interface should enable experimental groups to identify which systems biology tool is best for their purpose, (iii) projects should be (funded and) carried out between the centre and those experimental groups to carry out the first phase of CSB on those projects.
CoSyB

The centre will elaborate, fine tune and validate CSB methods on a number of well-defined experimental model systems. Among the latter there may be living cells where the centre should take the international lead (e.g. hepatocyte, L. lactis) and organisms where Europe can play an equal major role (e.g. E. coli). Methodology may include: silicon cell modelling, control analysis methodology, flux analysis methods, integrative bioinformatics.
CoSyB
The centre will also coordinate the advising of standards for modelling, computation, data analysis, data storage, experimental methodology. It will also be involved in the generating of a 'sample base', in which of a number of model organisms under well defined conditions, standard samples will be taken, on a large scale, to be aliquoted and stored, for assay by all interested scientific groups now and later.
The centre will also be a catalyst for the training of human capital.

Ralf Herwig
EMI -CD
– a platform for data integration and modeling of complex biological processes

1. Introduction

Genome research has seen fundamental technical breakthroughs in recent years such as the sequencing of the human genome (Lander et al. 2002, Hattori et al. 2000) and the genome of other species serving as experimental model systems. The main sociological and economical impact of genome research is the molecular understanding of major human diseases and the development of new therapies and medicals for the combat of these diseases. However, despite the fact that there was a nearly three-fold increase of pharmaceutical investment in R&D in the time period 1992-2001 from 11.5-30.5 billion USD, the number of newly filed molecular entities has been fairly constant as pointed out by several pharmaceutical researchers recently (BIO 2003, PhRMA report 2001).
One possible reason for this development might be the fact that analytical methods and tools are not yet significantly installed in the drug development process. While bioinformatics methods are well incorporated in the first part of this process (drug target discovery), this is not the case for the later stages. In particular, the simulation and modelling of biological processes such as disease-relevant signaling pathways and metabolic processes are under-developed in drug target validation. Nevertheless, computational methods are needed here. In contrast to the early 90s where target discovery was a main problem, nowadays the number of potential drug targets has increased to a large extent leading to an unfeasible number of targets and to excessive costs in drug development. For example, the R&D costs per drug have increased from 95 million USD in 1982 to almost 880 million USD in 2000. A fundamental challenge is thus, to search through this exhaustive set of targets and separate feasible from unfeasible ones. Here, in silico experiments can be the basis for a successful screening within the drug discovery process and the entire drug development process should be accompanied by bioinformatics and systems biology approaches especially by the introduction of simulation techniques and experimental design in all phases of the process.
Furthermore, the need for integration rules and methods is fundamental in current functional genomics research (Kanehisa and Bork 2003). Multiple databases exist, a variety of experimental techniques have produced gene and proteome expression data from various tissues and samples and important disease-relevant pathways have been investigated. Information on promoter regions and transcription factors is available for a lot of genes as well as sequence information. This information - although extremely helpful - cannot be utilized in a sufficient way because of the lack of integrative analysis tools.
In our project we attempt to develop a software platform that is able to meet some of the above requirements. The software platform bases on three layers and will connect and implement several modules necessary for the in silico modelling process. In the first layer information is gathered on the biological objects under analysis and experimental measurements on these objects are integrated. An analysis layer will translate this knowledge into biological networks. Using probabilistic learning methods (e.g. Bayesian networks) these networks will be expanded in light of all available data on the objects. In a simulation and modelling layer these network hypotheses will be evaluated and predictions of experiments will be produced which have a direct feed-back to the forthcoming experimental design and experimental verifications.

2. EMI-CD – platform technology

The analysis of processes involved in the course of multi-genic diseases has to cope with data from diverse experimental (functional genomics) platforms such as gene expression data (DNA arrays, RT-PCR), protein expression data (MALDI-TOF, 2D-gels), functional sequence data (gene ontologies,

annotation databases), physiological data (patient information, phenotype information), environmental factors and many others. Thus main elements of the software platform will target data integration and data standardisation.
EMI-CD platform is designed in a modular way. Main modules are listed below.

Database integration. The platform will have an interface to the SRS annotation system (LION Bioscience Ltd.). This will allow exhaustive information mining of data from many diverse sources such as functional sequence data, gene and protein annotation, genome annotation, regulatory and metabolic pathway data, ontologies, protein families, biological reagents, protein structure including information about active sites, binding sites etc. Furthermore, we will incorporate inference of knowledge on given pathways, in particular the pathway database developed in the EBI-ENSEMBL group. This database provides interaction data of groups of genes in a qualitative way (www.genomeknowledge.org) and will be used to populate the models and to generate hypotheses. A further web-accessible information system to be linked is the *Genome Matrix* developed at the RZPD Berlin in co-operation with the Max-Planck-Institute for Molecular Genetics (www.genome-matrix.org). *Genome Matrix* provides gene-based visualization with links to various databases and resources. Genes can be visualized in chromosomal neighbourhood as well as throughout different organisms. This allows cross-species comparisons for example conservation of gene function in human, mouse, rat, C.elegans, Arabidopsis, Yeast and others.

Experimental data integration. A critical issue is the integration of experimental data. Typically, the objects of a biological system are measured on different levels of cellular information and with different biotechnical procedures. Here, our approach consists of three steps (MicroDiscovery GmbH Berlin). In the first step experimental data will be normalised and correlated on the single object level. As a recent example, we set up a gene expression catalogue of mouse orthologues to human chromosome 21 genes (Gitton et al. 2002). This catalogue contains transcriptome data measured with RT-PCR, in silico EST-mining, and localised expression images by WISH technology (http://chr21.molgen.mpg.de/hsa21/). These data have been correlated and evaluated with respect to their consistency. Additionally, available database information of the biological object will be collected. Although useful, the linking of databases and information resources is only the first step of data integration. In order to gain new knowledge, we have to develop integration rules and methods for translating these data into a biological object network (for example a genetic network). Here, data from different techniques must be integrated and correlated and data conflicts must be handled.

Data analysis. Various models and analysis techniques have been suggested, primarily focusing on gene expression data. Prominent current expression analysis tools are based on clustering (e.g. Herwig et al. 1999). Such analyses successfully reveal genes that are co-regulated, but not their regulatory relations. More advanced approaches rely on mathematical models of the regulation process. Different models at various levels of detail have been suggested. These include Boolean, qualitative, linear, differential equation and detailed biochemical models. The model will include genetic and metabolic regulatory systems, multi-cellular signaling, phenotypes under different stimulations, protein-protein interactions and more. The ultimate goal of the inference process is to derive novel hypotheses on the mechanistic nature of the disease under study. 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 GENESYS software developed by the Computer Science group of the Tel-Aviv University and will be connected to the system (Tanay and Shamir, 2001).
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 object-oriented simulation tool is implemented as a *Python*-product into the Zope web application server environment. It entails hierarchically structured models of biological systems in relation to their cytological cell structure. Biological objects can have different attributes (e.g. initial concentration,

sequence data, kinetic constants) and actions that represent a single reaction or groups of similar reactions. Each action holds the stoichiometry and a kinetic law. Based on this data ODE-systems can be created automatically and used for simulations. The model population is done by a web interface based on different information resources. An automated connection to the KEGG database allows including information about metabolic pathways and compounds. Furthermore, the import/export of models via SBML enables compatibility with other systems. The system allows the incorporation of metabolic networks and signaling 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.

3. Outlook - Impact on Systems Biology

The purpose of EMI-CD is to provide a software platform that is complex enough in order to cope with various experimental techniques for the discovery of gene function and the understanding of disease processes. A main issue (upon implementation of the platform) will be to co-operate with experimental projects on the design of experiments for combined strategies on the combat of human diseases (for example diabetes).

Compatibility with other systems will also be an issue. By the use of SBML models can be interchanged between different systems. A further issue is the scaling of the platform to large systems (whole cell models). At the current stage systems with a few thousand reactions are computationally feasible. However, more advanced ODE solver and the increase in computer hardware are needed. EMI-CD will be an open system for the integration of advanced analysis tools and other database systems. For example, a future issue (planned research activity) is the connection of the platform to GRID-technology.

4. Literature

Gitton, Y., Dahmane, N., Baik, S., Altaba, R. (Group 1), Neidhardt, L., Scholze, M., Herrmann, B.G. (Group 2), Kahlem, P., Ben Kahla, A., Schrinner, S., Yildirimman. R., Herwig, R., Lehrach, H., Yaspo, M.L. (Group 3) (2002) A gene expression map of Human Chromosome 21 orthologs in the mouse: a step towards a molecular understanding of the Down Syndrome phenotype. *Nature*, **420:**586-590. (all groups contributed equally).

Hattori, M. et al. (2000) The DNA Sequence of Chromosome 21. *Nature* **405:** 311-319.

Herwig, R., Poustka, A., Müller, C., Bull, C., Lehrach, H., and O'Brien, J. (1999) Large-scale clustering of cDNA fingerprinting data. *Genome Research* **9:** 1093-1105.

Kanehisa, M., and Bork P. (2003) Bioinformatics in the post-sequence era. *Nature Genetics* **33**: 305-310.

Lander, E.S., et al. (2001) Initial sequencing and analysis of the human genome. *Nature* **409:**860-921.

Tanay, A., and Shamir, R. (2001) Computational Expansion of Genetic Networks *Bioinformatics* **17**:S270-S278.

Luis Serrano -

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

I COMBIO: Objectives

- The major objective of our proposal is 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.

II COMBIO: Objectives

- Global quantitative understanding of the p53/Mdmd2 network.
- Global quantitative understanding of the dynamics of spindle formation.
- Construction of a database for simulation and visualization tools
- Production of a standard guide for experimentalists showing which simulation tools should be used, depending on the problems been addressed and on the information available.

III COMBIO: Participants

- *European Molecular Biology Laboratory EMBL.*
- Luis Serrano, Francois Nedelec, Isabelle Vernos
- *Centro Nacional Biotecnologia CNB*
- Alfonso Valencia
- *Weizmann Institute*
- Uri Alon
- *Universite Libre Bruxelles ULBCR*
- Shoshana Wodak, Marceline Kauffman
- *Centro Nacional Investigaciones Oncologicas CNIO*
- Cayetano Gonzalez, Amancio Carnero
- *University of Göttingen. UKG*
- Erik Windenger
- *Budapest University of Technology and Economics BUTECR*
- Bella Novak
- *BIOBASE GmbH*

IV COMBIO: Experimental Systems

- **p53/Mdmd2 (Amancio Carnero, Uri Alon, Luis Serrano)**.

IV.1

IV.2

- **Dynamics of spindle formation (Cayetano Gonzalez, Isabelle Vernos, Francois Nedelec, Luis Serrano).**

V COMBIO: Database & Visualization

- Shoshana Wodak
- Alfonso Valencia
- Erik Windenger
- BIOBASE GMBH
- In the framework of this project, these partners will pool their tools and expertise to produce a public domain resource, which handles all the information compiled in this project on metabolic pathways, protein-protein interactions, gene regulation and signal transduction network relevant to the processes of interest. This resource will feature a modular web-based user interface, which will enable flexibly querying and visualising the information (including interactive display of

meaningful pathway charts), custom-building of analyses tools and interfaces to simulation programs, as well as tools for the annotation of data on interactions and pathways

VI Databases used in Combio

- aMAZE, a public domain database system, implementing a formal model, which embodies general rules for associating individual biological entities and interactions into large complex networks of cellular processes (van Helden et al, 2000, *Biol Chem* **381**(9-10), 921-35; van Helden et al, 2001, *Briefings in Bioinformatics* **2**(1), 98-93). This system can deal with a large variety of processes such as metabolic pathways, protein-protein interactions, gene regulation and transport, as well as signal transduction.
- (BIOBASE) will provide with sets of relevant interaction as they are stored already and as they will be additionally updated in the TRANSPATH database during the project (commercial system for representing information on gene regulation). In doing so, particular attention will be paid on a clear documentation of the quality of the information manually extracted from literature.

VII **COMBIO:** Testing of different simulation strategies

- Uri Alon
- Marcelle Kaufmann
- Bella Novak
- Francois Nedelec
- Luis Serrano
- Mathematical modelling, together with simulations and computational approaches, can provide the necessary framework for integrating data and gaining insights into the dynamical and functional properties of complex networks. A good evaluation of the best-suited approaches for the modelling of a given biological process is in general, however, a difficult task for most biologists. An important effort will be devoted in this project to provide a number of selection criteria facilitating an appropriate choice according to the objectives and available level of information.

The SSA EUSYSBIO – The take-off of European Systems Biology

Petra Wolff

Projektträger Jülich PTJ (Project Management Organisation),
Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany

Introduction

Thanks to the recent spectacular advances in the "-omics" disciplines and in information technology (IT), the biosciences are heading for another revolution: the in silico simulation of complex life processes. This branch of research, termed systems biology (SB), combines concepts from molecular biology, engineering sciences, mathematics and IT in a holistic approach to complex biological systems, for example living cells. SB is currently being promoted intensively in ambitious funding initiatives in particular in the United States and Japan.

So far there is no competitive Europe-wide networked research in SB. However, there are several renowned research groups in the SB field across Europe. It is necessary to close this gap as soon as possible and to include relevant activities in related disciplines.

Objectives

EUSYSBIO makes an active contribution to establishing a European Research Area for Systems Biology (SB). Within the framework of FP6, the proposed project is to provide a first component for an internationally powerful network which will secure a top position for Europe in this new research area.

Recent years have shown that the fragmentation of the European research landscape is one reason why Europe has been very slow in catching up with the United States in the field of the life sciences. While existing competences were effectively concentrated in large research clusters on the other side of the Atlantic and have developed into prospering and efficient centres as a result of generous public and private support, structural weaknesses and divergent national funding structures have paralysed the establishment of a European counterweight – despite the fact that Europe has rich resources in terms of highly qualified personnel and public funding.

In Europe, first national research initiatives have been launched recently by the German BMBF[4], the Academy of Finland and the Dutch NWO[5]. As a result of the start of FP6 and EUSYSBIO, existing gaps in the European research landscape in the field of SB can progressively be closed. In order to achieve this goal, EUSYSBIO provides for politically, administratively, industrially and scientifically oriented WPs.

Work programme

An important initial step within the framework of EUSYSBIO will be a **widely acceptable, clear definition of SB** which is accepted by the scientific community. Although the term "SB" has experienced a real renaissance during the last few years, it is still not clear what the specific scientific methods, the goals to be pursued and the characteristic features behind this concept are. In particular

[4] Bundesministerium für Bildung und Forschung/Federal Ministry of Education and Research

[5] The Netherlands Organisation for Scientific Research

for the layman, the dense jungle of unintelligible terms, *e.g.* SB, systems biotechnology, biocomputing, computational biology, *in silico* biology and theoretical biology, is to be cleared to provide better orientation. On the one hand the definition must leave sufficient scope and flexibility for the development of SB, while on the other hand it must, as a definition, be identifiable and comprehensible for non-experts and SB experts alike. What is, however, **most important is that all participating actors should speak a common language** and that the representatives of the diverse disciplines involved can identify with SB. This will form an important prerequisite for creating, expanding and strengthening a European SB network.

Yet another component of EUSYSBIO is a **Status Quo Survey** as well as an **intensive international benchmarking** to enable further action platforms to be set up. **EUSYSBIO's aim is to identify available scientific competences in Europe, to concentrate and network these to create synergies** and thereby to launch a broad-based SB offensive on the basis of existing scientific excellence.

The results of the survey and the benchmarking will be made available to the interested public as a brochure as well as online. Once the stakeholders from science, politics and industry have been identified, they must be brought together in **pan-European platforms**. These forums are to be established within the framework of EUSYSBIO and, if possible, to be converted into effective long-term structures on conclusion of the project.

At the scientific level, a network of experts was built up in the course of the Expression-of-Interest phase (ESBIGH[6]) which unites the best brains in the field of SB and relevant disciplines such as biology, IT and engineering sciences under one roof. **The establishment and extension of this network of scientists** will be further promoted within the framework of EUSYSBIO. One aim is to also include competences from **Eastern European countries, EU candidate countries, Russia, countries of the Western Balkans and East Asia** on a long term basis, as these countries are expected to have great potential in terms of highly qualified scientists (mathematics, bioinformatics, systems engineering, etc.).

In order to achieve this goal, EUSYSBIO will promote the **international networking of various activities** in the field of SB (European and non-European ones), and this will include the **National Contact Points for FP6 all over Europe** (NCPs) as well as relevant activities in other priority fields.

Furthermore, EUSYSBIO will set up a **Policy Maker Forum**, within which representatives of relevant national and European research funding organisations can exchange their views and deliberate on strategic and funding policy considerations regarding the further development of SB. EUSYSBIO will make a decisive **contribution to co-ordinating national and international SB activities**. A basic prerequisite are **regular exchanges of information between all the partners**, which are to be ensured by means of workshops organised by the forum and supported by a **web-based EUSYSBIO information platform**.

An **attractive, user-friendly EUSYSBIO homepage** will provide news and information on SB and also facilitate access to other web portals. In this way a broad interested public (including the press, young scientists, etc.) will be given easy access to information on contacts, international/national/EU activities, events, continuing education measures and results of EUSYSBIO.

---

6 European Systems Biology Initiative: Genomics for Health; see:

http://www.systembiology.net/esbigh/description.html

SB is a young branch of science, which due to its cross-discipline and cross-border character and the combination of computer science with biology arouses the interest of young scientists in particular. This is confirmed by the good participation of students, doctoral students and postdocs in conferences (*e.g*. ICSB), tutorials and workshops. It is an optimal prerequisite for opening up a new, innovative field of science, which needs an inflow of new brains and ideas. The HFSP[7] programme has also dedicated its current call for proposals to the topic of complex biological systems. **The training opportunities integrated into EUSYSBIO offer an FEBS[8] course and a Marie-Curie fellowship programme on SB**, which are in the long term to be developed into a major EU training programme supplementing existing training measures effectively.

As in all other fields of the modern life sciences, SMEs play a central role in providing innovative services, platform technologies and new products. There are many European SMEs today which are active in the "-omics" disciplines or in IT and satisfy the great demand for customised individual solutions in science and industry. In the environment of SB there are some internationally well-positioned SMEs, (e.g. Physiome Sciences Inc., Lynx, Lion Bioscience, etc.) which offer intelligent IT solutions, simulation software, HT systems[9] and successfully occupy specific niches in the international market. Big companies in the field of IT such as IBM and Sony and in the field of the pharma industry (e.g. Eli Lilly, Glaxo Smith Kline Beecham) increasingly engage themselves in this field. **SB will offer interesting prospects for SMEs, so that the industry is to be involved in SB early on**. EUSYSBIO will pursue this goal by setting up an **SME platform** to ensure the **transfer of information** between relevant European SME associations and interest groups as well as European SMEs and big companies.

As a highlight of EUSYSBIO, the **International Conference on Systems Biology ICSB 2004** will offer an international forum for all participants of EUSYSBIO in autumn 2004. Within the framework of ICSB, the results of the project will be presented to a broad public in the form of talks and print materials. ICSB will be the largest SB conference, in which about 500 guests from the international community of SB and industry will take part. Young scientists are to be addressed by specific tutorials, poster sessions and lecture series oriented to their specific needs and interests. It is planned to organise a get-together of all EUSYSBIO actors in the margins of ICSB to enable EUSYSBIO participants to exchange their experience and to deliberate on a continuation, if any, of the activities launched in the form of suitable measures within the framework of FP6.

[7] Human Frontier Science Program; see: http://www.hfsp.org/

[8] Federation of European Biochemical Societies (FEBS)

[9] High Throughput

The German Research and Funding Program "Systems Biology"

Frank Laplace[1] and Petra Wolff[2]

[1] Federal Ministry of Education and Research
Heinemannstr. 2, D-53175 Bonn-Bad Godesberg, Germany
[2] Projektträger Jülich PTJ (Project Management Organisation),
Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany

Introduction

The 'Systems of Life – Systems Biology' programme represents a new approach to research funding with a view to establishing systems biology in Germany and tapping its potential for future research and development in science and industry. To this end, an interdisciplinary network of centres of excellence will be developed which will weave the biosciences together with systems studies, engineering sciences, computer science and mathematics to form systems biology. To ensure the medium- and long-term focus of the network of centres of excellence the individual research activities should be conducted in collaborative projects (i.e. projects involving science and industry) and concentrate on the hepatocyte as a model system. This type of cell was selected because of the central function it fulfils in higher-order organisms and its great importance for medicine. This includes the numerous essential metabolic processes that take place in the liver, the catabolism of drugs and toxic substances, metabolic adaptations, the transport and processing of substances as well as differentiation and the ability to regenerate. The expectation linked to the 'Systems of Life – Systems Biology' programme is that in future methods of quantitative biology will be used more extensively in all disciplines of the biosciences.

Definition of systems biology underlying the programme 'Systems of Life – Systems Biology'

Over the past 30 years biology has been characterised by a qualitative and descriptive approach designed to investigate molecular details. But in order to understand complex systems properties such as optimal regulation, adaptation and memory both the components of a system and their interactions need to be studied.

Systems biology will develop has chosen this holistic approach. It aims to look into biological processes at the systems level by analysing the complex networks which produce complex functionality. This task requires a shift towards quantitative biology. It is only by means of a quantitative description of a system's components and their interactions that the behaviour observed can be understood. To attain this objective it is also necessary to take a theory-based approach to a complexity which can no longer be understood intuitively. This is why mathematical modelling plays such a central role in systems biology. Consequently, the effort to develop a quantitative understanding of biology at the systems level is based on an interdisciplinary approach combining concepts borrowed from biology, computer science and systems studies. The ultimate goal of systems biology is to develop virtual representations of cells and entire organisms. These representations will then enable computer experiments to be conducted in analogy to experiments involving real biological systems. This can pave the way towards predictive biology which – among other things - will help to understand and treat diseases in man.

Research under the Systems of Life – Systems Biology programme will focus on the hepatocyte system

The model system of the BMBF's funding activity on systems biology is the human hepatocyte. This model was selected with regard to its central function in higher organisms. The liver is a highly complicated "biochemical factory" which synthesizes, converts or breaks down more than 10,000 substances a day. This includes numerous processes related to the utilization of food as well as other essential metabolic functions. The liver supplies the organism with necessary substances such as proteins, carbohydrates and fats, which are absorbed from food. The hepatocytes are largely responsible for the many different metabolic pathways and their control. They make up about 90 per cent of the liver mass. They were chosen as the model system because of their various vital functions and their manifold potential applications in medicine, pharmaceutical research and other areas.

The liver fulfils many essential functions. As a storage organ, it accumulates reserves of important substances. From the organism it eliminates toxic substances and pathogens ingested, for example, with food and converts them into harmless substances. As a gland producing nearly one litre of bile per day, the liver supports the digestive process. Many metabolic waste products are released into the intestine and eliminated together with the bile while numerous fat-soluble substances are made water-soluble in the liver through chemical conversion processes finally released with the urine. An important property of the liver is its regenerative ability which is essential for natural healing processes as well as for organ and tissue engineering.

Structure and Organisation of 'Systems of Life – Systems Biology' in Regard to Contents
The activities of the 'Systems of Life – Systems Biology' network of competence are divided into three thematic **modules**. In **module 1** – development of methods – tools and methods will be developed for generating quantitative data for systems analysis. In this process it is particularly important that actual experiments and mathematical and computer- assisted modelling go hand in hand right from the beginning. Another key aspect is the establishment of standard conditions. And finally, free access has to be ensured to the data generated within this network of centres of excellence.

In **module 2** – modelling – the algorithms and computer models needed for systems biology will be developed. These include the establishment of a central bioinformatics platform which, among other things, will have to ensure standardised data filing and accessibility.

Finally, in **module 3** – cell biology – the biological systems studied by the network of centres of excellence will be established. Here, the focus will be on developing suitable cell lines which in the medium term will permit concentration on the hepatocyte system.

One of the great challenges and, at the same time, one of the essential success factors of the research programme is the integration of the expertise of different disciplines into the various collaborative projects. Traditionally, research has been structured mostly along the lines of different disciplines, while interdisciplinarity requires special efforts. A second major success factor is close interaction between standardised data generation and computational modelling. A special **organisational structure** was chosen to take up these challenges and to ensure a targeted and efficient implementation of the programme in the medium and long term. The activities of the network of competence are controlled by an international steering committee supported by a project co-ordinator whose task it is above all to ensure the exchange of information among the working groups receiving grants. Applications for project funding under the programme are assessed by an international referee panel.

Complementing other BMBF research activities
The 'Systems of Life – Systems Biology' programme complements already existing BMBF research activities that were set up under the 'Biotechnology – Using and Shaping its Opportunities' framework programme. The German Human Genome Project which was launched in 1996 as a joint initiative of the BMBF, the DFG and industry is now in its second phase (1999 to 2002) and has shifted its focus to functional genome analysis. It centres on the functional analysis of medically relevant human key genes and of the genes of model organisms that are necessary to understand the human genes. A National Genome Research Network (NGFN) is being built by further pooling, cross-linking and expanding the capabilities of the best-performing scientific, clinical and industrial partners. In this network five core institutes will closely co-operate with five disease-oriented genome networks. In addition, bioinformatics and proteome research are integrated into the network as platform technologies. The methodology developed for functional genome analysis – e.g. high-throughput technologies for functional analysis – and the knowledge thus gained of the function of genetic and physiological networks form an important basis for systems biology. Additional networks of centres of excellence in genome research are created under the GenoMik research initiative which focuses on

microorganisms, and the GABI plant genome project ‘Genome Analysis of the Plant Biological System’.

The development of techniques and methods for genome and proteome research is complemented and optimised by BMBF’s priority funding of proteomics research. The methodological findings gained here can be immediately used in protein network analyses which are required for systems biology. This also applies to the Nanobiotechnology research priority which covers the interface between physics, biology, chemistry and the engineering sciences and which, among other things, aims to develop completely new measuring technologies for the biosciences. Finally, the Bioinformatics Training and Technology Initiative contributes to broadening the computerscience knowledge base needed for systems biology in Germany.
The ‘Systems of Life – Systems Biology’ initiative will also provide an important methodological foundation for exploring and developing new approaches to the prevention and treatment of human diseases. This research exercise also complements activities carried out under the BMBF's Health Research Programme. The Systems of Life – Systems Biology research and funding programme is also meant to help shift systems biology more into the focus of modern bioscientific research in Europe. The interdisciplinary approach of systems biology seems to be ideally suited for the creation of European research networks.

Implementation of the Programme ‘Systems of Life – Systems Biology’
During the first funding phase, the BMBF-financed network "Systems of life systems biology" will probably consist of three major collaborative projects working on their own specific questions on an interdisciplinary basis. They will be supported by the platforms for cell biology and computer science/modelling, which partly pursue separate projects and perform development tasks in major areas which go beyond the topics of individual collaborations. Important priorities will initially be the development of suitable hepatocyte cell culture systems and the establishment of an overarching data, modelling and simulation platform for the overall collaboration.
Within the framework of their specific research goal, the collaborative projects will consist of research groups comprising experts in cell biology, methods development and modelling. Both the research groups of the individual collaborative projects and the researchers of the platforms will be expected to practise a high degree of cooperation and interdisciplinary communication to be able to meet demands of the complex topics dealt with in this interdisciplinary funding priority on the modelling of the human hepatic cell. The partners will be supported by an intranet ensuring fast and interactive exchanges of information. Furthermore, standardisation and working protocols, SOPs and other documents of general interest and relevance will be available to all the partners in the competence network. In this way it will be possible to set up a functional national network of competence on systems biology, which is to be incorporated into an international context in the medium term.
To provide support with the implementation of the research projects and the recommendations of the international steering body, a project coordinator was appointed at the DECHEMA e.V. in Frankfurt to act as interface between the scientists of the competence network and the steering body. The members of the independent steering body were appointed by the Federal Minister of Education and Research and charged, *inter alia*, with the task of actively steering the scientific development of the German competence network on systems biology.
Progress of the Programme ‘Systems of Life – Systems Biology’
The first projects of the programme ‘Systems of Life – Systems Biology’ will start at the 01.01.2004. Networks will be devoted to the topics defined by the modules listed above. Platforms in bioinformatics and cell biology will be formed. Further information of project details will be available at http://www.systembiologie.de.

Computational Systems Biology
Workshop
Brussels 10-11 September 2003

Norbert Hübner
Max-Delbrück-Center for Molecular Medicine (MDC)
Berlin-Buch, Germany

===
Systems Biology

- Genomics of transcriptional regulation
- Genetical Genomics

Physiological Genomics
===
Functional Genomics in Engineered ES cells
FunGenES
Main objectives:

- fundamental knowledge of the key biological process of
- cell differentiation and lineage commitment with
- identification of potential novel targets for therapeutic
- intervention
- novel cellular and molecular tools to characterize gene
- function in specialized cell populations and potential use

for small molecule testing
===
Development of different organs from the mammalian germ layers
====
Systems Biology

- Genomics of transcriptional regulation
- Genetical Genomics
- Physiological Genomics

===
Self renewal and early commitment

- Maintenance of ES cell pluripotency in-*vitro* requires the
- continuous supression of differentiation.
- This blockade is sustained through signalling pathways
- which are likely to up-regulate the expression of genes that
- promote self-renewal and proliferation and to down-

- regulate the expression of genes that otherwise commit
ES cells to differentiation.

===

Identification and characterization of novel genes and signal transduction pathways involved in maintenance of -and exit from- pluripotency in mouse ES cells

- Characterization of the *LIF* transcriptome
- Characterization of the *STAT3* transcriptome
- Role of PI3K dependent signalling in the control of self-renewal
- Identification of self-renewal genes using novel inducible
- expression screen
- Identification of genes involved in the exit of pluripotency
  - e.g. Retinoic Acid Response Elements

===

Identification of novel genes involved in the early commit-ment of ES cell into ectoderm, mesoderm, and endoderm lineages

- Identification of early commitment genes by means of selective
- cloning of differentially expressed cDNAs
- Identification of novel genes encoding modulators of *Oct4*
- expression
- Identification of early commitment genes by means of an *esi*RNA
- based strategy
- Chromation modifications that regulate differentiation and lineage
- commitment of mouse ES cells
  - ChIP on Chip

===

Systems Biology: Genomics of transcriptional regulation
Goals -

- ✓ identify the transcriptional regulatory networks active in unicellular
- ✓ and multicellular organisms
- ✓ first approached at the static level of a circuit diagram
- ✓ when more information available as a dynamic process that varies
- ✓ with input and time
- ✓ focus efforts on identifying targets of DNA binding TF and the sites to which factors bind productively as well as understanding the effect of binding on RNA expression

===

Systems Biology: Genomics of transcriptional regulation

Datasets -

- ✓ TF binding to target sites (e.g. ChIP/chip analysis)

✓ Global RNA expression profiles at cell-type specific resolution

- Accurate models of binding specificities of TFs

✓ Effects on expression of loss of function and overexpression of TFs

✓ Interaction partners for all TFs

✓ Chromatin structure and methylation status at genome-wide scale

✓ Comparative sequence analysis –
>draft of many species, along with better annotation of sequences

===

Systems Biology: Genomics of transcriptional regulation

Future tools - experimental

- Intergenic region arrays for all model organisms
- TF characterization tools including:
    - libraries of expression constructs of tagged TFs
    - antibodies for every TF
- Better methods for defining Position Weight Matrices
- High throughput experimental validation methods for interactions of TFs with targets

Approaches using genetic and physiological variation

Systems Biology: Genomics of transcriptional regulation

Future tools – computational

- Improvement in analysing genome wide data:
  e.g. expression data (normalization stategies, influence of polymorphic RNA, significance issues etc...)
- Better algorithms for:
    - identifying binding sites
    - deducing regulatory networks
    - making predictions from a deduced regulatory networks

Integration of existing algorithms

===

Systems Biology

- Genomics of transcriptional regulation
- Genetical Genomics

Physiological Genomics

Computational Systems Biology – personal view

Stefan Hohmann, Göteborg University

This text has been written after the Computational Systems Biology workshop (Sep 10 and 11, 2003) and hence reflects also impressions and conclusions from the discussions during those two days.

Definition

Like Functional Genomics in its beginnings, also (Computational) Systems Biology as a field is interpreted differently by different people. It is often regarded as a further step of Bioinformatics and Functional Genomics. For instance, it is thought that Systems Biology encompasses technologies that bring system and structure into the huge amount of functional genomics data. While Systems Biology should make use of those data, their organisation is, however, the task of Bioinformatics. In addition, Systems Biology is often defined as a discipline that strives for a complete understanding of whole cells and organisms. While it is difficult to comprehend what a "complete" understanding may encompass, a global understanding of how different subcellular systems interact and function in context is certainly of interest to Systems Biology.

Possibly a useful description of Systems Biology derives from its actual goals. Those are to understand the structure and function of biological systems that are composed of a certain number of interacting biomolecules, cells or even organisms. In other words, Systems Biology strives at the understanding of the logic and the elucidation of the functional rules of modules or systems, rather than the individual parts of those.

Based on this definition, Systems Biology is inherently multidisciplinary and requires the input from biomedical experimental research as well as from mathematics, computer sciences, physics and engineering. More specifically, Systems Biology makes use of mathematical models (computer replicas of the system) that are based as much as possible on actual data in order to understand systems properties such as feedback loops, robustness, bistability and more. An important property of the mathematical models is that they can be used to predict properties of uncharacterised systems components, predict the results of experiments and help phrasing hypotheses, thereby assisting experimental planning, reducing the number of experiments and opening up for a number of possible applications such as in drug development, diagnosis, breeding and genetic engineering.

See also: www.systemsbiology.org

Why has Systems Biology become a topic now?

The use of mathematical models in biological research is not at all new. However, for many years those models commonly had little if any footing on actual data and therefore lacked realistic use. In fact, they were regarded by experimentalists as a playground for mathematicians and being completely useless. What has changed in the last few years?

- The availability of global data, such as gene expression and proteomics data that provided information on most or all components of a module/system.
- The emerging (though still not general) interest of biologists to collaborate with mathematicians and researchers from other disciplines, which partly has been driven by relevant programmes from different funding agencies, including the EC.
- The realisation that Systems Biology approaches can help advancing biomedical research and allow addressing research questions that cannot be targeted by experimentation alone.
- Persons that drive and shape the field and gave it the name Systems Biology, such as Leroy Hood, Hiroaki Kitano, Roger Brent, Hans Westerhoff and others.

What Systems Biology can do and possibly deliver

Systems Biology works with mathematical models that precisely replicate the structure and function of the relevant module/system under study. This means, to the best possible extent should the model be based on experimental data and it should be able to simulate as precisely as possible the actual operation of the system. The cooperation of experimentalists and mathematicians should result in iterative improvement of the model and hence the understanding of the system. In other words, the model should in mathematical form contain all available experimental knowledge and connect individual data to a functional unit. As such, the model has predictive abilities, i.e. it opens for the possibility for in silico experimentation to test alterations to the system or new perturbations that have not previously been tested in experiments. Based on these requirements, Systems Biology approaches can be applied for the following:

- Planning of experiments, thereby optimising the design of biological experiments, reduce their number and hence make research more cost-effective and targeted.
- Elucidating properties of biological modules/systems that cannot be understood on the basis of experimental data alone.
- Identifying components of the system hitherto unknown from experimental work.
- Help to understand the basis for diseases and diseases processes.
- Assist to identify the "weak spots" in systems, i.e. the possible targets for pharmacological intervention: drug target discovery and drug design.
- Help to determine, eventually on an individual basis, the best timing and mode of drug application to cure diseases.
- Help designing approaches of genetic engineering or breeding to optimise crops of biotechnologically relevant microorganisms.

QUASI – A Systems Biology project funded by the EC

QUASI is a STREP presently under contract negotiation and was submitted in response to the first call for proposals under FP6. Although not submitted as a Systems Biology project, it has important elements that are completely in line with the goals of Systems Biology, as defined above.

The project aims at a better understanding of the dynamic operation of so-called MAP kinase signalling pathways. Such pathways play important roles in the control of cell division, cell morphogenesis, stress responses and development in all eukaryotes. They are involved in diseases such as cancer and inflammatory diseases and are potential drug targets.

QUASI consists of an experimental component (four groups) and a bioinformatics component (two groups).

The experimental part aims at collecting quantitative data on the dynamic operation of the signalling pathways in the model system baker's yeast. Data are to be collected in time courses. Where possible, quantitative, time-dependent data on the relative proportion of systems components in different cellular compartments are to be collected. Preferably, all individual steps in signalling should be monitored. Collecting quantitative, time-dependent and spatial data is an important aspect where Systems Biology employs dynamic models.

The bioinformatics part of the project aims at developing and iteratively improve dynamic mathematical models of the MAP kinase system. This is to be done in close collaboration between mathematicians and experimental biologists. These models are expected to assist experimentation and should be possible to apply to MAP kinase systems in all organisms. In addition, a work package of the project aims at developing illustration tools to present and visualise the mathematical models.

Taken together, QUASI is a Systems Biology project because:

- It collects and makes use of quantitative and time-dependent and spatial data.
- It uses mathematical models.
- It makes use of existing as well as functional genomics data and, in addition, collects new data to feed mathematical models and to iteratively improve those
- It uses model organisms and studies well-defined modules
- It consists of close collaboration between experimentalists and mathematicians
- It aims at understanding systems properties

Needs and actions

It can be expected that:

- Systems Biology approaches become an integral part of biological/molecular biological research over the coming years.
- Systems Biology will become important to fully exploit the potential of genomics and functional genomics.
- Systems Biology will become highly important in drug target identification, drug design, assessment of side effects, drug approval and application to patients.
- Systems Biology will become an important tool for breeding and genetic engineering of crops, farm animals and microorganisms.

For these reasons it is important to develop the area aggressively.

Europe could potentially be in a leading position since it has a tradition in Systems Biology approaches. But at this point Europe lags behind because Japan and the US have already invested heavily (publicly and privately) and have built relevant infrastructures (such as the ERATO Kitano project in Japan, the Institute for Systems Biology in the US, and others). A present problem in Europe is fragmentation of the area, which is apparent already when it comes to a definition of the field. For this reason it may be advisable to define the type of possible future Systems Biology projects explicitly (see example below). Fragmentation is also apparent when it comes to funding: while some countries like Germany, Finland and the Netherlands have already put programmes in place, other countries lag behind.

What kind of action is needed in FP6?

- Success stories: well-defined Systems Biology projects that testify the power of the approach. We hope QUASI could become one.
- A visible larger project, either and NoE or an IP, see example call text below.
- Actions to coordinate at the EU level national initiatives and encourage such initiatives where they do not yet exist. As an SSA will be funded and a call for a CA has been published, the necessary action seems to be in place.
- Training, especially of researchers trained both in experimental and mathematical research. This can be achieved through Marie Curie Actions (Networks, EST…).

The following is an example call text for a NoE/IP in FP6

Elucidation of how system properties arise in defined cellular modules.

The objective is to enable researchers to study properties and dynamic operation of complex biological modules/systems.
Projects should make use of existing data as well as experimental and computational approaches to understand the properties and operation of cellular modules/systems in model organism. Among deliverables should be tools for predictive in silico experimentation to use the full potential of genomics and post-genomics.

What could be done for FP7 and beyond?

- Call for several projects (IP) to achieve silicon replicas of larger modules (metabolism, signalling, trafficking, organelles, cell cycle, gene expression, replication, cytoskeleton) in model organisms.
- Extensive coordination and management of the approaches.
- A vision to strive for whole cell projects in FP8.
- Extensive support for training.
- Requirement for Systems Biology components in all relevant biomedical projects.
- Thereby establishing a mid and long term vision and sustainable funding perspectives for the area.

Systems Biology in Sweden

Systems Biology within Sweden is itself still fragmented. Principally, there are groups that try to achieve a systems understanding based on publicly available global data (e.g. Tegner in Linköping, Aurell in Kista, Petersen in Lund) as well as teams that combine experimental research and mathematical modelling (Ehrenberg in Uppsala using *E. coli* and the Göteborg groups using yeast). Recently, these groups have started a Swedish Systems Biology initiative (www.systembiologi.org) that organises workshops and plans courses in the area open to students from all organisations involved in the network. In addition, the interest group discusses with funding agencies to put in place relevant programmes. The 3rd International Conference on Systems Biology was held in Stockholm in December 2002 (www.ki.se/icsb2002/).

In Göteborg, Systems Biology is visible in research and in training within the following platforms:

- The Research School for Genomics and Bioinformatics (funding: Ministry of Education and Research). The School offers undergraduate programmes that include Systems Biology and funds PhD students.
- Additional funding for Systems Biology studies comes from the Foundation for Strategic Research and the Wallenberg Functional Genomics initiative (SWEGENE), although this funding is not earmarked for Systems Biology.
- Within the research programme of the School, there are three pair-student projects in place, one on metabolism and two on signal transduction. In these projects, one student and one supervisor each from experimental biology and mathematics/computer sciences collaborate on a defined biological question. The approach works very well.
- The Göteborg Yeast Centre, which encompasses eleven groups and about 50 researchers, has explicit research goals and visions in Systems Biology. It organised the XXI International Conference on Yeast Genetics and Molecular Biology in 2003, which had a workshop and a plenary session on Systems Biology.

Two workshops will be in November 2003, in which Swedish research and training in Systems Biology will be discussed with the goal to add structure and common goals to ongoing activities.

Computational Systems Biology (CSB)
Its future in Europe

WORKSHOP ORGANISED BY
DG RESEARCH of the European Commission,
in Brussels 10-11 September 2003

SESSION 2. FUTURE REQUIREMENTS

Cellular networks: new tools and approaches

Lilia Alberghina
Dept. of Biotechnology and Biosciences
University of Milano-Bicocca, Milano, Italy

Integrated biological processes given by a large number of components, such as metabolism, signal transduction and cell cycle, need to be described in quantitative functional terms if one wants to understand their control circuits and to be able to predict their behaviour at changing conditions by simulation (1).
This approach, called systems biology (2-3), is surely needed to give structure to the wealth of information coming from genetic and biochemical investigations and more recently from genomic and post-genomic findings. Drug discovery and in general health care are expected to become much more effective and less costly if able to rely on modelling of cellular functions at the molecular level (4-7).

A systems biology approach to cell cycle control.

Although the understanding of cell cycle control is very important for basic science and for drug discovery, the enormous number of data available in many experimental systems is far from giving a clear picture.
In budding yeast it has been established that at least 15% of the 6000 gene products are involved in cell cycle (8). One of the more physiologically relevant control function is given by the requirements for a critical cell size to enter S phase; this fact has been known for more than 25 years but so far no satisfactory molecular mechanism has been proposed.
We developed a model for which the cell sizer control involves cyclin dependent kinase Cdk1, cyclins Cln3, Cln1,2 and Clb5,6, and inhibitors (Cki) Far1 and Sic1. We collected data available from the literature and data banks and assayed during systematic perturbations the levels of relevant molecules. In this way we have been able to derive a map for the cell sizer network (9-10). To simulate the Start network we built up a system of nonlinear ordinary differential equations, obtained from the biochemical model by means of standard chemical kinetics. The model is constitued by the equations for 27 proteins (or complexes) plus an equation for the total mass; there are involved about 100 parameters. Simulations were carried out on a Linux platform using a variable step Runge-Kutta algorithm. When available, experimental data were used as a starting point for the simulations. The simulation correctly predicted known cell cycle deletion mutants phenotypes and in its current implementation the model appears able to rationalize major features of Start execution (10).
In order to further test the role of Sic1 in the control mechanism we derived by homology modelling the 3D structure of the complex Cdk1/Clb5/Sic1. Sic1 has nine consensus sequences for Cdk1 phosphorylation, that are involved in setting the degradation of the Cki, and one consensus sequence for CK2, kinase known to be involved both in the G1 to S and in the G2 to M transitions. We determined the affinity constants of unphosphorylated and

CK2-phosphorylated peptides to a Cyclin/Cdk complex by BIAcore analysis, and collected evidences of the stronger inhibitory activity of CK2 phosphorylated Sic1 in the Cyclin/Cdk activity on histone. The simulation of active complex (Clb5/Cdk1) formation suggests a potential role of CK2 phosphorylation of Sic1 for the control of the onset of DNA replication (11).
These findings support the notion that, for an effective systems biology approach, modelling, computation and tailored “wet” experiments need to actively interact.

Which new tools and approaches are required?

Most of the current knowledge of biological processes is qualitative. The identification of the sequence of DNA and proteins, the function of biological molecules, the identification of protein-protein interactions, the reconstruction of metabolic or signal transduction pathways, the patterns of gene expression obtained by DNA microarrays, are all examples of qualitative description.
As it has recently noted by Nobel laureate Paul Nurse, “… to explain logical and informational processes on a cellular level, therefore, we need to devise **new ways to obtain and analyse data**, particularly those generated by genomic and post-genomic studies” (ref 11, bold type added for emphasis).
A model of network topology is a hypothesis about the real interactions. It must be confirmed or refined using experimental data, and will be used to suggest new experiments and predict their results as well as to detect inconsistencies. Since the network model itself is not sufficient to make quantitative predictions about the behaviour of the regulatory system under study, the network topology must be translated into a mathematical description that allows quantitative testing of the hypotheses about network topology and dynamics. Modelling and simulation thus need quantitative data, like concentration and absolute number of relevant molecules present in each cell compartment (for instance nucleus, cytoplasm, mitochondria), the affinity or dissociations constants for interactions, the rate constant of biological reactions and the extent of material flow. These data should be obtained at least as function of different physiological states and of time. An evolution of experimental technologies should be fostered in order to be able to measure as many molecular properties as possible in real time and, at least for some set of data (i.e. concentration of ions in cell compartments) it would be necessary to develop non invasive methods.

Standards to generate quantitative data

There is a compelling need to elaborate standards for data generation, which covers both the methodology of analysis and the data quality assessment. Besides, standardisation will allow the comparison of findings in different laboratories for the same or for connected subsystems, and therefore would greatly speed up the pace of systems biology investigation. Standardisation should extend also to the methods employed to induce systematic perturbations of systems. Genetic methods (for instance deletion, expression and mutation of a gene supposed to participate to a given network) are generally used. Genetic methods have been shown to be more informative when applied to cells in different growth conditions (see for instance ref. 13). The standardisation of the analysis of the pattern of response to the perturbation is also required.

The first set of data that need to be standardized is number and/or concentration, phosphorylation and other molecular modifications, localisation and interaction of other molecules, activity levels, and kinetic constants of the relevant molecules of a network..

Standards for modelling and simulation

Computers models of biological systems may have different levels of definition and rely on different approaches (14). Quite often we find chemical kinetics models that represent cellular processes as systems of chemical equations, mathematically expressed as differential equations. It is emerging the need to develop a friendly and unambiguous symbolic representation of the cellular networks to be modelled and simulated.

Development of standards for the model description so that the components included, the computer terminology, the units of measurement, the reaction specifications and the corresponding computer code used in the modules will be compatible when combined together. The Systems Biology Markup Language (SBML) and Systems Biology Workbench are possible starting points.

Development of software tools, in high-level, multi-platform, non-proprietary languages, that are capable of scale-up to the whole-cell simulation, and that can combine the modules into the whole cell model. This is a challenging aspect of the project, because of the size of the simulation and its very wide range of concentration and time scales.

Development of visualisation tools that can display not only the network structure of the model, but can also overlay the results of analyses and simulations. Web access to these visualisations. Already existing platforms such as the proprietary *Gene Network Sciences* visualization tool can be considered to accelerates process by creating dynamic computer models of living cells. GNS aims to combine the power of mathematics, computation, experimental molecular biology, and bioinformatics to bring genomics data to its most valuable end, making drug discovery predictive.

In conclusion

Systems biology is an area of scientific discovery raising great interest and with yet undefined borders. The computational aspect of systems biology should integrate the standard bioinformatics approaches in a frame of modelling and simulation.

Computational Systems Biology will surely require new standardized ways to obtain and analyze experimental data to yield a clear representation of the logical network controlling cell behaviour. New symbolic representations of cellular networks and new software for the visualization of complex cellular dynamics are required. Besides, it would be of great interest to launch a project for the systems biology description of a whole cell. The budding yeast is very suitable for this purpose being the best known eukaryotic cell, and on the other hand it shows homology for the molecular basis of important functions with human cells.

The achievement of these aims will make systems biology valuable for postgenomics research and for drug discovery.

References

1. Hartwell, L.H., J.J. Hopfield, S. Leibler, and Murray, A.W. From molecular to modular cell biology. *Nature*. **402**, C47-52 (1999).
2. Kitano, H. Systems biology: a brief overview. *Science*. **295**, 1662 (2002).
3. Kitano, H. Looking beyond the details: a rise in system-oriented approaches in genetics and molecular biology. *Curr Genet*. **41**,1 (2002).
4. Henry, C.M. Systems biology. *Chem.& Eng. News* **81**, 45 (2003).
5. Kitano, H. Computational systems biology. *Nature*. **420**, 206 (2002).
6. Collins, F.S. et al. A vision for the future of genomics research. *Nature*. **422**, 835 (2003).
7. European Federation of Pharmaceuticals Sciences – New Medicine Faster – Report, July 2003

8. Spellman P.T. et al. Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol Biol Cell.* **9**, 3273 (1998).
9. Alberghina L. et al, 2002, ICSB3 – Proceedings of the 3rd International Conference on Systems Biology
10. Alberghina L., Rossi R.L., Querin L., Wanke V., Paleari S and Vanoni M.. Cln3, Far1 and Sic1 as determinants of a cell sizer network controlling entrance into S phase in budding yeast. *Submitted* 2003.
11. Barberis M., De Gioia L., Ruzzene M., Sarno S., Marin O., Coccetti P., Fantucci P., Pinna L.A., Vanoni M. and Alberghina L. CK2 phosphorylation regulates inhibitory activity of the yeast cyclin-dependent kinase inhibitor Sic1. *Submitted* 2003.
12. Nurse, P. Systems biology: understanding cells. *Nature*. **424**, 883 (2003).
13. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929-34 (2001).
14. German Federal Ministry of Education and Research (BMBF). Systems of life. Systems biology. (2002).

Simulating what cannot be simulated
Olaf Wolkenhauer
www.sbi.uni-rostock.de www.sbi.uni-rostock.de

Experience & Previous Work

Protein Identification & Sequence Analysis: Bayesian mass spec data analysis Information geometry of protein sequences Microarray Data Analysis & Modelling: Clustering & classification Gene network modelling Dynamic Pathway Modelling: Signal transduction pathways Switching, regulation, and control … we are interested in mathematical modelling & data analysis, with applications to molecular- & cell biology. … we are interested in mathematical modelling & data analysis, with applications to molecular- & cell biology.

Experimental Design and Quality Control

Two types of error introduced in µArrays: Systematic All measurements equally affected Source detectable Random Every measurement affected in a different way Can be accounted for with replicated measurements Normalisation: Removes systematic errors Replication: Removes random variations Adds statistical significance to results … uncertainty can be reduced, through clever experimental designs!

Systems Biology

The Central Dogma of Systems Biology:

It is systems dynamics that gives rise to biological function, regulationand control. The Systems Biological Approach: Signal- & systems-orientedapproaches applied to intra- & inter-cellular dynamics. DATA DATA INFORMATION INFORMATION MODEL MODEL SIMULATION SIMULATION EXPERIMENT EXPERIMENT DESIGN DESIGN … moving forward from molecular characterisation and cataloguing. The Central Question of Systems Biology: How do cells, genes and their products interact and reactto environmental stimuli? … requires systematic perturbation studies.

Systems Biology takes Genomics towards its natural conclusion: an understanding of cellular dynamics. Systems Biology takes Genomics towards its natural conclusion: an understanding of cellular dynamics. The Systems Approach: Causation is the principle of explanation of change in the realm of matter. Causation is a relationship, not between things, but between changesof statesof things. Motivation (and illusion?): Understanding physiological effects through modelling from the molecular and cellular level upwards. DNA mRNA Proteins Enzymes/Metabolites Reaction Networks Cell Physiology The aim of science is not things in themselves but the relations between things; outside these relations there is no reality knowable. The aim of science is not things in themselves but the relations between things; outside these relations there is no reality knowable. Henri Poincaré

History of Systems Biology

1929: W.B.Cannon: Feedback regulation in organisms: homeostasis. 1945: L.Bertalanffy: Theory of the organism as an open system. 1948: N.Wiener: Cybernetics – control and communication in the animal and the machines. 1958: R.Ashby: Adaptive, self- organising behaviour: organisms as machines. 1958: R.Rosen: Metabolism-Repair Systems: why organisms are not machines. 1968: M.Mesarovic: 1970: F.Jacob & J.Monod: Cellular cybernetics – regulatory proteins. 1975: L.Segel: Enzyme kinetics. 1975: J.Kacser: Interactions & Dynamics. 1978: J.G.Miller: Living Systems Theory – from cells to supranational systems. “The real advance [..] will come about only when biologists start asking questions which are based on systems- theoretic concepts [..] then we will [..] have [..] a field of systems biology.” “The real advance [..] will come about only when biologists start asking questions which are based on systems- theoretic concepts [..] then we will [..] have [..] a field of systems biology.” Closely related developments: Complexity studies. Pattern formation. Physiological modelling. Recent important books: 1996: H.Heinrich & S.Schuster: The regulation of cellular systems. 1996: A.Goldbeter: Biochemical oscillations and cellular rhythms. 1997: D.Fell: Understanding the control of metabolism.

Bioinformatics vs. Systems Biology

Mininggenomic data we can only describe the direction and intensity of co-variation. The systems approach investigates functional relationshipsand thereby allows hypotheses about causal entailment . We want to do the latter but practise the former; why? The cell is a dynamic system!

Bioinformatics vs. Systems Biology

We need to move away from just cataloguing components, data, and information We need better technologies to quantify cellular dynamics (accurately at high resolution). The cell is a complexdynamic system! To understand intra- & inter-cellular dynamics… We need new ways of thinking. … and for this … … and subsequently …

What is a complex system?

Many variables Nonlinear relationships between variables Difficult to observe/measure/quantify … leading to uncertainty in the analysis.

===

Mining- vs. Systems Approach

… it is systems dynamics that give rise to biological function! … it is systems dynamics that give rise to biological function! A system is defined by inputs and outputs: For living systems, the present depends on the past: … unstable, unbounded. Adding negative feedback:

====

We can't always do what we want…

linear parametric model (can explain underlying principles if applicable…) nonlinear non-parametric model (lacks explanatory power)

===

In the theory of dynamic systems we generally have to make a decision whether to regard the process as a deterministic non-linear system but with a negligible stochastic component or to assume that the nonlinearity to be only a small perturbation of an essentially linear stochastic process. Genuine nonlinear stochastic processes have not yet been shown to be applicable for practical time-series analysis.

===

Dynamic Pathway Modelling

nonlinear deterministic stochastic linear linear non-linear Breakdown of the superposition principle: the whole is more than the sum of its parts.

===

The Purpose of a Systems Approach

a dynamic system! Modelling product enzyme complex substrate Concentration [uM] time [sec] Simulation:

===

The Ras–Raf–MEK–ERK Module

1. Ras activated by growth factor receptors.
2. Ras binds to Raf-1 kinase.
3. Activation and recruitment of Raf-1.
4. Raf-1 phosphorylates and activates MEK kinase.
5. MEK-PP phosphorylates and activates ERK.
6. ERK translocates to the nucleus.
7. ERK-PP regulates gene expression by phosphorylation of ranscription factors.

The kinase cascade controls cell differentiation and proliferation of various cell types. growth factors (extra-cellular signals) Activation Kinetics of Pathway Components (transcription factor) (G-protein) (MAP kinase) nucleus membrane cytosol

===

Dynamic Pathway Modelling

ODE models are not factual representations but mathematical abstractions. ODE models are not factual representations but mathematical abstractions. Assumptions: Law of mass action Molecules move independently Small concentrations creation proportional to A/B collisions decay proportional to number of molecules present Dynamic Pathway Modelling?www.sbi.uni-rostock.de www.sbi.uni-rostock.de Mathematical Models as Mediators For interdisciplinary collaborations in Systems Biology,

keep expectations for mathematical models realistic: they will not be accurate they can help testing and generating hypotheses they can help designing experiments … in engineering we learn most from models that fail! … the modelling process itself is valuable and should be at the heart of interdisciplinary collaborations. Mathematical modelling becomes more important as the current battle of equipment turns into a battle of brains…

===

Dynamic Pathway Modelling

Challenges Too many variables, too few data Nonlinearity Identification & quantification of feedback loops Formal analysis for very large systems: … phase-plane & bifurcation analysis. … dimensionality reduction. Not a well-stirred reactor… Larger diffusion times -> partial diff. equations Compartments -> more variables Randomness – a matter of chance? Few molecules -> stochastic model Many molecules -> deterministic model Model validation Hypothesis testing vs. generating hypotheses

===

Challenges for Modelling & Simulation

Dynamics at different time scales: Regulationis the maintenance of constant conditions w.r.t. perturbations. Controlis the ability to make changes as necessary. Regulation vs Control:

===

Cellular Weather Forecasting

As the complexityof a system increases, our ability to make precise and yet significant statements about its behaviour diminishes until a threshold is reached beyond which precisionand significance(or relevance) become almost exclusive characteristics. As the complexityof a system increases, our ability to make precise and yet significant statements about its behaviour diminishes until a threshold is reached beyond which precisionand significance(or relevance) become almost exclusive characteristics. (Zadeh’s Uncertainty Principle)

1. How to identify the model structure: key variables and heir relationship.
2. Methodologies for parameter estimation.
3. Experimental and formal methods for model validation.
4. Identification of feedback, and circularity from xperimental data.
5. Modular representations and simulation of large scale ynamic systems.
6. Investigations into the stability and robustness of ellular systems.
7. Visualization and fusion of information, integration of odels and simulators.
8. Scaling models across time scales and description evels (from genes, transcripts, and proteins to cells and organisms).

Principal Challenges in Systems Biology:

===

Bottlenecks: Cellular Weather Forecasting

Bottlenecks:Complexity Observability Uncertainty (Dimensionality reduction) (Quantification, and instrumentation) (Experimental design, replication) … to approach these challenges, Systems Biology needs: … and doesn’t require: • People, with ideas, pen, paper and a PC. • Good collaborators. ?Expensive hardware ?Laboratories … systems biologists can’t get enough of it: (data that is)

===

Systems Biology’s Central Dogma

Instead of cataloguing genes as causal agents for some protein, function, or phenotype, we should relate observations to sequences of events. …it is system dynamicsthat gives rise to biological function!

Molecular Simulations of Membrane Proteins: Towards a Virtual Membrane

Mark S. P. Sansom
http://sansom.biop.ox.ac.uk
mark@biop.ox.ac.uk

Complex Systems: MD Simulations of Biomolecules

water, ions minimum system 50,000 to 100,000 atoms proteins & phospholipids e. g. potassium channel environment cell interior biology

‹ Describe the forces on all atoms: F = dU(x)/ dx
‹ Integrate: F = ma (a few million times…)
‹ Result: positions of all atoms for ~10 ns
‹ Experimental (static) structure ? in vivo dynamics
? The challenge – to relate dynamics to function
===

MD Simulations: Why?

‹X-ray structure
static,average structure at 100 K
‹MD simulation multi-nanosecond dynamics at 300 K
‹The challenge:to relate dynamics to biological function
–interpolation,extrapolation &in silico experiments
===

Roles for Biomolecular Simulation

‹Physicochemical extrapolation: from static structure to dynamic physiological properties
‹ Homology modelling & beyond: from static bacterial structures to dynamic mammalian homologues
‹ Large scale simulations and emergent complexity: towards a virtual membrane
‹ Comparative dynamics: conservation of dynamic properties across protein families
‹ Integrative biomolecular simulations: from QM to meso- scale modelling
===

Bacterial Outer MembraneProteins

‹Numerous structures known –X-ray &NMR
‹Potential antibiotic and vaccine targets
‹Current simulations:
OmpA (pore);OmpX & OpcA (recognition proteins);PagP,OmpT &OMPLA (enzymes);FhuA &FepA(transporters)
‹Towards a virtual outer membrane (vOM)
===

OmpA:Dynamics vs. Environment

‹Multiple, comparative simulations – on 10 to 50 ns timescale
‹ Need to relate dynamics in experiments to in vivo function
‹Small changes in flexibility – can open the central pore
Bond & Sansom (2003) J Mol Biol 329: 1035
===

More OMPs: Building a Library

‹ OpcA from Neisseria meningitidis: a pathogen
‹ Recognition protein: binding to target cells
‹ Possible ion channel: test pore formation by MD
‹ OMPLA – an outer membrane lipase
‹ Simulations: calcium- free monomer, Ca 2+ - bound dimer, Ca 2+ -bound dimer- inhibitor
‹ Differences in dynamic stability due to local ordering of water around Ca 2+
===

High Throughput Simulations of OMPs

‹Biosimulation database entries for an entire class of membrane proteins: cf. BioSimGRID
‹ Robust protocols for high throughput simulation.

‹ Comparative analyses of e. g. membrane protein dynamics vs. environment: cf. OmpA
‹ A first step towards a ‘virtual outer membrane’: cf. vOM
‹ Current simulations: OmpA – pore OmpX, OpcA – recognition OmpT, OMPLA, PagP – enzymes FhuA, FepA - transporters
==========

?Towards a Virtual Outer Membrane (vOM)

OmpT OmpX OmpA OmpF PhoE FhuA P i TolC LamB FhuD MalE P i BP OMPLA OpcA d - d - d - d - d + P i TonB
? First step towards computational systems biology – a suitable system
‹ Bacterial OMs – 5 or 6 proteins = 90% of protein content
‹ Structures or good homology models of proteins are available
‹ Complex lipid – outer leaflet is lipopolysaccharide (LPS)
‹ Minimum system size ca. 2. 5x10 6 atoms; simulation times ca. 50 ns cf. current FhuA – 80, 000 atoms & 10 ns – need HPCx
===

Coarse-Graining

‹ MD (atomistic) simulation – analyse dynamics
‹ Parameterise coarse- grain simulations
‹ Large (time, length, component) scale simulations
‹ Emergent properties
===

Managing MD Data:BioSimGRID

‹ www. biosimgrid. org
‹ Distributed database environment
‹ Grid/ Web services using GT3/ OGSA infrastructure
‹ Software tools for interrogation and data- mining
‹ Generic analysis tools
‹ Annotation of simulation data QM drug binding protein motions drug diffusion

Multiscale Biomolecular Simulations

‹ Membrane bound enzymes – major drug targets (cf. ibruprofen, anti- depressants, endocannabinoids)
‹ Complex multi- scale problem: QM/ MM; ligand binding; membrane/ protein fluctuations; diffusive motion of substrates/ drugs in multiple phases
‹ Need for GRID- based integrated simulations

?Computational Challenges

? Need to integrate HPC, cluster & database resources
? A ‘classical’ E- science problem…IntBioSim HPC Linux cluster BioSimGRID database

Conclusions:HPC & GRID Resources

‹ Simulations and systems biology: integrating molecular and cellular descriptions of biological function
‹ Both capability and capacity HPC resources are essential
‹ Very large simulations require improved scalability of codes
‹ Multi- scale simulations require heterogeneous GRID- enabled resources

Phil Biggin Carmen Domene Alessandro Grottesi Andrew Hung Daniele Bemporad Shozeb Haider Kaihsu Tai Bing Wu George Patargias Oliver Beckstein Yalini Pathy Pete Bond Jonathan Cuthbertson Sundeep Deol Jeff Campbell Loredana Vaccaro Jennifer Johnston Katherine Cox Robert d’Rozario John Holyoake Andy Pang BBSRC DTI EPSRC GSK IBM & HPCx MRC OeSC (EPSRC & DTI) OSC (JIF) The Wellcome Trust Peter Tieleman (U Calgary) Richard Law (UCSD) Joanne Bright (ANU) Marc Baaden (Paris) BioSimGRID Leo Caves (York) Simon Cox (Southampton) Jon Essex (Southampton) Paul Jeffreys (Oxford) Charles Laughton (Nottingham) David Moss (Birkbeck) Oliver Smart (Birmingham)

Present Challenges in Systems Biology: A personal summary of the EC workshop on CSB

Uwe Sauer
Institute of Biotechnology, ETH Zürich, sauer@biotech.biol.ethz.ch
http://www.biotech.biol.ethz.ch/sauer/sauer.html

The essence of systems biology is quantitative understanding of the interaction of components of biological systems, which are not represented in the components themselves [1]. The necessary in silico representations of complex subsystems/modules of cells/organisms must enable quantitative predictions of system behavior. At an advanced stage with defined and verified models, such predictions will inevitably lead to new scientific discoveries and will have many important applications not only in medical/pharmacological research but eventually also in patient treatment.

In sharp contrast to bioinformatics, systems biology relies on iterative cycles of model construction/refinement, predictions, experimental design, and experimentation. Existing data, in particular genomic and postgenomic data, are instrumental for initial model construction. Different from verbal interpretation, however, quantitative and comparable data sets are necessary for model validation/falsification. Hence, consistent data sets obtained for a single strain under well-defined environmental conditions are a prerequisite for model construction. This matter was discussed by various attendants at the CSB workshop under the topic standardisation and deserves some clarification.

Experimental standardisation

Standardisation of quantitative heterogeneous data sets pertains primarily to their direct comparability. Since biological systems are often sensitive to the exact environmental conditions, their quantitative system responses are not directly comparable for quantitative modelling if, for example, different conditions (often subtle and unnoticed) were used to generate data sets in different labs. For the initial phase of model testing and hypothesis generation, it is thus of utmost importance to rely on consistent and standardized data sets. Beyond the use of one strain and defined physiological conditions, this includes standardisation of system perturbations (e. g. genetic or environmental modifications), well-defined and verified analytical methods, and consistent statistical data treatment. Appropriate control mechanisms to verify data comparability and reliability should be part of systems biology projects in research networks. This standardisation is crucial for the initial phase of systems biology to allow identification of faithful models and parameter sets. It should be understood though that once a suitable model is defined, it can certainly deal with non-standardized data. In fact, the identification of data in large heterogeneous sets that are quantitatively or structurally inconsistent with other data or the present model is a hallmark of systems biology, and will become of great value.

Computational standardisation

A separate issue of standardisation relates to the computations themselves. Here standardisation is necessary at the level of unambiguous (and simple) representation of the networks and components modeled, model description (including reaction specification, measurement units, etc.), data storage and retrival, and the computer codes. The latter assumes a particular importance because models developed by different networks/groups represent typically modules of cellular operation that should be compatible with each other. Hence, development and use of consistent use of multi-platform, non-proprietary programming languages such as SBML is a priority. The ultimate goal are modular combinations of models and routine applications of ‘standard’ models in non-specialist (experimental) labs. Hence, the development of open-source, professional software (and maintenance) should be encouraged.

Beyond present ‘omics:

Functional data At present, compositional transcriptome, proteome, and metabolome studies dominate large-scale functional analyses. The missing link in contemporary functional analyses, however, is the capacity to observe the output of the true units of function [2]. Such functional data

may be, for example, exact cellular localization of proteins, their interaction in supramolecular structures, or reliable protein-protein interaction data. While the definition of function is somewhat fluid and a matter of controversy, there was a broad consensus that simply collecting 'omics data is insufficient.

In linking genes and proteins to higher-level biological functions, the molecular fluxes through fully assembled biochemical networks determine the systemic phenotype in metabolic research [3]. The capacity to quantitatively observe this whole network operation by methods of metabolic flux analysis based on 13C-labeling experiments, thus provides a global perspective of the integrated, system-wide regulation at the transcriptional, translational, and metabolic level. Such quantitative functional information is highly important for systems biology.

Data integration was consistently identified as a top priority. At the first level, consistent (same strain/conditions) and quality-controlled large-scale 'omics data sets must be made available via databases. This includes also consistent statistical data treatment for the heterologous data sets that goes beyond the current ad hoc practice. In collaboration with experimentalists, this is the realm of bioinformatics. At the next level, however, these heterologous data sets must be integrated into predictive models of some detail that allow to identify inconsistencies, systematic experimental errors, and important connections between certain subsets of heterogeneous data; all of which become then priority targets for further experimentation.

Computation vs. experimentation

In sharp contrast to functional genomics, systems biology does not follow a large-scale data collection and analysis scheme. Computation and experimentation are simultaneously occurring and integrated components of systems biology research. Models may be build from publicly available data to indicate – with lower confidence of course – the most important next experiments for a given experimental subsystem. Initially quantitative, inspired guess experimentation may be the major effort of most projects, but eventually model-derived hypotheses will become increasingly important for experimental design. In the intermediate and long run, systems biology will significantly reduce novel experimentation because computations identify pivotal missing components for quantitative understanding of the fully assembled system or module.

Fragmented research in Europe:

What is required from the EC? Fragementation includes research and funding in different countries but also know-how and approaches in different scientific fields. How can a fruitful environment be created?

The CSB workshop has helped to promote a common understanding among leaders of different scientific disciples. It is obvious that substantial EC funding, possibly in combination with national funding agencies, is necessary to foster systems biology in Europe. At short term, STREPs, NoEs, and teaching networks are necessary. STREPs are an immediate possibility for well-defined projects of small, interdisciplinary research teams. Such teams are in some sense the core for larger, high-quality network-based projects. NoEs and Teaching networks are important to overcome fragmentation at all levels and to supply appropriately educated scientists, respectively. At an intermediate scale, larger research networks are important (IPs). Appropriate calls for IPs in systems biology make only sense, however, if a reasonable volume is made available. As a consequence of the broad applicability and the involved tools, one or two IPs are insufficient to advance systems biology in Europe.

How to select for high-leverage projects?

Systems biology is a nascent field, hence cannot be expected to yield applied benefits such as novel drug targets etc. immediately. Instead preference should be given to strategic projects that provide conceptual computational and experimental advances, using relevant model systems. Selection criteria should include:

? Relevance of the model system

? experimental accessibility of the model system for hypothesis testing

? available European expertise

? possibility to establish European leadership.

Potential model systems

? S. cerevisiae – model eukaryote, excellent exp. accessibility, presently the spearhead of European systems biology with a clear lead versus the US and Japan, existing European networks, strong industrial interest in biotech and in pharma as a model.

? B. subtilis – the gram-positive model microbe, excellent exp. accessibility, strong biotech industry interest, history of EC funding with excellent, established networks in place.

? E. coli – probably the best known microbe, excellent exp. accessibility, projects should be tied to the International E. coli Alliance (IECA) [4] to position Europe within this world-wide program and to ensure that Europe has access to the conceptual advances made in this top-notch project.

? Filamentous fungi – strong European networks, history of EC funding, strong biotech interest. Disadvantage: limited experimental accessibility, additional levels of complexity, lack of a clear model case for higher cells.

? Neurons – was discussed as an example of a higher cell type with a competitive situation for Europe, highly interesting but low exp. accessibility.

? Hepatocytes – given the strong funding in Germany, EC projects may aim at connecting this nucleus to other top European groups in the field.

Immediate needs

? Coherent, high quality data sets as a basis for model construction

? quantitative dynamic data sets for time-dependent changes

? absolute concentrations of proteins (and their modification) and mRNAs

? new experimental tools for functional analysis (in situ proteomics, reliable protein-protein interactions, metabolic fluxes etc)

? heterogeneous data integration

? software and model standards

? new modelling concepts

? advance from analytical to predictive modelling

1. Kitano H: Systems biology: a brief overview. Science 2002, 295:1662-1664.

2. Bailey JE: Lessons from metabolic engineering for functional genomics and drug discovery. Nature Biotechnol. 1999, 17:616-618.

3. Hellerstein MK: In vivo measurement of fluxes through metabolic pathways: The missing link in functional genomics and pharmaceutical research. Annu. Rev. Nutr. 2003, 23:379-402.

4. Holden C: Cell biology: alliance launched to model E. coli. Science 2002, 297:1459-1460.

Systems Biology at TNO

Age K.Smilde[1,2] and Jan van der Greef[1,3]

[1] *TNO Nutrition and Food Research, Zeist, The Netherlands*
[2] *University of Amsterdam, Amsterdam, The Netherlands*
[3] *Leiden Amsterdam Center for Drugs Research, Leiden, The Netherlands*

Introduction

Systems Biology has already a long tradition in TNO. Already from the 80-ies, TNO invests in measuring in bodyfluids (profiling) and developing bioinformatics tools to handle the resulting (complex) data. In 1999, the Systems Biology concept was developed at TNO and this resulted a.o. in the creation of Beyond Genomics Inc, a leading company in the USA applying Systems Biology concepts in the pharmaceutical industries.

View on Systems Biology

The view of TNO on Systems Biology is to integrate transcriptome, proteome and metabolome data to form (nonlinear, dynamic) system models of organisms. Based on these models, biomarkers can be selected (nutrition, health, diseases) or new microbial production routes explored. This concept is applied to microbial systems, but also to model systems as mice, rat, guinea pigs, and to human beings.

Current Systems Biology at TNO

Several years ago, the Board of Directors of TNO decided to embrace the Systems Biology concept in their theme Quality of Life. This resulted in a concerted effort and a separate (large) budget to spend for a period of 3-4 years. This Systems Biology programme involves six TNO institutes:

1. Nutrition & Food Research
2. Prevention & Health
3. Environment, Energy & Process Innovation
4. Technical & Physical Services
5. Prins Maurits Laboratory
6. Strategy, Technology & Policy.

Structure and organization of the Systems Biology programmame

The driving force of the research is the biology. The technology consists of state-of-the-art analytical instrumentation, such as nuclear magnetic resonance (NMR), mass spectrometry (MS), liquid- and gas chromatography (LC, GC) and combinations thereof. The informatics part consists of bioinformatics, data analysis & pattern recognition, data base building and software engineering. The programme consists of individual projects, run by projectleaders. These projects are supervised by programme leaders to ensure coherence. The programme leaders are:

- B. van Ommen (Biology)
- E. Verheij (Technology)
- A.K. Smilde (Informatics)
- J. van der Greef (Overall)

These programme leaders meet regularly to ensure the integration of the three areas.

Fields of applications

The fields of application of the developed Systems Biology concepts and tools are:

- Medical Systems Biology
- Nutritional Systems Biology
- Microbial Systems Biology
- Plant Systems Biology

The tools developed are generic and can be applied to all these areas.

Relationships with other initiatives
The TNO Systems Biology initiative relates to other initiatives in this area. Specifically, TNO participates in:

- The Center for Medical Systems Biology at Leiden; one of the Genomics Centers in the Netherlands
- Kluyver Centre for Genomics of Industrial Fermentation at Delft; another one of the Genomics Centers in the Netherlands
- European Nutrigenomics Organization (NUGO); a recently approved EU Network on nutrigenomics

Moreover, TNO has contracts with companies to apply the Systems Biology concepts.

Needs in Computational Systems Biology

A lot of computational tools and methods have to be developed in Systems Biology. To name a few, tools and methods have to be developed to:

- Build system models
- Integrate transcriptome, proteome and metabolome data
- Find patterns
- Explore (nonlinear) dynamics
- Select biomarkers
- Link data analysis results to biology

CSB Workshop, Brussels - September 10 -11, 2003

Karl Kuchler Vienna Biocenter
http://www.at.embnet.org/molg/kuchler/
kaku@mol.univie.ac.at

➲Computational Systems Biology
THE emerging research area in biomedical research

===

CSB is hot, BUT there are problems.....

- ➲ Insufficient cross-talk of biology, mathematic & engineering
- ➲ Many CSB activities - fragmented area
- ➲ Lack of integration and coordination of current CSB activities
- ➲ Only few national funding programs exist in Europe
- ➲ Little participation of Europe in other CSB activities (USA, Asia)
- ➲ Student training not widely available at Universities

CSB potential very high, but exploitation difficult (SMEs...)

===

The immediate needs of CSB in Europe.....

- ➲ Competitive CSB demands FP6, 7... (ERC?) and national support
- ➲ Connection of existing national funding programs - ERA-NET
- ➲ Database and communication platform for European scientists
- ➲ Identify researchers in member states as CSB contact points
- ➲ Human ressources - Ph.Ds, post-doctoral levels & junior groups
- ➲ Identify CSB topics and appropriate model systems

Standard protocols for data aquisition and common tools

===

How can we address problems and needs of CSB..

- ➲ Human ressources to enable competitive CSB development
- ➲ Limit models systems (bacteria, yeast, mammalian systems)
- ➲ Establish national programs in member states - academic
- ➲ Involve SMEs and Big Pharma (research & funding)
- ➲ Networking and coordination of ongoing CSB activities
- ➲ CSB funding should use bottom-up and top-down approaches

One answer could be - EUSYSBIO - and (hopefully) ESBIGH

===

Training of human ressources - a key for European CSB.....

===

A bright future for European CSB.....

- ➲ Concentrate and network CSB research activities
- ➲ CSB needs to think “European“
- ➲ Critical mass in funding and scientists - think “Big“

Establish Europe as competitive key player in global CSB

===

Summary thoughts....

- Workshop helped convey importance of CSB to EC - FP7
- EC funding (think BIG) might induce national activities where non-existing (i.e. via ERA-Net).
- Multidisciplinarity is definition of CSB - proposal review by EC!
- Training & Mobility (CSB Ph.D. in MC, support hi-le conferences)
- National CSB centers as training sites - Projects in CSB
- Realistic - Choose model systems where experimental tools are availailable and thus data generation is possible!!!!!!!!!!!
- NoE for coordination of European activities

===

Alfonso Valencia - Possibilities and limitations of the genomic information: biodegradation networks as a case study.

ref: ©2003 EUROPEAN MOLECULAR BIOLOGY ORGANIZATION
EMBO reports VOL 4 | NO 10 | 2003

Microbial biodegradation of environmental pollutants is a field of growing importance because of its potential use in bioremediation and biocatalysis. We have studied the characteristics of the global biodegradation network that is brought about by all the known chemical reactions that are implicated in this process, regardless of their microbial hosts. This combination produces an efficient and integrated suprametabolism, with properties similar to those that define metabolic nets in single organisms. The characteristics of this network support an evolutionary scenario in which all reactions evolved outwards from the central metabolism. The properties of the global biodegradation network have implications for predicting the fate of current and future environmental pollutants.

Comparative and Functional Genomics
Comp Funct Genom 2002; 3: 000–000.
Published online in Wiley InterScience (www.interscience.wiley.com). **DOI:** 10.1002/cfg.224
Feature

Conference Report: ESF program on 'Integrated Approaches for Functional Genomics' workshop on 'Modelling of Molecular Networks'

Hotel Alixares, Granada, Spain, 12–14 June 2002

Paulino Gomez-Puertas [1] and Alfonso Valencia [2] *
[1] *Center for Astrobiology, Madrid, Spain*
[2] *CNB–CSIC, Madrid, Spain*
**Correspondence to:*
Alfonso Valencia, Protein Design Group, CNB–CSIC, Centro Nacional de Biotecnologia, Cantoblanco, 28049 Madrid, Spain.
E-mail: valencia@cnb.uam.es

The rapid pace of genome sequencing and new high-throughput methods are offering an unprecedented opportunity for investigating how individual genes and gene products cooperate to build up complex cellular structures and perform elaborate processes that enable cells and organisms to live and reproduce. The diagrams of cell regulatory net-works that are being produced look increasingly complex, and it becomes impossible to use mere intuition to make predictions about their behaviour. Thus, the need for new integrative approaches is becoming paramount. These approaches range from systematic integration of large amounts of data, to efficient querying tools, to rigorous statistical analyses, and dynamic modelling. Such character-ization of whole biological processes is becoming known as 'systems biology', and it will have a predictable impact on our knowledge of biological systems.

Shoshana Wodak -
Bioinformatics requirements for Systems Biology
Universite Libre de Bruxelles, Belgium
shosh@ucmb.ulb.ac.be

Beyond Genomes:
Finds out how genes and Proteins interact to give rise to cellular processes
===
Challenges of the post-genomic era
Curation & management of the fast growing bodyof primary data
Deciphering sequence information in terms of biological function
-Gene assignment
-Infer function by homology
-Comparative genomics
-Experimental analyses:
* Transcriptome
* Proteome
Gaining understanding of cellular processes
Analysing & simulating cellular processes
===
Only a small fraction is chartered: known metabolic, regulatory & signal transduction pathways..
Cellular systems: Huge networks of thousands of molecules interconnected via thousands of interactions
===
Information on function
Public databases, such as: GenBank, EMBL, SWISS-PROT,PDB, contain information on individual genes and proteins;

Information on function appears as annotation in text form, and cannot be readily analysed by a computer.
===
Information on function
Several DB's on metabolic pathways
Several databases on enzyme function
2-3 DB's on transcriptional regulation
No DB's on gene regulation networks
No DB combining data on metabolism and gene regulation
1-2 fledging DB's on signal transduction
Only scant information on transport
SRS (T.Etzold, Lion Biosciences/EMBL-EBI), :
Links information from different sources on metabolic pathways, enzyme function & regulation
===
What are major needs ?
Represent the functional and physical interactions between all biochemical entities mapped onto their cellular and tissue locations:
Metabolism + regulation + signal transactions
All organisms
Represent the parameters(rates, affinities, concentrations) associated with these interactions, and offer means of tracking any piece of data to its source (e.g.literature ref.); description of incomplete knowledge.
* Tools for querying, displaying & analysing automatically the structure of network of interactions at various levels of granularity.
* Tools for data curation and annotation***

* Interfaces to simulation packages

===

Database embodying a rich enough representation to enable different applications All these applications are important for improving our ability to understand biological function, hence to assign function & identify important genes.

===

aMAZE: Representing biological function as networks of molecules & interactions work started at the EBI-EMBL, now at ULB, funded by the Government of the Brussels Region

A rich Object Oriented model:

- integrating protein-protein interactions into their functional context

-capable of handling different types of processes: metabolism, regulation transport, signal transduction, etc.

-representation of quantitative parameters

===

Example of signal transduction pathway

gepasi.dbs.aber.ac.uk/metab/signal/signal.html

===

Conclusions

The concept of one gene- one function does not hold anymore The (biological) function of a gene is the role (or roles) that it plays in the cell. It is defined by the various interactions that it and its products make with other molecules in the cell. Hence, it requires a systems level description

Ultimately the function of a gene will not be looked up in a catalogue, but obtained through a query to databases that contain the necessary information not only on all the players but also on all their interactions with their space and time dependence.

->Need for new types of databases and tools to analyse the information in them. These tools go all the way from simple queries to static and dynamic analysis of the network of interactions

Bioinformatics for Systems Biology:

Rich representations of interactions and processes

- Adequately represent data collected from experiments
- Handle complex network structure and interactions and their space & time dependence
- Should contain quantitative information
- Suitable for some analyses
- “views” (data structures) for other analyses can be generated
- Object Oriented approach is one valid alternative

Computational models

- Capable investigating dynamic properties
- Can take their input data from the rich representations

(as shown here by the example on circadian rhythm)

Eero Vuorio
11.09.2003

Forum of Genomics Research Managers & CSB

An EC-catalysed initiative
Aims:
CO-ORDINATION OF NATIONAL GENOMICS INITIATIVES

SYNERGIES BETWEEN NATIONAL GENOMICS PROGRAMMES

Secretariat:
A STRATEGIC ACCOMPANYING MEASURE [SSA]: COGENE (CO-ORDINATION OF GENOMICS RESEARCH ACROSS EUROPE)

COGENE

Five work packages
WEBSITE (AND A NEWSLETTER) FOR INFORMATION EXCHANGE BETWEEN FUNDING ORGANISATIONS & SCIENTISTS

A SURVEY OF NATIONAL GENOMICS PROGRAMMES (INCLUDING KEY FUNDING ORGANISATIONS, RESEARCH UNITS AND INFRASTRUCTURES)

INFORMING THE PUBLIC

WORKSHOP ON POPULATION GENOMICS

WORKSHOP ON PHARMACOGENOMICS

COGENE

Achievements
A FORUM FOR INFORMATION EXCHANGE BETWEEN FUNDING ORGANISATIONS, E.G. WHEN ESTABLISHING NEW RESEARCH PROGRAMMES

WORKSHOP ON POPULATION GENOMICS BROUGHT TOGETHER RESEARCH MANAGERS AND LEADING SCIENTISTS TO DISCUSS FUTURE NEEDS OF RESEARCH ON GENETIC EPIDEMIOLOGY (E.G. POPULATION COHORTS, STANDARDISATION OF BIOBANKS AND LEGAL RESTRICTIONS FOR USE OF BIOBANKED MATERIAL)

COGENE

Achievements
Workshop on population genomics also resulted in a survey of existing (and future) population cohorts and biobanks for genetic studies on common diseases with multigenic background, and triggered a transnational investigator-driven initiative for their coordination, which now has expanded to a global initiative.
Collaboration between national funding agencies has also resulted in transnational initiatives between Nordic countries; EUSYSBIO

Other activities of the Forum

An information platform also for other initiatives
GENOMES OTHER THAN HUMAN (MOUSE, MICROBES, PLANTS)

PRIME - A EUROPEAN PROGRAMME IN MOUSE FUNCTIONAL GENOMICS

BIOINFORMATICS

SUPPORT FOR ESSENTIAL RESOURCES (INFRASTRUCTURES) FOR FUNDAMENTAL GENOMICS

Other activities of the Forum

Identification of national research activities (programmes) of European interest suitable for EC co-ordination action
SYSTEMS BIOLOGY

PLANT GENOMICS

A co-ordinated call for proposals on a specific topic?

Challenges in transnational research funding

Marked heterogeneity of research funding organisations in member countries
PRACTICAL ISSUES PROVIDE UNBELIEVABLE OBSTACLES (SCHEDULING JOINT CALLS, AGREEING ON PEER REVIEW, TRANSFER OF FUNDING ACROSS BORDERS, NATIONAL TRADITIONS, ETC)

EUROPEAN SCIENCE FOUNDATION (EUROCORES SCHEME)

European Molecular Biology Laboratory (EMBL)
A transnational research institution of 17 European countries

Eero Vuorio-

The FORUM for Genome Programmes Managers and the need for CSB

RESEARCH PROGRAMME ON SYSTEMS BIOLOGY AND BIOINFORMATICS

1 FOREWORD

The planning of the Research Programme on Systems Biology and Bioinformatics has been performed in close co-operation between all four Research Councils of the Academy of Finland, together with the National Technology Agency Tekes.

Since the mid-1990s, the Research Programmes of the Academy of Finland have provided targeted support to research, networking, researcher training and infrastructures in the field of biotechnology and molecular biology in Finland. The Research Programmes on Genomics, Cell Biology, Molecular Epidemiology and Evolution, Structural Biology and Biological Functions (”Life 2000”) as well as the Research Programme on Mathematical Methods and Modelling in the Sciences comprise a continuum providing long-standing support to the various aspects of molecular biology and biotechnology. The Research Programme on Systems Biology and Bioinformatics is a logical extension of this support.

Research and researcher training in bioinformatics have also been supported through these Research Programmes, and through targeted funding from the Ministry of Education and from the National Technology Agency Tekes. This support has helped to build the national infrastructure for gene expression profiling by microarrays and proteomics, and for structural biology. Two recent surveys[10,11] on bioinformatics and biotechnology in Finland indicated a need for a research programme to bring together scattered research capacity and to increase co-operation between different scientific disciplines, along with a need for intellectual and property right and legal services as well as faster application and commercialisation of research results.

Due to very rapid developments and competition in these fields, it is important to target support to systems biology, bioinformatics and computational biology. Together with the strengths of the Finnish research base in information technology, biometry, biomathematics, population genetics and epidemiology, the interdisciplinary nature of the Research Programme will provide considerable added value.

2 BACKGROUND

Publication of the preliminary nucleotide sequence of the human genome at the turn of the millennium was one of the milestones in modern biology. Yet this information package of 3 000 000 000 nucleotides marks only the beginning for modern ”postgenomic” research on molecular genetics and life sciences in general. A characteristic feature of such research is the generation of increasing amounts of raw data requiring advanced informatics services and tools to process this information into biological knowledge. In addition to the draft nucleotide sequence of the human and mouse genomes, those of several other eukaryotic model organisms, e.g. *Caenorhabditis elegans*, *Drosophila melanogaster*, *Arabidopsis thaliana*, *Saccharomyces cerevisiae*, *Saccharomyces pombe* and those of more than 100 bacteria and archaea are currently known. This offers unique possibilities for comparative biological analyses. Analysis, organisation and mathematical modelling of large amounts of data present a major challenge to modern bioinformatics. Examples of research fields utilising high throughput methodology and thereby producing huge data sets include structural biology, molecular modelling and simulations and genetics of multigenic traits and diseases.

Multidisciplinarity and integration are characteristic features of postgenomic research. Genes, gene products, their regulatory networks and interactions with environment must be analysed as components of higher order structures, metabolic pathways or entire cells and organisms. This type of

[10] Bioinformatiikka Suomessa, Teknologiakatsaus 129/2002, Tekes

[11] Biotechnology in Finland. Impact of public research funding and strategies for the future - Evaluation Report. Publications of the Academy of Finland 11/02

an integrative and holistic approach has been termed **systems biology**. Research defined as systems biology is characteristically multidisciplinary and dependent on **bioinformatics**, the computer-assisted analysis of biological data. Close collaboration of biologists, biochemists, physiologists, chemists and physicists with computational biologists and mathematicians is needed for the characterisation and modelling of the complex interactions of genes, proteins and metabolic processes.

A range of ethical, political and economical constraints limit the generation, processing and application of biological information about populations and individuals. These challenges are not only faced by researchers, but are also important to decision makers and laymen. Storage and use of human genetic information and manipulation of genomes pose ethical questions and challenges, necessitating research on the ethical, social and cultural dimensions of bioinformatics and systems biology.

3 THE SCOPE OF THE PROGRAMME

In order to understand the complex biological systems, knowledge of the molecular characteristics of individual components or phenomena is not enough. A holistic view and integrative, multidisciplinary approach is needed to study the complex interactions between components and networks.

Examples of research fields that the Research Programme on Systems Biology and Bioinformatics will cover:

Structural biology

Research support will be available for structural analysis of biomolecules and their complexes with preference for multidisciplinary proposals involving several research fields, ranging from chemistry, biochemistry and physics to computer science and medicine. The field has rapidly evolved from analysis of individual gene-protein-systems to high-throughput structural genomics. This development is likely to result in elucidation of thousands of structures within the next few years, which will require high-performance bioinformatics. The resolved structures will also form the foundations for analysis of molecular interactions, drug development and design, proteomics and functional genomics, as well as for commercial applications.

Functional genomics and proteomics

Support will be available for postgenomic research on humans as well as on all model organisms whose genomics are sufficiently well known. The covered research areas include genomics, transcriptomics, proteomics and metabolomics, and the techniques used in these fields. Again, the goal is to support multidisciplinary approaches for the elucidation of the interplay of cellular and subcellular structures as well as metabolic pathways, and their function. This information is needed e.g. to engineer microbes, plant and animal cells into cell factories. Microarrays and proteomics are key techniques to produce raw data from these complex processes and their interactions, the interpretation of which requires improved tools and know-how of bioinformatics.

Molecular genetics

Support will be available for research on the genetics of multifactorial human diseases, on gene-environment interactions and on model organisms. All these topics are dependent on the development of new tools for bioinformatics due e.g. to the huge amount of raw data produced by the single nucleotide polymorphism (SNP) analyses and microarrays. Combined with the homogenous health care system, population history, several nationwide registries and an overall supportive atmosphere, the Finnish population offers advantages for molecular genetic analyses of polygenic diseases and for studies on the interaction of genes, environmental factors and life style in the pathogenesis of several common diseases. Results from such studies are expected to pave the way for pharmacogenomics and theranostics, which are based on novel diagnostics and development of targeted (personalised) medical treatments.

Bioinformatics, biomathematics, and computational biology

The Research Programme supports basic research and method development in bioinformatics, biomathematics and computational biology. The methods of bioinformatics are based on computer science, statistics, and mathematical modelling. Because the treatment of the central themes of systems biology leads to computationally intensive problems, the development of efficient algorithms is crucial for successful research. Collecting, storing, handling, sharing and analysis of large amounts of data require mathematical and statistical modelling as well as the development of visualization methods. In addition to the locally collected and produced data, biological information is stored all

over the world in data banks, on Internet sites, etc. The extraction of important information from extensive databases is possible only using data mining, pattern recognition, classification methods and other related techniques. The organization and coordination of these databases is an important and challenging task. Bioinformatics is not solely data analysis. The ultimate goal is to understand the system or process that produces the data. This is possible only by mathematical modelling.

Ethical, social and cultural aspects

The Research Programme on Systems Biology and Bioinformatics will also provide support to research on the ethical and socio-economic aspects of systems biology and bioinformatics. The information obtained through biological experimentation and subsequent biomathematical analyses is likely to result in fundamental changes in our conceptual thinking of disease, health, health policy, diagnostics and prognostics. Application of the information produced involves important political and economic interests and challenges. Data and privacy protection becomes an increasingly important issue as health information accumulates in different registries. Research on these topics is needed to maintain public confidence in biological research.

4 OBJECTIVES OF THE PROGRAMME

The main objective of the Research Programme is to promote an integrative and holistic approach in research on biological processes at the systems level. Multidisciplinarity, interdisciplinarity and transdisciplinarity are essential characteristics of the Programme, with bioinformatics envisioned to play a central integrating role in the projects.

More specifically, the Research Programme aims to

create new knowledge through high-quality, multidisciplinary collaborative research in the field of systems biology and bioinformatics,

promote efficient and synergistic use of the existing resources and infrastructures,

develop research environments, methodologies and co-operation of researchers,

promote efficient researcher training and mobility of researchers, taking the multidisciplinary nature of the Programme into account,

promote application of technologies both in basic research across disciplines, and in research and development aiming to protected intellectual property and commercialisation of research results,

increase information on and knowledge of the ethical and socio-cultural dimensions of systems biology and its applications among researchers and in society.

Dissemination of the research results is considered very important in order to increase the impact of the Programme.

5 IMPLEMENTATION OF THE PROGRAMME

The Systems Biology and Bioinformatics Research Programme is scheduled to run for four years from 2004 through to 2007. The Programme is coordinated by the Academy of Finland and implemented jointly by the Academy of Finland and the National Technology Agency Tekes. Other funding agencies may join in at a later stage. Each funding agency uses its own procedures and criteria in making their funding decisions.

Academy of Finland

The Board of the Academy of Finland has allocated EUR 9 million to the Programme. Projects to be funded by the Academy of Finland may receive a four-year funding starting from 1 January 2004 and ending by 31 December 2007. The projects will be evaluated using the following criteria: relevance of the project to the Programme, scientific quality and innovativeness of the research plan, feasibility of the research plan, competence and expertise of the applicant and the research team, national and international networks, and research and training environment.

National Technology Agency Tekes

Tekes has reserved the minimum of EUR 1.35 million for academic projects in the Programme. Tekes may consider additional research and development funding for companies, if suitable applications are filed. Besides the general criteria of the Programme, Tekes will emphasize collaboration with industrial partners.

The Research Programme is managed and supervised by a Steering Group, assisted by the Programme Manager. The Steering Group is chaired by Professor Marja Makarow (Research Council for Health).

The members are Professor Annele Hatakka, (vice chair; Research Council for Biosciences and Environment), Professor Mats Gyllenberg (Research Council for Natural Sciences and Engineering), Professor Juha Sihvola (Research Council for Culture and Society), Professor Peter Slotte (Research Council for Biosciences and Environment), Professor Eero Vuorio (Research Council for Health), Senior Technology Adviser, Docent Erja Heikkinen (National Technology Agency Tekes) and Senior Technology Adviser Pentti Nummi (National Technology Agency Tekes). In addition, Scientific Secretary Timo Sareneva, PhD, Scientific Secretary Janica Ylikarjula, D.Sc. (Tech) and Scientific Secretary Helena Vänskä, M.A., and Programme Manager Sirpa Nuotio, PhD, from the Administrative Office of the Academy of Finland are involved in the Steering Group. The Steering Group will invite additional experts as advisors.

The research projects that receive funding through the Research Programme, are expected to work in close co-operation with each other and to contribute to networking and training of researchers both nationally and internationally.

Particular attention will be given to the administration and coordination of the Programme. Joint seminars, workshops, training courses and electronic communication will be used to reach this goal. The added value and scientific impact of the Programme are dependent on efficient cooperation and communication between scientists working in different fields. The Research Programme will be evaluated during 2008.

6 INTERNATIONAL CO-OPERATION

The Research Programme aims to strengthen the European Research Area through networking and cooperation of scientists in Finland and in other countries.

The applicants are encouraged to initiate international collaborations. To expedite the transfer of know-how into Finland, support can also be provided for Finnish researchers and research teams moving temporarily abroad to carry out collaborative research in leading international institutions as well as for foreign scientists to work in Finland. Opportunities for international co-operation will be promoted also later on during the Programme.

The Academy of Finland is actively searching for international collaboration with other funding agencies interested in targeting funding to systems biology and/or bioinformatics. An example is the networking of the Research Programme with the "Systems of Life - Systems Biology" programme of the Federal Ministry for Education and Research in Germany (http://www.systembiologie.de/index.php).

7 APPLICATION PROCEDURES AND DEADLINES

Applications for participation in the Research Programme on Systems Biology and Bioinformatics will be processed in two stages. In the first stage, all applications are submitted to the Academy of Finland. In the second stage, the application shall be submitted either to the Academy of Finland or/and to the National Technology AgencyTekes, as will be proposed by the Steering Group.

Steven M. Foord -
The plans and needs of Large Pharmaceutical Industry for CSB

Not Data mining

Genetics
physical map to helping patients
Human sequence variation is due to ~10 million variations- most are SNPs
Segments of the human genome show very limited haplotype diversity- they are inherited as ‘blocks’.
Some markers report a lot of sequence and others report on very little.
A ‘haplotype map’ will make genotyping a more efficient means of ‘decoding individual traits’.

Conclusions
Function
Use ‘omics’ to best effect
Simplify options for pathway discovery

The vital role of Computational Systems Biology in future Neurobiological research

Nicolas Le Novère, Computational Neurobiology, EMBL-EBI, Hinxton, UK;
Tel: 01-45-68-88-44; Fax: 01-45-68-88-36; e-mail: lenov@ebi.ac.uk

Study of cellular signaling occupies a particular place in contemporary research in Biology. The large number of molecular partners involved is not approached but maybe in the case of gene regulation. However, the intricacy of signaling pathways is much larger than the one we can infer from our current knowledge of transcription control. A significant percentage of the 25,000 genes idendified in the human genome code for proteins directly or indirectly involved in the pathways of intercellular communication. In particular, the number of different pharmacological receptors generated from those proteins is enormous (let's remenber that a receptor can be formed by several different proteic subunits, opening the door to a very large combinatorial diversity). Chemical synapses between neurons constitute a specific type of intercellular signaling. Specialised in the conversion of electrical signals into chemical ones, and vice-versa, those structures are essential to the function of our nervous system, in other words, to the life of any complex animals. What we believed we knew about those intercellular structures has been completely disintegrated in the last decade.

Recent evolution of the notion of chemical synapse

The classical concept, still widely used, of the so-called chemical synapse is based on the idea that a presynaptic terminal button releases only one neurotransmitter. Post-synaptic densities are considered as stable clusters, made up of a single type of receptor, homogeneously distributed. As a consequence, theoretical models used to simulate the function of those "ideal" synapses are often rather simple.

On the contrary, recent reseaches in molecular and cellular neurobiology have shown that chemical synapses are highly complex dynamical structures. A single terminal button can sometimes release several neurotransmitters in the synaptic cleft. Accordingly the post-synaptic part can exhibit densities made up of different receptors, sometimes with opposite cellular effects (such as nicotinic acetylcholine receptors, cationic channels, and glycine receptors, anionic channels. Postsynaptic domains are sometimes highly convoluted, and within these domains the receptors can be heterogeneously distributed. Both the structure of the post-synaptic membrane and the differential location of receptors is likely to strongly influence their cellular function. Moreover, the receptors are embedded in complex multimolecular assemblies. The synapse is now understood in terms of a complex supra-macromolecular assembly, encompassing proteins of the pre- and postsynaptic sides, but also part of the cytoskeleton and the basal lamina. This organisation could be crucial for the synaptic function. Moreover, the whole structure is dynamic and evolves, for instance under the control of the neuronal activity. Receptors diffuse laterally in the plasma membrane, and get trapped in or escape from the post-synaptic domain. The figure shows the movements of an individual glutamate metabotropic receptor. The red spot represent the post-synaptic domain). The diffusion accross the plasma membrane depends on the clustering state of the receptors, the composition of the juxta-membranous compartment, and even on the lipidic composition of the plasma membrane. The stabilization of neurotransmitter receptors at synaptic sites appears as a reversible process which is likely to be governed both by the topology of the synaptic apparatus as well as by the affinity of the receptors for their stabilizing elements. In addition to the horizontal diffusion, there is a continuous vertical flux of receptors, newly inserted in the plasma membrane or removed by internalisation. The direction of those vertical movements depend on the molecular identity of the receptors.

The complexity described above was totally unexpected a decade ago, and calls for a deep change in the representation we give of the neuronal transmission at the sub-cellular level. In particular it is of utmost importance to take into account the fine structure of synapses and their dynamics if we want to understand their actual mechanisms of action, both at the pre- and post-synaptic sides.

Enters *Systems Biology*

Two types of modeling have been used to tentatively understand and reproduce biological phenomenons, in particular at the molecular and cellular levels.

Analytic reductionism concentrates on a very small portion of living, like a molecule or a simple signaling pathway, isolated from its initial environment, and try to describe it as accurately as possible (opposing simplifying reductionism, which sought to simplify the problem up to the point only a representative backbone remains). The main problem of this approach is the real possibility that the investigation of an isolated system would provide parameters unrelated to its actual function when it is embedded in a larger interacting network. In addition, the ignorance of regulation mechanisms, such as feedback loops, does not permit to understand the role of the object within its cell environment. Finaly, the correction of a phenotype generated by the dysfunction of a node can require the modification of another node of the same network rather than a modification of the dysfunctioning node itself (See the example of the subthalamic nucleus tetanus in the treatment of Parkinson's disease).

At the opposite of the analytic reductionism lies the "blackbox" approach, largely used in Computational Neurosciences, that sought to isolate functional modules and to reproduce the global behaviour of the system. One consider only inputs and

outputs of each module, using transformation laws to link both. This approach grounds the formal neuron network models. While those methods gave acceptable results in many cases, they suffer form the lack of understanding of the underlying biological system. In addition of the intellectual frustration, that triggers a rigidity uncompatible with the current knowledge in Cell Biology. The laws which are used have been determined in particular cases, and then extrapolated. One result is that unforecast phenomena cannot emerge. A good example is the supposed unidirectional propagation of action potentials along the axons. Today , one knows that action potentials also propagate to dendrites, and this back-propagation plays a crucial role in plasticity phenomena such as Long-Term Depression and Long-Term potentiation. Moreover, modulatory systems, such as acetylcholine, dopamine or serotine systems are largely ignored in those models. Finaly, as they do not possess accurate representations of molecular networks, those models cannot simulate cellular deficits, and *a fortiori* help to design patches.

The availability of complete genomic sequences support the expectation that complex biological phenomena and systems can be understood completely. Rapid development of methods to decipher protein-protein interactions, and quantify the activities of macromolecules in vivo permits to envision the construction of virtual cells. With the avent of "*Systems Biology*", one can use th analytical descriptions of each elementary brick to reconstruct entire systems. This is really a change of paradigm, from a phenomenologic to a mecanistic philosophy. On the contrary of the approaches described above, such models permit to take into account the role of each molecular actor, and to identify, for instance, sub-cellular bottlenecks, robust elements, or instead sensitive steps.

Of course, there is absolutely no way to understand such large systems just by juxtaposing what we know about all the components. Only numerical simulations can reproduce and order the diversity of possible dynamical behaviours. Such networks of reactions can present stable, periodic or chaotic activities, according to the initial conditions and the kinetics of inputs. In such a situation, paper and pen have definitively to be replaced but the computer.

Systems Biology and neuronal signaling: the DopaNet project

The usefulness of Systems Biology models depends a lot on the quality of numerical data used. Although there exist several methods of parameter estimation/optimisation, their accuracy will be inversely proportional to the uncertainty of the model. It is therefore necessary to access large quantities of quantitative data of assessed quality, regarding every cell phenomena. .

DopaNet (http://www.dopanet.org), a project recently selected as a network of the European Science Foundation, aims to mobilise the European scientific expertise in Molecular and Cellular Neurobiology, in order to investigate precisely and quantitatively all the aspects of neurotransmission in a specific neuronal system. The selected system is made up of the dopaminergique neuron of the mesencephalon together with two of its targets, the GABAergique medium spiny neuron of the striatum, and the pyramidal glutamategique neuron of the prefrontal cortex. Those neurons are part of a bigger system called basal ganglia, which plays a major role in many emotional and cognitive functions, and as such is affected in numerous neuropathologies. The destruction of the dopaminergic neuron triggers Parkinson's disease. The destruction of the medium spiny neuron causes Huntington chorea. The activity of the pyramidal neuron of the prefrontal cortex is strongly affected in certain forms of schizophrenia. Finaly the system is at the very center of most of the drug addictions (opioïds, cocaïne, amphétamine, nicotine, cafein etc.). Because of this physiologic and medical importance, a large quantity of data is already available in the literature, and a portion of it could be reusable. The system is massively parallel, with a conservation of topology between cortical, striatal and mesencephalic structures. Hence, from elementary models, we can envision a scale-up toward more physiologic situtations.

The members of the consortium will have to study the various modalities of neuronal signaling regarding the molecules, the supra-macromolecular assemblises, the nervous cell, and interacting neurons. Those data will be incorporated within numerical models. The ultimate goal of the DopaNet project is to build realistic simulations of signaling networks at the levels of the synapse, the neurone and the micro-circuit. Such models should permit a better understanding of the neuronal signaling, but will also seek to reproduce neuronal disorders *in silico*. Hence, one could decipher how a particular molecular defect can trigger a pathological phenotype, and predict the effects of pharmacological treatments. The work of the consortium are just starting, and are mainly limited by the absence of significant funding. However, several projects already took off, among which the construction of a database of functional data about the proteins involved in signaling pathways, and the design of an ontology specific for the neuronal cell. Functional simulations should progressively take place at the EMBL-EBI.

Annex A: Selected bibliography on Computational Systems Biology

US Bhalla, R Iyengar. Emergent properties of networks of biological signaling pathways. (1999) *Science* 283: 381-387.

ME. Csete, JC Doyle. Reverse Engineering of Biological Complexity. (2002) *Science* 295: 1664-1669.

Most of the ideas of Systems-Biology are coming from cybernetics, and this review takes the point of view of the engineer. Although analogy should be used with caution (one cannot afford to sacrificed aircrafts to test them, while the oyster lay down millions of egg, most of them doomed), the "legome" carries many challenging ideas.

Davidson et al. A Genomic Regulatory Network for Development. *Science* 295: 1669-1678.

An example of Systems-Biology investigation, linking large-scale experiments and simulations.

H Kitano. Computational systems biology. (2003) *Nature* 420: 206-210.

A pretty general, and recent, review on the subject.

HH. McAdams, A Arkin. Stochastic mechanisms in gene expression. (1997) *Proc Natl Acad Sci* 94: 814-819.

One of the classical papers on stochastic simulation of genetic regulation, using the lytic/lysogenic balance of the lambda phage.

D Noble. Modeling the heart -from genes to cells to the whole organ. (2002) *Science* 295: 1678-1682.

Systems Biology is not just about molecule interactions. The concept is scalable, and the work of Dennis Noble and Peter Hunter shows that one can also build virtual organs from elementary cells.

L Sánchez, D Thieffry. A logical analysis of the Drosophila Gap-gene System. (2001) *J Theor Biol* 211: 115-141.

An application of the logical formalism of genetic networks. Logical algebra is an excellent alternative to the differential calculus when one has very limited quantitative knowledge about parameters.

JJ Tyson, K Chen, B Novak. Network dynamics and cell physiology. (2001) *Nat Rev Mol Cell Biol* 2: 908-916.

A review describing the modelisation of cell-cycle, a test case in Systems Biology.

G von Dassow, E Meir, EM. Munro, GM. Odell. The segment polarity network is a robust developmental module. (2000) *Nature* 406: 188-192.

Beside the conclusions of the paper on the robustness of a network, that is the conservation of the final result across a wide-range of parameter values, this article carries an important message on "tuning". If parameters are randomly chosen, and half the values of each parameter are just fine, a network involving 30 parameters will exhibit the right behaviour in only 1 in a billion parameter sets! (0.5^{30})

Annex B: Selected simulation tools used in Computational Systems Biology

E-CELL (http://www.e-cell.org/)

One of the first effort to build a software able to simulate a complete cell. It possesses a graphical user interface. The simulation are run using systems of ordinary differential equations. The last version allows to define several compartments in the cell. Read/write Systems Biology Markup Language (SBML).

M Tomita et al.. E-CELL: software environment for whole-cell simulation. *Bioinformatics* 15: 72-84.

Genesis (http://www.genesis-sim.org/)

The most elaborate whole-neuron simulator. It allows the construction of complex 3D neuronal shapes. It is mainly devoted to simulate electrical phenomenons.

Wilson, M. A., Bhalla, U. S., Uhley, J. D., and Bower, J. M. GENESIS: A system for simulating neural networks. In: Advances in Neural Information Processing Systems. D. Touretzky, editor. Morgan Kaufmann, San Mateo, CA. pp. 485-492. (1989)

Gepasi (http://www.gepasi.org/)
One of the earliest simulators based on systems of differential equations, it is a standard in the field. It comes with a graphical user interface. The last versions include several parameter optimisation approaches (gradient descent, simulated

annealing, genetic algorithms). A parallelised version, called COPASI is under development. Read/write Systems Biology Markup Language (SBML).

P Mendes. GEPASI: a software package for modelling the dynamics, steady states and control of biochemical and other systems. (1993) *Comput Applic Biosci* 9: 563-571.

P Mendes, Kell DB. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. (1998) *Bioinformatics* 14: 869-883.

Jarnac (http://www.cds.caltech.edu/~hsauro/Jarnac.htm)
Initially called SCAMP II, it is the evolution of one of the earliest simulators based on systems of differential equations. It comes with a graphical user interface, which permit to draw the network of reaction. Read/write Systems Biology Markup Language (SBML).

Sauro H.M and Fell D.A. SCAMP: A metabolic simulator and control analysis program. (1991) *Mathl. Comput. Modelling*, 15: 15-28

MCell (http://www.mcell.cnl.salk.edu/)
One of the more elaborate simulator of cellular processes. Developed to simulate synaptic events, it allows for complex 3D shapes, and the representation of every single molecules, including ions.

TM Bartol, BR Land, EE Salpeter, MM Salpeter. Monte Carlo simulation of miniature endplate current generation in the vertebrate neuromuscular junction. (1991) *Biophys J* 59: 1290-1307.

StochSim (http://www.zoo.cam.ac.uk/comp-cell/StochSim.html)
General simulator of (bio)chemical reactions, its main interest resides in the possibility to simulate multistate reactants, thus dramatically reducing the number of reactions to simulate. The stochastic algorithm is molecule-based rather than reaction-based (classical Gillespie algorithm). Each molecule can thus be simulated, allowing for spatial simulations. It comes with a graphical user interface. Read/write Systems Biology Markup Language (SBML).
CL Morton-Firth, Bray D. Predicting temporal fluctuation in an intracellular signalling pathway. (1998) *J Theor Biol* 192: 117-128.
N Le Novère, TS Shimizu. STOCHSIM: modelling of stochastic biomolecular processes. (2001) *Bioinformatics* 17: 575 576.

Systems Biology Workbench (http://sbw.sourceforge.net/)
Software platform developed by the ERATO-Kitano project in Systems Biology. It is a client-server distributed infrastructure, entirely modular. One can launch one or several servers, and then "plug" various modules such as model drawing applications, simulation engines, data analysis tools. The various modules use a Remote Process Control (RPC) system based on the Systems Biology Markup Language (SBML).

Hucka, M., Finney, A., Sauro, H., Bolouri, H., Doyle, J., Kitano, H. The ERATO Systems Biology Workbench: AN Integrated Environment for Multiscale and Multitheoretic Simulations in Systems Biology. Foundations in System Biology, ed, Hiroaki Kitano, MIT Press, (2001)

BioSPICE (https://community.biospice.org/)
Software platform developed with the help of a very large funding from the US DARPA. As SBW, it is a modular framework. It is accepts three different communication systems between modules, the Open Agent Architecture (http://www.ai.sri.com/oaa/), SBW, and the direct use of NetBeans (http://www.netbeans.org)

Garvey TD, Lincoln P, Pedersen CJ, Martin D, Johnson M. BioSPICE: access to the most current computational tools for biologists. (2003) *OMICS* 7: 411-420.

the Virtual Cell (http://www.nrcam.uchc.edu/)
Very elaborate whole-cell simulation software. Run entirely through the internet. Read/write Systems Biology Markup Language (SBML).
J. Schaff, C. Fink, B. Slepchenko, J. Carson, L. Loew. A , A general computational framework for modeling cellular structure and function. (1997) *Biophys J.*, 73: 1135-1146

Jaap Heringa - Bioinformatics for CSB at the Free University of Amsterdam

The Human Genome
What's in a genome?

Noble (2002) "Genes code for protein sequences. They do not explicitly code for the interactions between proteins.."
Information resides at level of protein interactions within context of subcellular, cellular, tissue, organ, and system structures.
Need to filter protein interaction data
We can therefore only compute and model interactions

CSB creates Bioinformatics challenges

Time dimension of system
Spatial information within system
Including the environment
Integration and modelling of heterogeneous data
Cell factory: get alternative modelling strategies in addition to concentration-dependent equitations
Because of vast data space in CSB, we need model-driven data collection

CSB versus bioinformatics

Ulam: "Don't ask what mathematics can do for biology but ask what biology can do for mathematics"

What can bioinformatics do for systems biology?
Integrative bioinformatics, data housekeeping, all basic methods
What can systems biology do for bioinformatics?
Enhanced inference methods ("super BLAST")

CSB versus bioinformatics

Requirement of standards for data, models, interactions, communication, etc.
"First standardise a car and then worry about inventing the wheel".

Databases

Structuring database while collecting data is difficult; data (structures) change all the time (A. Valencia yesterday)

Integration

Data integration

Need to integrate methods (models)
Is difficult (M. Sansom yesterday)

Structuring database while collecting data is difficult; data (structures) change all the time (A. Valencia yesterday)

Computational Systems Biology; a European spearhead

A vision; Hans V. Westerhoff, BioCentrum Amsterdam

Why Systems Biology

Culminating in complete genome sequences and genomics, molecular biology, biochemistry and biophysics have led to appreciable understanding of the macromolecules of living cells and to an impressive number of tools. The tools enable one to obtain much more such information when needed. However, obtaining *all* information about *all* molecules in *all* organisms remains too costly, and may limit the seeing of the forest for the ever-increasing number of trees. What appears needed is a focus on the original and true issues, such as the understanding of how living organisms function, of how they sometimes dysfunction (such as in disease), and how their function can be improved both in therapeutical and in biotechnological settings.

What seems to limit the understanding of function now is the phenomenon that much of the function of living organism comes about in the complex interactions of the macromolecules. It is the (lack of) understandings of these complex interactions, i.e. the systems biology more than the molecular biology, that is now limiting progress.

What is Systems Biology: an operational definition

There are various definitions of Systems Biology [cf. www.systembiology.net]. Yet, it is not a vague discipline. Systems Biology is neither the Biology of Systems (which is Physiology), nor the physical-chemistry and mathematics of their components (modern molecular biology), it is the in between. It focuses on the new properties, important for biological function, that arise in the interaction of the components of Biological systems, i.e. that are not present in those components in isolation.

The in-between can be at various levels of biological organization. One that is particularly acute, thanks to the explosive advances in genomics, addresses the level between macromolecules and the simplest form of autonomous life, i.e. single living cells, such as microorganisms and tumor cells. At the level between organisms and ecology, there is another example of Systems Biology. We shall focus on the former example.

Deliverables of Systems Biology

For good reasons, research agencies require scientific activities to lead to results that are important for society. Without guaranteeing delivery within two years, we here mention a number of deliverables:

Discovery of new scientific principles that govern at the system level and not at the molecular level (: high quality science leading to Nobel prizes, Fields medals)

Insight in the pathology of multifactorial diseases (e.g. cancer, type-II diabetes, rheumatoid arthritis, heart failure, infectious diseases) and in the diverse pathologies of unifactorial diseases against the backdrop of polymorphisms.

Multifactorial, subtle therapies for various diseases.

New drugs deriving from network-based drug design

New drugs and strategies to combat multidrug and antibiotic resistance

Computer models (and -replica) of patients helping to manage their disease and to pretest and optimize therapies.

Much reduced frequency of animal experimentation through substitution by computer models

Increased insight of the public in genomics research through layman-accessible computer simulations (and ‘games’) of living organisms, of research issues and of therapy and biotechnology

Support of companies *vis-à-vis* regulatory agencies such as the FDA which will soon require computer model validation in addition to experimental validation of drug descriptions

Insight in how functional systems can work that are evoluationarily stable; inspiration for man made society and ethics discussions

What is needed for Systems Biology?

Computation

New behavior of systems relative to their components arises through the nonlinear interactions of the latter. Such nonlinearity cannot be understood through the standard intuition, but requires assistance by computations.

Experimentation

The behavior of nonlinear systems depends on their operating point and on the magnitudes of their parameter values. These need to be determined quantitatively and sufficiently accurately, which has not been the priority of molecular biology until now. Therefore a new line of experimentation is needed, part of which should be directed towards experimentation inside living cells.

Conceptual advances

A computer replica of a living organism has the tendency of being equally unintelligible as the original. Therefore new conceptual tools are needed to facilitate the understanding of biological complexity. Already existing examples of such tools include metabolic and hierarchical control analysis, modularization concepts (such as elementary modes), stability analysis, but more will need to be developed.

Their integration

The above lines of Systems Biology will need to develop in an integrated manner in a procedure that also includes discovery, hypothesis, validation and falsification.

Model Systems

Systems Biology requires the integration of much and precise information about a system, which is difficult and expensive to obtain, and which requires the collaboration of much man power from many, diverse disciplines. Because of this breadth, focus of much of the activity on a limited number of models systems is required. The model systems should be selected on the basis of: (i) experimental accessibility, (ii) possibility to obtain the information needed by the computation system biology, (iii) relevance, (iv) the existence of scientifically exciting systems biology issues in the, (v) the possibility that Europe can contribute substantially, possible in a leadership role.

Systems Biology is a discipline in development. Although appreciable roots of Systems Biology lie in Europe, North America and Japan have taken strong positions already. Therefore, we should distinguish model systems in which Europe might take the lead, and model systems where Europe may be an equal partner to the Japanese and the Americans.

Where Europe may lead:

Lactococcus lactis (model prokaryote; simple model system; substantial Dutch, Danish and French initiatives); thorough industrial (biotech) interest

S. cerevisiae (best-known eulkaryote; first sequenced eukaryote, largely thanks to a European effort; much of the system biology of this organism reside in Europe); thorough industrial (biotech) interest

The hepatocyte (model mammalian cell; large German initiative; tremendous interest pharmaceutical industry)

Metabolic and Hierarchical Control Analysis (conceptual method; historically led by European groups)

Silicon cell (computer replica of parts of living cell; here in the sense of *precise* replica Europe leads)

Where Europe should participate in world-wide programmes

Escherichia coli (best known model prokaryote; existing world-wide International Alliance for *E. coli* Alliance; IEcA)

Virtual cell: modeling tools for molecular cell systems biology connected to the SBML initiative

Computational approaches

Europe can certainly make a leadership contribution in Computational Systems Biology. Recent American work has excelled in applying exiting engineering views to cell biology. They have lacked specificity. They did not always address reality.
Because of its tradition of collaboration between research groups, Europe may well lead in the challenge of making computational systems biology contact to the reality of molecular cell biology. The following types of programs could be characteristic:
The silicon cell: precise replica of living cells. For the creation of these through collaboration is required between computional scientists and experimnetall biologist.
Interactive Systems Biology: modeling activity in direct contact with experimentation; computation aided experimental design; experimentation based model optimization
Concepts for systems biology: theoretical tools phrased in terms of molecular cell biology have always been developed most in Europe. Further developments of these for the new systems biology are a good bonanza for Europe (examples: Control Analysis, non equilibrium thermodynamics)

What requires EU support?

Coordination
Postdoc grants
Training grants
Network of excellence:
Many types of support are needed. Substantial support is already becoming available from national governments, many of which have understood the potential of Systems Biology for the European health and wealth. The bulk of Systems Biology does not require EU support therefore. What is needed, is support for coordination of the national activities. Such support should not only comprise support for visits, but also support for salaries of transdisciplinary and transnational scientists, i.e. scientists that venture to a laboratory in a different EU country with a different aspect of Systems Biology (e.g. an experimental cell biologist going for one or two years to a lab in a different country to models his system of interest; postdoc or sabbatical, or mobility). In addition there should be a network of excellence coordinating the best national CSB initiatives. And there should be substantial support for training activities.

Cédric Notredame

Digging up the foundations of in-silico cell simulation

Short description

The simulation of living organisms has become an increasingly popular topic [1], owing to recent availability of genomic, expression and interaction data generated by large-scale projects. With this data at hand, time has come to tackle large-scale simulation and merge all the many pieces of the puzzle into one single unified model of life. The successful assembly of this model will probably be regarded as a major milestone of modern biology.

The pace of activity in field of large-scale biology and data acquisition clearly makes it a priority to start looking for efficient and realistic modeling solutions. Yet, it is also worthwhile asking whether available data is already abundant enough to simulate a living organism? One may argue that too many genes still lack clear functional attributes. For instance, half of the E.Coli proteome remains of unknown function. This data will have to be acquired, either through curation and merging of existing data collections or by experimental means.

While maintaining a homogenous database of functional knowledge could be done in a SwissProt manner, the gathering of experimental data will require some new kind of rationalization. For instance, one could use static bioinformatics methods to flag important genes so that they can be characterized in priority and incorporated within complete models. Large-scale functional characterization could use a targetDB-like strategy, to fuel the emergence of a collaborative, yet still competitive, environment. Such a context may also be a good opportunity to systematically confront wet and dry-lab results, and convince wet lab people of the relevance of computer-based techniques.

Well-maintained data collections will undoubtedly play an important role in the building of complex models, and as pointed out by a previous EU report (Bioinformatics Workshop Report, June 2003), care should be taken in maintaining these basic structures. We also argue here that the current results of comparative genomics should be used as a first milestone for defining the goal of complex system simulation. These simulations must deliver more than static bioinformatics does. This will only be achieved if efficient means of assessing complex models reliability are made available. For this purpose, one may consider the creation of an established institution whose mission would be to run the appropriate experiments so that models predictions can be verified.

We should also consider developing adequate static bioinformatics methods as a cheaper mean of model assessment. For instance, if a model predicts an important synergy between two metabolic pathways, one should start doing comparative genomics before making an in-vivo assessment of the new hypotheses.

What is the main research area?

Multiple genome comparisons.

How complex is the biological system?

A bacterial genome with probably less than 10.000 different macro-molecular species and 100.000 metabolites species all arranged in about 200 pathways.

How did you choose your research area?

Evolution and the comparison of close and distantly related biological structures is a very powerful mean of identifying and understanding important mechanisms. Starting from multiple sequence comparison I naturally moved to complete genome comparisons.

HOW IMPORTANT IS IT: What are the main contributions expected?

I expect complex system simulation to reveal new unknown mechanisms, or reveal the importance of conditionally express genes that could become new drug targets. Comparative genomics will play a key role in rationalizing the targets choice for functional characterization, it will also constitute an efficient in-silico means of validating resulting hypothesis, looking for the evolutionary trace associated with newly uncovered mechanisms.

Do you have enough data

The lack of functional data is a major limitation. The careful choice of a few hundred targets and their functional characterization must be made a priority. Only 50 % of a typical bacterial proteome is associated with functional attributes, not to mention the scarcity of functional data related to the RNAome. We need an active gathering of functional information at all levels, from biochemistry (which enzymatic activity?) to integrated function (which metabolic pathway, which interactor, which regulation?). We also need a better understanding of the non-enzymatic proteins, much harder to study and characterize but probably just as important for proper modeling of cellular processes.

Gathering this data in a coordinated and efficient manner will require new experimental designs. The ideal would be an international collaborative/competitive framework, like targetDB, where results obtained worldwide are collected and made available to the community in real time. This would help rationalizing strategies, avoid useless duplications and stimulate healthy competition in key areas. Furthermore, this strategy would contribute to the emergence of a complete and consensual definition of the notion of biological function.

Static bioinformatics can help a lot. For instance, comparative genomics could be used to prioritize functional targets so that the community focuses on the genes predicted to be important for our understanding of cellular processes. This framework would also be an ideal benchmark for double blind evaluation of bioinformatics methods, and given a targetDB strategy the community could set up a CASP-like competition. In CASP[2], bioinformatics groups compete to solve the structure of proteins that are being experimentally solved, while in the ***functional CASP***, bioinformatics groups would compete to predict the function of a target that is otherwise being experimentally characterized.

This type of objective confrontations between in-vitro and in-silico biology will certainly pave the way to establish of a bond of trust between experimentalists and dry lab biologists.

Is the problem tractable computationally?

Providing all the required data would be available, methodological progresses at the mathematical level are still needed. For instance, we need models that can take into account complex discrete phenomenon such as interaction, translocation or cell compartmentalization.

How far can your project go in terms of experimental method? (Explain, Predict, Control)

Results that can be achieved with comparative genomics and static bioinformatics must set the lower bound of our expectations from complex system simulation. These new simulations must aim at explaining and controlling phenomenon beyond the scope of simple analysis.
Explain
Comparative genomics makes it possible to discover new genes, new metabolic pathways, and thus explain specific adaptations. We expect complex modeling to go further, and explain why combination of seemingly unrelated cellular function can yield new adaptative abilities to an

organism. As far as comparative genomics is concerned this should result in new hypothesis of gene co-evolution easily tested in-silico, using standard static bioinformatics methods.

Predict

In a recent study, our laboratory used comparative genomics to design a cell-free media suitable for growing an intracellular pathogen [3]. This prediction was made on the basis of genome-genome comparisons that revealed the absence of several bio-synthesis pathways. We expect complex modeling to go further and reveal subtle sensitivities of a pathogen to basic media parameters such as pH, temperature, salinity and so on.

In a context were many research groups are bound to work on similar predictive models, a standard benchmarking procedure will be needed. Using the predicted concentration of a few key metabolites will be the fastest way to assess the validity of new models. Accurate testing of these concentrations is a complex task. Leaving it to a specialized institution would be the best way to ensure reliability, objectivity and cost efficiency. Its mission at the European level would be to do in vitro testing of models. This institution would be a biological equivalent of a synchrotron facility for X-ray diffraction.

Control

From a comparative genomics perspective, control can be obtained by identifying a gene conserved over many species. These conserved genes often make interesting drug targets. For instance, one may identify new targets for antibiotics by looking at genes that are conserved among bacteria but very different from their closest relatives in human. This strategy has successfully been used by several post genomics projects [4].

I expect complex models to go further and reveal conditionally essential genes. These genes, non essential under normal conditions, will be revealed by the model to be key players during the infection. As such they may become new drug targets.

Choose right software tools for problems complexity level

So far, I have been using static bioinformatics, with an emphasis on multiple comparisons. These methodologies still need to be developed in order to ease the spread of experimental data on sequences not yet functionally characterized.

Dealing with experimental data, I would expect statistical languages such as R or S to become increasingly important. I would also expect the use of Hidden Markov Model and Bayesian statistics to develop significantly, and go well beyond the area of sequence analysis.

Database requirements

It is essential to help knowledge building. Models and simulations must be distributed in such a way that they can easily be used and adapted by other members of the community. Standards must be defined for platforms, languages and other conceptual tools, to alleviate the current state of anarchy. For instance, a rapid scan of the current situation reveals that most tools use different languages, and sometimes exotic compilers or interpreters. Many of these simulation workbench also require Matlab, a tool whose license is well beyond the financial capacities of most laboratories. Supplying the community with a proper and affordable bench-work is a priority.

Standards must also be defined so that updating a model with new parameters or new molecular specie becomes trivial. Of course, this will also require an active reflection on the way a protein, a gene, an RNA or any other molecular species can be described within a database.

What do you need, and what is available, in the areas of hardware and human resources?

A framework for generating functional information would be the first requisite. This should be a distributed effort, where individual laboratories would register and stick to some common guidelines.

This effort would be coordinated by a European agency that could also run a CASP like competition for protein function prediction.

A bench-work Cheap, easy to run and distribute, evolutive and whose maintenance would be insured on the long-term by an established institution.

A structure where models can be tested will be instrumental. A framework must be set up where models can be evaluated in-silico and experimentally according to their predictions. The experimental validation should be made feasible by using simple criteria such as key metabolites concentration. For the sake of reliability and objectivity, this model testing aspect would better be left to some specialized organism that would receive the conditions, run an experiment, measure the various parameters and confront them to the prediction, in what could be double blind experiments.

Massive computing power may help, but only when enough data will be available to run realistic simulations. Yet, from the moment simulations will become successful at the cell level, a rapid exponential increase in computing power requirement is almost guaranteed. I would expect this need to arise in 5 to 10 years and at this point in time, mature GRID-like technology may prove an asset.

1. Reed, J.L. and B.O. Palsson, *Thirteen years of building constraint-based in silico models of Escherichia coli.* J Bacteriol, 2003. **185**(9): p. 2692-9.
2. Bourne, P.E., *CASP and CAFASP experiments and their findings.* Methods Biochem Anal, 2003. **44**: p. 501-7.
3. Renesto, P., et al., *Genome-based design of a cell-free culture medium for Tropheryma whipplei.* Lancet, 2003. **362**(9382): p. 447-9.
4. Claverie, J.M., et al., *In Ssarch of new anti-bacterial target genes: a comparative/structural genomics approach.* Comb Chem High Throughput Screen, 2002. **5**(7): p. 511-22.

COMMENTARY ON CSB Meeting
William Saurin
Genomining

1 What is CSB.

Describing what an emerging field isn't is always easier than to describe what it is.

Computational Systems Biology (CSB) is not Bio-informatics which is mainly concerned with information retrieval or with information extraction. By information retrieval I mean finding biological information in huge amount of data by building systems (e.g. blast, fasta or SRS) or databases (e.g. the embl nucleotide data bank, swissprot or unigene). By information extraction I am referring to those programs which elucidate new information about unidentified biological features (e.g. genscan or clustalw).

A CSB study must have a dynamic dimension: data gathered must not only describe temporal co-occurrences of measures for various biological phenomenon or objects, they must somehow describe the trajectory of a biological system through time.

2 CSB goals.

CSB aims at building models representing systems or subsystem of living organisms (e.g. a cell, a couple of cells interacting, a molecular subsystem accomplishing a given function, or a metabolic pathway). It is expected that those models can describe and predict the dynamic behaviour of such systems.

The predictivity of the models that are desired in CSB must not be achieved by sacrificing the understanding of those systems. Though it can be legitimate to use "black boxes" in the process of building these models, a complete explicitation must be striven for.

A reasonable objective of CSB research programs could be to model living systems or organisms the way they are and also the way they could be. Question like "what happens if a signaling pathway is changed for another?" could be answered by CSB models. In principle, independent of the ethical issues involved, CSB should be able to aid in the design of artificial organisms meeting some arbitrary specification. This can be seen as a test of the understanding of the modeled systems.

3 Some difficulties.

The relations between experimentalists and theoreticians are not clear in biology, both communities have different agendas and time frames. To overcome this difficulty one could not incite laboratories of both kinds to collaborate. One could also incite researchers of both kinds to work in the same laboratories or the laboratories to recruit researchers of both kinds.

CSB aims at describing the mechanisms that explain dynamic phenotypes of living organisms. However, such phenotypes must be considered only if they have some functional meaning for a given organism. The paradox resides in the fact that it is much easier to define what a function is at a high conceptual level (e.g. reproduction, nutrition...) than at the level of genes or proteins: is the function

of the lactose repressor "to bind the operator in absence of inducer", or "to block the synthesis of proteins of the lac operon proteins in absence of substrate"? It is clear that the second proposition makes more sense than the first one, but it is also a far less objectivable proposition. The first proposition says "what" the lac repressor can be used for, the second one explain the role it plays in a (rather simple) biological system. We are very far from having a rational and objectively defined nomenclature of biological functions, and indeed, one can even wonder whether it should be the case.

4 CSB needs

CSB will bear upon rather important computer program developments. One can hope that these developments will be made with highly professional standards but we must keep in mind that most of the program developments in academic research laboratories are performed by graduate students or by post-docs that are mainly interested in obtaining biological results and not in creating robust software. This is probably a field where SME could bring some service to the academic community.

Data standardisation is another difficult point. It is clear that CSB research programs will have to integrate very important volumes of data and one can expect that they will come from many different experimental benches. It will be important to define standards, but we must keep in mind that premature definition of standards can lead to some important observations being missed. DNA sequence is rather well standardized, however, at the opposite end of the spectrum, protein-protein interactions data are far from being homogeneous. An intermediary possibility between data standardisation and total anarchy could be to define Standard Operating Procedures (SOPs) and to attach to each piece of data the reference to the SOP used to acquire it.

5 Expected industrial repercussions.

We can expect that the repercussions of CSB will be felt in several industries.

In the pharmaceutical industry, CSB should lead to better and earlier predictions of possible adverse effects of drugs. In the agro-food industry, a better understanding of the biological systems used in production should lead to increased control of the bio processes involved. The chemical industry should expect the design of new bio processes for the synthesis of known or even new compounds.

We must bear in mind that these are long-term repercussions, and that any short-term expectation may be met with deception.

Ilias Iakovidis

eHealth – Past and future activities of the Commission

Directorate General Information Society, Components and subsystems. Applications.
eHealth Unit

1. From Medical Informatics to eHealth

Medical informatics: Methods and systems for the storage, retrieval, sharing and optimal use of biomedical data, information, and knowledge.

Called *Health Telematics* during the era of telecommunication: Emphasis on regional networks and telemedicine applications

Called *eHealth* during the era of Internet: Emphasis on the Internet based applications and patient empowerment

eHealth Mission: Improve access, quality and cost efficiency of health care through new paradigm of "*Person-Centered health delivery systems*".

2. BIOINFOMED Study

In November of 2001 a study was launched to continue the findings of the conference of December 14,2001: *Synergy between Research in Medical Informatics, Bio-Informatics and Neuro-Informatics*.

30 experts worked 1 year to present the potential of the synergy between Medical Informatics and Bioinformatics and proposed a roadmap for collaboration called ***"Synergy between Medical Informatics and Bio-Informatics: Facilitating Genomic Medicine for future healthcare"***

Medical informatics	***Bio-informatics***
Electronic Health Records	Functional Genomics
Medical Imaging	Proteomics
Clinical Decision Support	Techniques
Telemedicine	Computational Biology
Biomedical-informatics	

3. HealthGRID & FP6

Application of the existing GRID and GRID-like technology in the Health sector for

- timely and secure access to (distributed) patient data
  ex.: Electronic Health records, Regional Health Information Networks
- interoperability of databases of heterogeneous content (biology and medicine) for research purposes
  ex. enabling new knowledge discovery (research, drug design), better guidance and information (healthcare professionals)
- computing intensive applications and knowledge discovery
  ex. imaging, simulation and modelling

www.healthgrid.org

4. HealthGRID applications in the *e*Health unit

*e*MOLECULE
Molecular biology databases - knowledge discovery
Molecular Medicine (e-Pharmacology)

*e*CELL
Pathway simulations, virtual cell - computing power

*e*INDIVIDUAL
Medical imaging
Combination of genetic and clinical data

*e*POPULATION
Environmental Influences

A1. WORKSHOP ORGANISATION

This workshop on "Computational Systems Biology (CSB) - Its future in Europe" was organised by the Research Directorate-General of the European Commission, in the context of a series of workshops supporting the European Research Area (ERA).

A group of experts was invited to meet and discuss this topic, and to provide a summary of the background, problem areas, current situation, and guidelines and options for action by the Commission and policy makers and organisations in Member states, and for researchers themselves.

A workshop "Terms of Reference" and documents and references were provided before the workshop.

Attendees submitted short, highly condensed summary papers of their contributions, which are included here with the workshop summary, giving their points of view. Some were submitted as powerpoint presentations, and the text has been abstracted, and some summaries are based on notes from the presentations. A questionnaire was also completed by several of the attendees concerning actual and proposed projects.

The workshop consisted of presentations by invited speakers. These presentations were followed by open discussion.

Members of the Commission services, who provided background information on relevant activities, also attended the workshop.

This workshop report was written and assembled by the Rapporteur and Chairpersons and edited by members of the Commission services, in particular the officers responsible for the workshop, based on summaries of the workshop discussions, inputs from the chairpersons and participants during and after the workshop, and the contents of the submitted papers.

The executive summary represents a large convergence of views. Where they occur, significant differences are explicitly presented as such.

A2. TERMS OF REFERENCE

❖ **Philosophy and Definition of Computational Systems Biology for this workshop**

- In the world of modern biology, especially molecular biology, huge amounts of data are being generated: digital(sequences), analogue(signals), visual(scanning, microscopy), written (articles).
- The analysis of systems can be as wide as all of biology, and computational means are needed.
- Therefore, there are many pitfalls:
  - Getting so wide a definition as to be meaningless.
  - Defining it so narrowly that research and computational tool development separates from data analysis, or that tool development much too broad for data (e.g. full cell simulation without knowledge of key processes and parameters.
- Leading to the following pragmatic approach: Let us consider the narrower area of:
  - **Computational Systems Biology instead of just Systems Biology; and do it:**
  - **in the context of Fundamental Genomics** (Fundamental knowledge and basic tools for Functional Genomics in all organisms), where our specific programme states: (see http://europa.eu.int/eur-lex/pri/en/oj/dat/2002/l_294/l_29420021029en00010043.pdf) Fundamental knowledge and basic tools for functional genomics in all organisms to foster the basic understanding of genomic information, by developing the knowledge base, tools and resources needed to decipher the function of genes and gene products relevant to human health and to explore their interactions with each other and with their environment. Research actions: ..... Bioinformatics to enable researchers to access efficient tools for managing and interpreting the ever increasing quantities of genome data and for making it available to the research community in an accessible and usable form. Research will focus on developing bioinformatic tools and resources for data storage, mining and processing; developing **computational biology** approaches for in silico prediction of gene function and for the simulation of complex regulatory networks;

❖ **Therefore, for this workshop:**

- **Offer speeches and 1-4 page written contribution and presentations** (see for example ftp://ftp.cordis.lu/pub/lifescihealth/docs/bioinf_workshoprpt_2003_06_30_final.pdf
- **Identify your area by filling in or referring to the attached QUESTIONNAIRE by specifying "***horses for courses (English idiom***")**
  - Choose the Research Area (*choose the racetrack*)
  - Choose and develop research tools, e.g. software systems (*pick the horse)*
  - Develop appropriate input (databases) and support for the system (computers, systems biologists and programmers, analytic solutions, building, budgets) (*equi, feed the horse*)
  - Operate the software tools on the data, compare with experiment, test hypothesis and compounds so as to explain, predict and control. (*Run the horse race and cross finish line)*
  - Refer to the scientific method for experiments: Explain, Predict, Control

❖ **Conclusions :**

- By looking at the elements of the problem we can identify most promising areas
- Many tools will be common to many systems
- A "systematic" and complete approach, from experiment to data to analysis is as important as the choice of area for study (e.g. which cell model?)
- For both researchers and policy makers, we need to find the right balance between resources and broadness of research program.

❖ **Expected Outcomes of Workshop**

- Input for Documents and Actions to produce a report and action plan
- Areas and topics for present and future calls for proposals
- How to co-ordinate, complement existing and future research projects
- How to develop links between interested groups
- Inform Commission staff of the state of the field

A3. AGENDA

FIRST DAY - Wednesday 10 September

10:00 INTRODUCTION
10:00 - 10:10 - Welcome and Introduction from organisers, and Introduction of participants
Opening speech
10:10 Manuel Hallen (Unit Head F.4 Fundamental Genomics) - CSB and Research in Fundamental Genomics; CSB-related conclusions of the bioinformatics workshop
Workshop Rapporteur Mark Sansom

10:25 - Session 1 Chair - Hans Westerhoff
SESSION 1 - THE FOUNDATIONS - Current and already planned activities
Unifying theme: state of the art: the relation between the existing biology knowledge base and computational systems biology capabilities, state of the art of simulation and analysis computer tools, review of existing or just funded national and EU projects that have a computational systems biology component; application of existing computational tools, what are the centres of gravity? how do the tools relate the experimental outcomes to the data processing facilities?; links to genomics and related research.
10:25 - 12:20
SPEAKERS, FOLLOWED BY DISCUSSION
10:25 Hans Westerhof - Chairman Introduction
10:30 Ralf Herwig- EMI-CD-Platform for data integration and modelling of complex biological processes
10:50 Luis Serrano - COMBIO - An integrative approach to cellular signalling and control processes: Bringing computational biology to the bench.
11:10 Petra Wolff (30 minutes) - Integrating CSB programmes: the EUSYSBIO SSA and The German Research and Funding Program "System Biology" (Frank Laplace and Petra Wolff)
11:40 Norbert Hübner - Bioinformatics in FUNGENES, an Integrated Project
12:00 Stefan Hohmann - CSB in QUASI, a STREP
12:20 Discussion

13:30 - Session 2 Chair - Ralf Herwig
SESSION 2 - FUTURE REQUIREMENTS
Unifying theme: scope of computational systems biology, "half-way house" software platforms analysis of the needs in the scientific communities that employ systems biology tools: what are the actual and future needs in respect to storage and retrieval of biological information, including computational means and algorithms? What are the effects of technologies such as array technologies and their implications for databases, standardisation, software; possible other application avenues of processed genetic information and their implications in the development of bioinformatic tools: is it towards medical application (including drug development), etc. - analysis of the possible scientific developments in the field and in new research areas: what research solution might influence the appropriate development of computational systems biology tools.
13:30 - 17:00
SPEAKERS, FOLLOWED BY DISCUSSION
13:30 Lilia Alberghina -Cellular networks: new tools and approaches
13:50 Olaf Wolkenhauer - Simulating what cannot be simulated
14:10 Mark Sansom - Molecular Simulations of Membrane Proteins: Towards a Virtual Membrane
14:30 Uwe Sauer - Computational and experimental approaches in systems-oriented metabolic research

14:50 Age Smilde - Systems Biology as a platform for biomarker discovery
15:10 Coffee
15:30 Karl Kuchler - Training and education for CSB
15:50 Alfonso Valencia - Possibilities and limitations of the genomic information: biodegradation networks as a case study.
16:10 Shoshana Wodak - Bioinformatics requirements for Systems Biology
16:30 Michael Cahill - Embryonic stem cell differentiation: Possible elements of a CA.
16:35 Discussion
17:00 Finish of today's session

SECOND DAY - Thursday 11 September
9:00 - Session 3 Chair - Luis Serrano

SESSION 3 - STRUCTURING EUROPEAN CSB RESEARCH IN THE FUTURE

Unifying theme: what research areas should be addressed; what are the opportunities; how can research best be co-ordinated?
9:00 - 12:00
SPEAKERS, FOLLOWED BY DISCUSSION
9:00 Administrative announcements
9:10 Eero Vuorio- The FORUM for Genome Programmes Managers and the need for CSB
9:30 Steven M. Foord - The plans and needs of Large Pharmaceutical Industry for CSB
9:50 Nicolas Le Novere - The vital role of CSB in future neurobiology research
10:10 Coffee
10:30 Jaap Heringa - Bioinformatics for CSB at the Free University of Amsterdam
10:50 Hans V Westerhoff - Molecular System Biology in Practice
11:10 Cedric Notredame - A new (h)ome for protein function
11:30 Discussion

12:00 - SESSION 4- SUMMARY AND CONCLUSIONS - Session 4 Chair - Fred Marcus
12:00 - 13:00 Summary by 3 Session Chairs and Rapporteur (15 min each)
13:00 - 14:00 Lunch Break
14:00 - 15:00 Discussion
Research Policy (EU, National, International), Research Topics, New Directions): What are the key policy issues under discussion and what is current thinking?
15:00 Official Meeting finishes

A4. QUESTIONNAIRE ON ACTUAL OR PROPOSED OR HYPOTHETICAL PROJECTS

A questionnaire was circulated to the participants and observers, to try to get a uniform view across projects, programmes, and proposed ideas. The following table (formated into two parts for the 11 replies). This questionaire provides an overview of the various approaches, and also provides a cross comparison of various approaches.

Computational Systems Biology Workshop	**QUESTIONNAIRE**	SESSION - 2	Session - 2	SESSION - 3	Session-1	SESSION - 1	SESSION - 1
	NAME OF PRESENTER	Lilia Alberghina	Mike Cahill	Steve Foord	Ralf Herwig (www.molgen.mpg.de/~lh_bioinf)	Frank Laplace and Petra Wolff	Stefan Hohmann
	TITLE OF PROJECT OR TOPIC	Systems Biology of yeast cell cycle	Embryonic stem cell differentiation for innovative target discovery and validation	Discovery of Novel Drug Targets	EMI-CD - a platform for data integration and modelling of complex biological processes	'Systems of Life – Systems Biology'	Current and already planned activities: CSB in QUASI, a signal transduction STREP
	SHORT DESCRIPTION --------	Identification of the molecular basis of the more relevant regulatory circuits of cell cycle in yeast.	Pluripotent embroyonic stem (ES) cells represent an ideal model for systems biology applications. They represent the primordial mammalian eukaryotic diploid condition, most similar to single celled eukaryotes. Differentiation of precursor cells via asymmetric cell divisions leads to different tissue-type-like cell culture models, representing neural, cardiomyocyte, adipocyte or others. Genetically homogenous, these derivatives are related to the initial ES cells by a tree of linear descent that can be subjected to cladistic analysis and targetted intervention by e.g RNAi experiments. As well as providing a foundation for subsequent mammalian systems biology, we anticipate the delivery of condensed signatures of specific postranslational modifcation isoforms of proteins profiling developmental ontogeny, physiological and pathophysiological molecular signalling, modes of action/validation of drugs and toxicity and side effects.	Pharmaceutical companies derive most of their revenue from targeting relatively few classes of well studied proteins. They have relied on a 50 year legacy of academic research for the validation of these targets. This legacy is all but used up and novel sources of information are being mined. The pharmaceutic industry used to get >90% of its biological information from outside but this is changing. In house genomics and genetics data has to be processed on a scale that is new to the industry and it needs to be both integrated and interpreted.	We develop a software platform for modelling and simulation of biological processes such as gene regulatory networks, signalling pathways, metabolic pathways etc. The platform will have an information layer where exhaustive information is gathered and evaluated with respect to the biological objects under analysis, an analysis layer where this information is translated into mathematical models (networks) and a forward modelling system, where in silico experiments are performed.	The programme represents a new approach to research funding with a view to establishing systems biology in Germany and tapping its potential for future research and development in science and industry. To this end, an interdisciplinary network of centres of excellence will be developed which will weave the biosciences together with systems studies, engineering sciences, computer science and mathematics to form systems biology.	QUASI has an experimental and a theoretical component. Experimental studies: collecting quantitative, time dependent and spatial data on all steps in MAP kinase signalling (including generation of new tools). Theoretical part: mathematical models are used to represent pathway operation, to develop/confirm hypotheses for experimental studies and to achieve an understanding of the overriding rules that govern dynamic pathway operation. Information design is used to visualise results.
MAIN TOPIC	**AREA TO DISCUSS**	**ANSWER FOR SPECIFIC PROJECT**	**ANSWER FOR SPECIFIC PROJECT**			**ANSWER FOR SPECIFIC PROJECT**	**ANSWER FOR SPECIFIC PROJECT**
? Choose the Research Area	**What is the main research area?**	Cell cycle/ Systems Biology.	Embryonic stem cell differentiation with correlational functional analysis	Drug Discovery	Functional Genomics and Bioinformatics	Research will focus on the hepatocyte system.	Signal transduction through MAP kinase pathways, using yeast as a model.
? Level of complexity	**How complex is the biological system? Gene expression, protein-protein interaction, etc.; Metabolic / signalling pathways / elements of cell cycle; Whole cell modelling; Multi-cell - Physiological systems; Entire multicellular organism; OTHER**	Elements of cell cycle regulation and their linking with selected signal transduction pathways and ultimately with selected elements of metabolism and cell growth.	ES cells are biologically as complicated as any other mammalian cells, permitting all conserved eukaryotic processes to be modelled (cell cycle, etc.). The challenge/opportunity is including for the first time relevant human cell systems, next to more accessible mouse systems. The complexity of the ES differentiation system is highly expandable, in biologically important and controllable directions. Signal transduction events alter gene expression, leading to asymmetrical cell division, characterised by epigenetic heritable chromosomal states, with associated changes in gene and protein expression, physiology, and cell morphology. Gene expression, protein-protein interaction, RNAi, secreted molecules, ion flux, chromosomal imprinting, etc.; Metabolic / signalling pathways / elements of cell cycle; Whole cell modelling; Future potential: Multi-cell - Physiological systems; Entire multicellular organism	Gene Expression; Biochemical and Physiological pathways; Pathology; Animal Models; Human genetic variation.	Gene expression on various diseases and model organisms, protein interaction, signalling pathways, metabolic pathways	The vision of the programme is bringing together scientists from different disciplines. The long term objective of their work will be the in silico model of a human cell. To achieve this goal decoding of metabolic and regulatory networks will be necessary to model and later simulate physiological pathways.	Gene expression, protein-protein interaction, metabolic and signalling pathways, subcellular organisation, morphogenic changes
? Criteria for choice	**How did you choose your research area?**	Interest in dynamics of the control of cell proliferation.	Based on economic and scientific criteria. Scientific: The asymetrical cell divsions of ES cells and their	Study the human condition whenever possible. Integrate what is known	The development and introduction of modelling tools is an essential step for understanding gene	In early 2001 the Biological Research and Technologies Division of the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF)	Complexity and importance for biomedical research, feasibility of genetic alterations

			derivatives offer an ideally structured system of study, as described above. Economic: Huge need in industrial R & D for innovative target, mode of action and side effect/toxicology screening on the level of expressed proteins, in particular in human systems.	from other species into that template.	function, cellular processes and pathogenesis	initiated a forward-looking process under the name of BioFora21, with a view to developing new funding strategies to meet the future requirements of modern biosciences. Renowned experts from scientific institutions and from industry provided expertise. In addition the results of the most important international systems biology forum – the Second International Conference on Systems Biology (ICSB 2001) were included in the assessment. In December 2001 as the result, the BMBF published the new Systems of Life – Systems Biology research programme.	
? How important is it?	**What are the main contributions expected? Basic research advancement (see fundamental genomics work programme for context); Implications for applications, e.g. health**	Cell cycle, being the series of events leading to cell duplication, being seen as a central topic to understand cellular function in both molecular and systems terms. In the longer run it is expected that the cell cycle model, appropriately wired to metabolic and signal transduction input/outputs can act as a skeleton to develop a first "replicating" yeast e-cell.	The systems biology approaches have the potential to eventually resolve urgent health-related issues, whose enormous complexity became only apparent after genomic/RNA data became available. (new drug development, drug safety, but also non-animal based general screening for e.g. toxicity profiles of chemicals, etc.)	Standards that enable data integration. The public domain has taken responsibility for the representation of genomic sequence, gene nomenclature and to a lesser extent the function of gene products.	Develop a modelling system that is able to integrate experimental data and that meets the complexity of biological systems. Rules and methods for data integration that connect the existing vast amount of information on the biological objects. Development of "learning" methods and testable hypotheses in order to gain new knowledge. Co-operation with clinical and experimental partners on the analysis of human diseases.	The ultimate goal of systems biology is to develop virtual representations of cells and entire organisms. These representations will then enable computer experiments to be conducted in analogy to experiments involving real biological systems. This can pave the way towards predictive biology which – among other things - will help to understand and treat diseases in man.	Better understanding of the dynamic and gradual operation of signalling pathways that have crucial importance for human health and treatment of many diseases. Generating concepts that could be used eventually in personalised treatment.
? Do we have enough data and understanding to solve the problem?	**Discuss the level of data input available: Genome, proteins, pp interaction, expression data?**	Many data available in terms of genetics, biochemistry and more recently of global omics data. Nevertheless many of the data, as good as they are, are not tailored for systems biology approach (different biological systems, qualitative and not quantitative, noisy ...). Comparison of differentierarchical levels of data (transcriptional vs protein vs protein modification/interaction, vs metabolic marker) completely lacking in current data sets.	Correlated data sets will be generated during the project, which in this type of interface density do not currently exist, which is highly advantageous. Rigorous adherance to data formats will be mandatory, and participants will be schooled to the standards required. Data types: Gene expression, proteomics (in particular posttranslational protein modifications, localisation and turnover rates), gene methylation, ion fluxes, and other parameters to be decided during project preparation. Use of data from other projects will be considered.	Any problems we have to solve will only become tractable if we are able to organise the many pieces of information at our disposal. Breakthoughs normally occur when old problems are adressed with new insights/technologies.	Two main problems: Firstly, data is extremely heterogeneous, i.e. produced on different levels of cellular information, produced with different biotechnical methods, in different labs, etc. Thus, data is often poorly correlated. A fundamental need therefore is the development of data integration methods. Secondly, data is (in most cases) not well-designed to solve the problem. There is a fundamental need to understand that data production and data analysis are two elements that have interactions and feed-back. Methods of experimental design must be introduced at the initial phase of the experiments.	All data types mentioned are available or will be genrated in the course of the programme. They will contribute to accieve th e golas of the funding initiative System of Life - Systems Biology	Data only available for input and output but only little for intermediate steps and barely quantitative data or data at single cell/single pathway level. This is addressed in the project.
?Is the problem tractable computationally?	**Is "traditional" bioinformatics enough? How far can we get with an analytic or traditional hypothesis approach? Why are CSB tools are appropriate?**	In a word, NO. What is needed is an engineering-like approach able to tackle complexity in terms of circuits and their regulation, identifying the core common elements and their behaviour.	Cellular processes are highly stochastic. We are unaware of currently existing bioinformatic processes that are sufficient.	Only in terms of the speed at which it can be queried if it is organised correctly. Most problems still require experimental solutions. However elegant solutions can usually be extrapolated to other experimental problems with the aid of computers.	What is traditional bioinformatics ? Most bioinformatics tools are explorative and non-predictive, that is the main limitation.	Certainly, traditional bioinformatics and modelling expertise have to be developed further. Some experts even discuss the necessity for a new mathematic to solve the upcoming problems. This would pose additional challenges for "traditional" bioinformatics.	Requires mathematical modelling, probably with different approaches, to understand pathway operation and to assess the rules of pathway function (feedback loops, robustness etc, which are important for pharmacological intervention). Theory reasonably developed for deterministic view of cell as containing a well-mixed large numbers of participating molecules; less well developed for stochastic effects due to few participating molecules or uneven intracellular distribution.
? Goals of Research (explain, predict control)	**How far can your project go in terms of experimental method?**	Yeast and cell cycle may be good models because (i) yeast is enormously developed for genetics, biochemistry and "omics"	Signal transduction events, proliferation, drug intervention, etc. during proliferation and at defined endpoints (e.g. neurons) can be exactly controlled by functionally correlated measurements (calcium and pH imaging, motility, extracellular small molecule release, morphology, etc.). Horizontal integration is achieved	The pharmaceutic industry suffers from a massive attrition rate with >80% of targets failing to produce an effective therapeutic. This	The tools that we develop will have an impact on experimental design of experiments. The ultimate goal is the definition and refinement of hypotheses related		Available experimental data are not sufficient to assess dynamic operation and the importance of gradual/small changes. Available experimental data

		and cell cycle may be the good size for tackling very complex interlaced circuits befor going into the extra complexity of the e-cell.	by cross checking results in mouse and human stem cells.The appropriate tuning of these systems provides an ideal basis for subsequent molecular analysis involving differential and quantitative genomics, tanscriptomics and proteomics technologies.The project will employ all experimental methods necessary, involving groups chosen on the basis of their outstanding excellence.	provides plenty of scope for improvement- it's starting to be explained and predicted but we still have to see whether it can be controlled	to human diseases. The feed-back loops in our platform (experimental data - network reconstruction - forward modelling and simulation) will help researchers to perform guided experiments (knock-outs , RNAi) and have a direct impact on drug target validation.		may provide yes/no answer but little insight into quantitative, time-dependent operation.
? Explain	**Will we know more than when we stared modelling? Do we just want to organise data? Is it a self-consistency check and completeness check on data? Can we gain understanding?**	The major task of Systems biology is going beyond simple interaction schemes by adding quantitative/temporal terms, i.e. dynamics. This should allow to make "not-so-obvious" predictions.	If the project goals are met we will gain understanding. In the field of Systems Biology "understanding" is not something that can be 100% achieved, but will be much more a continuous process over the next 10-30 years that requires a commitment of resources with that time perspective. Processes leading to sequential asymmetrical cell divisions will be explained by the study. Processes involved in toxicology will be explained. Characterstics responsible for stem cell potential or differentiated cell-types will be explained.	About 50% of drugs fail because they have unforseen side effects.We are not short of experience in this area- just data. The remainder because they don't work as well as the market demands. One solution to this would be to better understand hetergeneity that underlies disease and the individuals response to medicines.	The proper organisation of data is the first step. The next steps will be to construct network models and to draw time-dependent simulations and perturbations of these networks and check if these can be verified with experimental data. In the light of these simulations we will gain new insights in the biological systems.	We will definitely gain new knowledge with the SB funding initiative. Goal is to increase the knowledge base in the fields hepatocyte cell culture, modelling and simulation, and to develop methods and tools to lay the foundation for an internationally connected national competence network in systems biology.	Gain understanding of role of feed-back loops, robustness, bistability etc, that is not possible with experimental data alone. Assess relevance of different stimuli/interaction pattern on signaling process.
? Predict	**If we e.g. change a gene, can we predict new expression data? In perturbed system can we predict perturbation and control experiment? What is involved in the transition from "healthy" to "diseased"?**	Within a systems biology approach the change from the healthy to the diseased staus does not necessarily simply involve disease gene products, but altered networks.	A model based on differention of ES cells from one species (e.g. mouse) can be used to predict events in ES cells from another species (e.g. human) or in an animal model (mouse experiments & clinical human cases). Differences between the response of the models provides a highly relevant additional level of "horizontal" integration of data obtained from different system. Moreover ES cell systems are perfectly suited for genetic analysis and manipulation (e.g. RNAi)	We will have the information to enable the study of individual human variation in many therapeutic senarios. Our ability to interpret any of this information will be limited by our lack of knowledge on how biochemical and physiologic functions relate to genes and their expression. We can't easily study the former but have information on the latter on a huge scale.	That is the challenge and it is totally dependent on good interactions of experimental and theoretical people.	The questions concerning the predictability of certain events in "cellular hepatic metabolism" are also topics within the research projects of platforms and collaborative projects of the funding initiative Systems for Life - Systems Biology.	Prediction and subsequent experimental confirmation is an important element of the project.
? Control	**In "diseased" or "ill" system, can we bring it under control with medicines, molecules, etc? Can we extend control of model system disease to human disease? Can we develop platforms for testing pharma medical concepts?**	A different way of thinking that may ultimately lead not only to more efficient identification of drug targets, but to different therapeutics approaches alltogether.	In the project exsisting and new platforms are integrated and interfaced to investigate physiological or pharmacological concepts in a comprehensive way with much higher relevance than previous "HTS" concepts. The generation of a relevant context on the level of posttranslationally modified protein isoforms will be the basis for new, second generation, HTS methods, which currently do not exsist.	Still gathering information. There are some promising case studies but we still have little idea of what significance looks like. The old story of nature versus nurture will always limit approaches based entirely on genetics. Similarly many wonderful drugs treat symptoms not disease.	The conservation of systems behaviour between species (e.g. human - mouse) is a critical challenge, in particular in the drug target validation process. But also the genetic variants of a drug target in different individuals. These are rather mid-term issues (5-10 years). They must be considered but they are not the primary goals of the current project.	The model system human hepatocyte aims at application oriented research in pharmaceutical industry. The long-term goals of the federal funding initiative Systems Biology are in silico model systems of hepatocytes to test pharmaceuticals and, more specific, model systems of hepatic differentiation and dedifferentiation as well as detoxification systems.	That is a long term goal and the tools from the project could be used for that.
? Choose right software tools for problems complexity level	**What modeling tools are you using, and how generic are they? Static bioinformatics; Static systems approach; Boolean switching; Chaotic systems; Time dependent, (differential equation); Time and space dependent; Time and space dependent plus multiple interacting systems (full cell model); Multi-cellular; Multiple physiological systems**	Dynamic modelling and custom made algorithms.	In general terms: Chaotic systems; Time dependent, (differential equation); Time and space dependent; Time and space dependent plus multiple interacting systems (full cell model). Specifically, there are a number of tools available that address at least parts of the problem (SBW comes to the mind). Long term possibilities: Multi-cellular; Multiple physiological systems, organism.	Simple analysis of structured data has the most business impact. The ability to deliver large amounts of genome related data in a comprehensible way to human intelligence pays with so many staff. More sophistical methods applied to defined and smaller scale	Explorative multivariate statistics (concept of co-regulation); gene regulatory networks; kinetic modelling (ODEs, Boolean); Learning methods (e.g. Bayesian statistics);	The model tools so far planned to be used in the platform 'bioinformatics' and in collaborative projects are mostly developed in specific laboratories and cover aspects of static, time- and space-dependent, and multiple interacting systems. New developments will be required to cope with the increasing complexity of the knowledge base over the time period of the projects.	Mainly differential equations, steady states and time simulations; Reaction-diffusion systems envisaged; Stochastic simulations for low number systems; Boolean and Petri-Nets where applicable; Own modeling and simulation environment under development
? For multiple interacting systems, e.g. (a cell, choose and design hierarchy of tools and interfaces	**How do you link the various models together? Boundary and interface conditions; Available data and its form; Different computations in different regions; Nature of "numerical experiment"**		Precise plans will be finalised during project planning. We would presently start from the software that is already there (eg., Systems Biology Workbench, SBML). As the work progresses, additional tools and standards will be developed as they become necessary.	Federated databases linked from common points- gene and genome based standards are employed. Industry standards employed for business practice e.g. Phase I, II, III	Exchange of models via SBML; automated population within our system is possible for example through KEGG database; planned activity is to integrate SRS system; we will develop a library of kinetic models and	At the beginning a modular approach to model hepatocytes will be followed. To be able to link thoee modules, standard data formats and procedures have yet to be defined. This will be one of the major tasks in the initial phase of the projects.	Its is intended to formulate one model in one mode of description (depending on the specific question to be answered). Interface conditions remain a problem to be tackled on the way to whole-cell

appropriate to problem, data, and desired solution					standardisation rules		modeling. Numerical experiments: in the case of ODE-models: simulation over time, check of steady states or critical dynamic behavior
? Database requirements	**Do your databases and software satisfy the following requirements? Standards in primary and/or secondary databases; Standard and compatible input protocols for programmes; Is there enough complete data to do the computation?; Experimental verification of conclusions.**	Data bases need to be cross-linked and better annotated. But many databases and experimental sets are devoted to a single cell element (RNA, protein etc.) , thus making more difficult to make correlations between data sets.	We follow standards as far as possible. However, current databases are not always properly cross-linked or annotated. Interfaces between databases and between the software employed need to be developed (XML/SBML-based, rather than the current specialized or even proprietary data formats).	Have to very robust- tend towards the conservative	This is an essential part of our project.	As mentioned above, the standardisation procedures have yet to be defined on all levels of the general data base. Complete data sets for computational modeling are not avaiable for hepatocytes at the moment. Experimental verification of modelling and simulaton results is probably the most difficult part in future systems biology research. For many questions experimental tools for quantitative description of the system components have still to be developed.	Data are quite sparse, but data and qualitative knowledge together allow for development of models, which can be improved iteratively with experimental verification/falsification of model predictions.
? Resources	**What do you need, and what is available, in the areas of hardware and human resources?**	People with different and complementary expertise (biochemists, molecular biologists, bioinformatics, systems engineers, etc.). Software and visualization tools to help explore concepts. More rigorous representation of biological networks and (semi)automatic generation of the simulation data set ready to be tested.	It is to be expected that the amount of data produced and the amount of computing power needed ultimately exceeds what is available. Therefore, data storage and computing facilities are needed. In terms of human resources, we need more people understanding both systems biology and programming - Education plays a central role here.	The trick is to engage the human resources that browse the data. I've never, realistically, been limited by resources. Most limitations are imposed through the benefits gained through consensus over such a broad organisation.	Hardware is not the problem. Probably, parallelising computation might be an issue in the next years. Data storage is sometimes expensive and this might cause problems. Literature resources should be provided in a computer-readable way that allow the extraction of information by automated methods. Training of young scientists for systems biology should be intensified by international training activities at universities and research institutes.	A central data storage and maintenance institution should be established that should possess scalable computational computing capacity. For the future, educational measures have to be developed to ensure qualified scientific personnel for systems biology research.	Available: Adequate computing hardware. Human resource, people interested and trained in biology, mathematics, and computer sciences. Needed: people understanding both experimental and theoretical work.

****

Computational Systems Biology Workshop		SESSION - 3	SESSION - 2	SESSION - 2	SESSION - 3	SESSION - 2
	NAME OF PRESENTER	Nicolas Le Novère (http://www.dopanet.org/)	Uwe Sauer	Age Smilde	Hans V. Westerhoff (www.systembiology.net)	Olaf Wolkenhauer (http://www.sbi.uni-rostock.de)
	TITLE OF PROJECT OR TOPIC	Systems Biology of the neuron	Systems-oriented metabolic research		Systems Biology from molecules to living cell; also: organizer of European Systems Biology for Genomics and Health (EoI)	
	SHORT DESCRIPTION -------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- --------	Realistic modeling of networks of signal transduction at the levels of the synapse, the neuron, and the micro-circuit. Those models will provide a deeper understanding of neuronal signaling, but also will aim at reproducing neuronal disorders in silico to understand how specific molecular and cellular abnormalities could generate pathological phenotypes, and to predict the effect of pharmacological treatments.	Computer modelling and quantitative experimental analysis of system-wide behavior, regulation, and dynamics in central carbon and nitrogen metabolism. What is the global regulatory control structure of metabolism; i. e. what is the relative importance of different regulatory mechanisms for the overal system performance?	TNO Systems Biology in Health, Nutrition and Microbial Systems. An integrative approach combining transcriptomics, proteomics and metabolomics. Computational tools to support this should be developed.	Day-to-day integration of computational, theoretical and experimental approaches to System Biology. Aim is to understand the funtionally important and exciting properties that occur in Systems of Macromolecules and not in those macromolecules in isolation. Existing methods that highlight the mechanism of emergence of system properties such as Metabolic Control Analysis are developed so as to include signal transduction, regulated gene expression, intercellular and interorganism communication. This has led to Hierarchical Control Analysis and Regulation Analysis. Experimentally we emphasize (i) 'vertical genomics' which measures mRNA, protein, enzyme activity, metabolites and fluxes for one functional property simultaneously, rather than measuring all mRNAs, or all proteins (whihc corresponds to the usual 'horizontal genomics'), experimentation in and with livinig cells under phyiological conditions, and experimental approaches that lead to results that can be used directly in computaitonal approaches (in	Mathematical modelling of the dynamics of gene expression and signal transduction.
MAIN TOPIC	**AREA TO DISCUSS**		**ANSWER FOR SPECIFIC PROJECT**			**ANSWER FOR SPECIFIC PROJECT**
? Choose the Research Area	**What is the main research area?**	Neurobiology	Central carbon and nitrogen metabolism in bacteria and yeast	System Biology in Health, Nutrition and Microbial systems	Molecular cell biology for experimental systems that are best for Systems Biology	Systems Biology
? Level of complexity	**How complex is the biological system? Gene expression, protein-protein interaction, etc.; Metabolic / signalling pathways / elements of cell cycle; Whole cell modelling; Multi-cell - Physiological systems; Entire multicellular organism; OTHER**	Gene Expression, Molecular interactions, Regulator and Signalling Pathways, Cell interactions	Gene expression, protein interaction, signalling pathways, metabolic, multi-cell-physiological systems	Gene expression, gene-protein-metabolite interactions. Biomarkers for diseases. Interaction between nutrition and health at the systems level. Exploring and understanding biorhythms. Exploiting gene-protein-metaboloite interactions for microbial processes.	Autonomous unicellular systems that are experimentally well accessible to SB: e.g. E. coli, S. cerevisiae, tumor cell lines. Analysis from macromolecules to cell function, including metabolic pathways, signal transduction, gene expression., spatial organization and the integration of these to produce function.	Gene expression, signalling pathways
? Criteria for choice	**How did you choose your research area?**	The experimental knowledge of neuronal signaling renders the usual models of neuronal function completely obsolete. Time has come to develop models that incorporate interactions between various neurotransmitter pathways, protein motions within membranes, and their interactions with cytoplasmic components	I felt it is important and suitable methods could be developed	Focus areas of TNO Systems Biology: Health, Nutrition and Microbials systems	Criteria: Fully accessible for genomics; much kinetic information already available, accessible for molecular genetic and metabolic manipulation; interesting emergent properties expected	The wonderful complexity of nature - because ist dynamics that cause biological function, regulation and control.
? How important is it?	**What are the main contributions expected? Basic research advancement (see fundamental genomics work programme for context); Implications for applications, e.g. health**	Basic science: Mechanisms of synaptic plasticity; Interactions between various neurotransmitters; Adaptative neuronal networks; Overlapping networks; learning and memory. Applications: Understanding of the basis of metabolic (e.g. Parkinson) and architectonic (e.g. Schizophrenia) defects, in silico Testing of drugs in realistic situations, Ground for new archictectures of chips (with or without living components)	Clearly fundamental research advancement at the core. Our system-wide, quantitative data help shape faithful models and understanding. There is a strong applied biotech component in the sense that being able to understand how metabolism functions as an intact system helps to engineer it at will. Beyond biotech, all pharmacological and medical applications with a strong metabolic component such as liver cells will benefit from both such data and models.	Health and nutrition, e.g. obesity and diabetes: early diagnostics and prevention. Satiety issues, toxicogenomics. Microbial systems can provide means for producing medicine and other important compounds	1. Proof of principle of Systems Biology; 2. Generation of first sizable Silicon Cell (computer replica of substantial part of a living cell); 3. Discovery of new scientific laws (Systems Biology principles); 4. Application of the discovered principles to multifactorial diseases such as cancer and type II diabetes; 5. Making partial Silicon Cells for these multifactorial diseases; 6. Application to, and Silicon cells for, cell factory biotechnology [yeast, L. lactis]; 7. Network based drug design; elaborated for sleeping sickness; 8. Increasing the usefulness of molecular biology and genomics for society by making the developed knowledge relevant for entire biological systems	Helping an understanding of intra- & inter-cellular dynamics. In the short term our models help experimental design, create and validate hypotheses. In the long term it provides a framework in which to understand the dynamic processes within the cell - the implications are fundemental to an understanding about what can go wrong in case of disease.
? Do we have	**Discuss the level of data input**	We don't and this is a huge problem. Systems	There are generally a lot but dispersed data available, and	No. Discuss what is needed and	Indeed genomics, but different from what is common practice now.	There are not enough data! We are

enough data and understanding to solve the problem?	**available: Genome, proteins, pp interaction, expression data?**	Biology of neuronal cells is impaired because we do not know where exactly are the proteins, in which amount, how they interact etc. A strong emphasis should be put on the funding of large scale data mining in model systems. Follow the example of the Alliance For Cellular Signalling.	we lack accurate quantitative data from one experimental system: strain, cultivation, …… Most of the presented data are at the compositional level, e. g. transcriptome and proteome. Thus, we lack additionally quantitative metabolic data on metabolite levels and in vivo reaction rates for a global understanding of metabolic control.	how to obtain it. What is necessary to build models, find biomarkers etc?	It is not so useful to determine everything at one genomic level (e.g. all mRNAs in a cell); it is much more useful to determine promoter activity, mRNA, protein, enzyme activity, metabolites, flux etc corresponding to one function; I call this vertical genomics). Interaction data are highly important, but above all data need to be more quantitative and precise and reliable. Much more kinetic data are needed. And data need to be generated that are necessary for the calculations; CSB driven experimental research is necessary.	lacking accurate quantitative measurements of protein concentrations measured for a large number of proteins (say 20) over a reasonable period of time (with sufficient time points). Ideally I would like to have the same experiment on at least two levels - transcriptome and proteome.
?Is the problem tractable computationally?	**Is "traditional" bioinformatics enough? How far can we get with an analytic or traditional hypothesis approach? Why are CSB tools are appropriate?**	Yes, with a (reasonably large) bit of work. We need more plastic software platforms, able to accomodate for various kind of data, and various levels of accuracy. We also need modular tools. For instance we should be able to run models using at the same time, logic, stochastic and deterministic algorithms.	Bioinformatics is not very helpful beyond 'cleaning' raw data. Traditional hypothesis-driven research will always be necessary but it would take for ever if a more complete understanding of complex systems is attempted in such a patchwork aproach. It should be understood that systems biology is hypothesis-driven research, only that hypotheses are model-based and generated in a systematic fashion. The problem lends itself naturally to computational/experimental collaborations. Comprehensive and structured hypotheses in such complex systems with large and multi-dimensional data sets can only be generated from computer models.	Developing computational systems biology tools is vital for progress in systems biology. An abundance of data is collected on a system, which should be intergrated in a systems model. This model can then be analyzed for its properties. This is more than bioinformatics: we have to think on a systems level and find the appropriate methodology to handle the data.	Traditional bioinformatics is one of the necessary components, but by far not enough. At least four new aspects are needed: 1. full appreciation of the complexity and nonlinearity of biological systems in the processing of experimental data to understanding, 2. full integration of information from all sources (databases, experimental results, physical chemical concepts and parameters, theory), 3. Integration into 'live' and precise mathematical models of reality (Silicon cells), 4. A dynamic analysis, computation, prediction, validation cycle.	Traditional bioinformatics doesn't provide much help - there is not so much a need for tools but for methodologies - clever ways to identify models from the data we have. Formal methods for the analysis are necessary to compensate for a lack of data.
? Goals of Research (explain, predict control)	**How far can your project go in terms of experimental method?**	My forthcoming group will develop a software platform to realistically model part of or entire neurons. We will use this platform to perform hypothesis-driven research. For instance, the detailed modeling of a synapse incorporating receptor movements should tell us if the horizontal (between synapes) and vertical (with intra-cellular membranous compartments) transfers affect the efficacy and the plasticity of synapses. Another example lies in the role of the dendritic spine: Is it primary a biochemical or an electrical compartment.	My group works on both the experimental and the computational aspects of metabolism. The major contribution being parallel quantification of in vivo carbon fluxes based on 13C-labeling experiments. These methods need to tailored for dynamic analyses and for higher cell types. Albeit an important aspect of metabolic research, flux data must be combined with complementary data on gene expression and regulation. In priciple, the data are there or could be generated easily, but computer models are the only realistic way to generate quantitative hypotheses for further experimental verification/falsification.	Developing biomarkers, understanding the impact of nutrition, understanding the relation between nutrition and health (e.g. diabetes!). Understanding microbial systems and ways to optimize those.	All the way. We have in pricnple all that seems needed and possible. What we need though is intensive interactions with other SB groups and amplification of the work force and common focus on subtopics.	Our models help the biologist with his understanding of nonlinear interactions in gene expression and signal transduction. It helps him to design his experiments, generates and validates hypotheses.
? Explain	**Will we know more than when we stared modelling? Do we just want to organise data? Is it a self-consistency check and completeness check on data? Can we gain understanding?**	Such a modeling fills a gap which cannot be tackle by experimental procedure. By its sheer nature, a biochemical experiment is (1) reductionist and (2) affects the system during the measurement (generally destroy it). The Systems Biology modeling permits to overcome that and studies emerging properties, non-linear behaviour and so on.	Systems biology is about understanding and has nothing to do with data organization - this is essentially bioinformatics. But understanding will not come automatically, since no single lab combines all required expertise. Thus, it will be necessary to create an appropriate environment for collaborations. If the research environment and the collaborations are set up in the right way, quantitative understanding of non-linear, dynamic interactions in complex metabolic systems and new causal relathionships will definetly result.	Making models of biological systems, including nonlinear dynamics, feedback control etc. Try to understand the system using this model	Much more than 50 % of the properties and laws of living organisms is in the systems biology. I.e. derives from the organization of the molecules rather than r3esidng in the iindividual molecules. SB will discover more general principles of biology than Molecular Biology has done. Because SB is closer to function, the discoveries will also be more directly applicable.	We are not just organising data! It is dynamics that creates biological function and control. These dynamic interactions are mostly nonlinear, i.e., counter intuitive.
? Predict	**If we e.g. change a gene, can we predict new expression data? In perturbed system can we predict perturbation and control experiment? What is involved in the transition from "healthy" to "diseased"?**	We can indeed predict cellular behaviours, and test the predictions, for instance via electrical recordings	Absolutely! Predictions are the goal and are already possible. The quality though depends on the models and on the data, which at this point are often not quantitative (enough). The iterative process of prediction (hypothesis generation), experimental analysis, and model refinement is the heart of systems biology. Simulation help identify and design critical experiments experiments, rather than running 'brain-less' and expensive discovery-driven projects.	E.g. predict whether an obese person is likely to develop diabetes.	We cannot do this now. We shall be able to do this once we have determined more paratnmeters quantiatively and once we have developed improved CSB approaches (fit to deal withg the nonlinearities of the problem). I belive this because we have been able to do so on a small scale for model cases of systems Biology (network based drug design in T. brucei; glycolysis in yeast, ammonia assimilation in E. coli). Transition between diseased and healthy is qite a realistic option; we have an example project on type II diabetes.	We can make predictions - but not quantitative (we don' have the data). The concept of a systematic pertubation is central to the systems approach: only with systematic pertubation studies we can identify causal entailment. This systems approach requires more expensive experiments but simulation can help designing these experiments.
? Control	**In "diseased" or "ill" system, can we bring it under control with medicines, molecules, etc? Can we extend control of model system disease to human disease? Can we develop platforms for testing pharma medical concepts?**	This is one of the main goal. This is why the DopaNet initiative chose its model system, responsible for Parkinson disease, Huntington chorea, Schizophrenia and drug addiction.	This is certainly one goal and can be done. One has to walk before running though, and model systems are invaluable to address complicated key questions. In some areas still tools need to be developed and this is best be done in a system one understands well. At a realistic level in European collaborations, a healthy mixture of fundamental model systems and platforms for testing pharma concepts appears to be feasible.	E.G. predict the diet that prevents the obese person to develop diabetes.	Yes. One should not be overambitious, but I see three lines that should already begin to lead somewhere: (1) control analysis of sickly cells [EU supported Eastern Europe program): we should be able to unbderstand the effect of low levels of heavy metal pollutants on cell function (2) Silicon cell models of parasites already suggest ays to manage them, (3) silicon cell models of humans although incomplete shold already help to suggest new clinical trials	Hope so - but I would not expect this to happen soon!
? Choose right software tools for problems	**What modeling tools are you using, and how generic are they? Static bioinformatics;**	Every single piece of software judged suitable will be used. But it is anticipated that a new large, carefuly designed, software platform will be	Differential equations, static systems approach, principal component analysis of mass data (Matlab). There is clearly a need to define common standards that allow to	(Nonlinear-) Dynamic models, systems identification., principal component analysis, multiway	Metabolic and hierachical control analysis, differential equations, time and space dependent; multiple interacting equaitons ; silicon cell type models; stability analysis, Fourier analysis, non	Dynamic modelling - time series analysis. Tools are Matlab, Mathematica, C++

complexity level	**Static systems approach; Boolean switching; Chaotic systems; Time dependent, (differential equation); Time and space dependent; Time and space dependent plus multiple interacting systems (full cell model); Multi-cellular; Multiple physiological systems**	necessary. This simulation environment shall use all the existing algorithms (e.g. PDE for calcium diffusion, ODE for signalling pathways, stochastic approach for receptor behaviours, logic rules for gene expression etc.)	exchange models.	analysis, preprocessing, multiset analysis, discriminant analysis, integration of data driven models and differential equations (e.g grey models).	equilibroium thermodynamics (-'Mosaic'), Principal Components Analysis, Monte Carlo modelling	
? For multiple interacting systems, e.g. (a cell, choose and design hierarchy of tools and interfaces appropriate to problem, data, and desired solution	**How do you link the various models together? Boundary and interface conditions; Available data and its form; Different computations in different regions; Nature of "numerical experiment"**	Markup languages (SBML, NeuroML); Ontologies (cf http://www.dopanet.org/ontology.html)		Systems Biology Informatics platform. Integrating data bases and bioinformatics/data analysis software tools. Systems Biology models.	We use and develop nonlienar modularity strategies; define modless in new ways with (nonlinear) boundary conditions. Our Silicon cells (www/siliconcell.net) are web based and fully usable through the web; SBML is our intended interface standard.	Systems Biology Workbench, SBML
? Database requirements	**Do your databases and software satisfy the following requirements? Standards in primary and/or secondary databases; Standard and compatible input protocols for programmes; Is there enough complete data to do the computation?; Experimental verification of conclusions.**	Interfaces to existing databases (BIND, MINT; DOCQS, KEGG, Genome Knowledge Base) should be developed. However, it is anticipated that customised databases, of higher accuracies will be developped (and we began the process actually).	Our own quantitative physiology and flux database is pretty much for in house use with a generalized statistical quality check. Again standardisation would be desirable but requires an appropriate organization.	We collect an abundance of data These have to be stored an accessed. Links to metabolic pathway maps and genomics data bases.	Not yet; work and activity is needed here; this event connects to defining standard experimental strategies and sample bases!	The main challenge is with methodologies - the mathematics to deal with the data we get. Databases have little to do with this.
? Resources	**What do you need, and what is available, in the areas of hardware and human resources?**	We need large funding initiatives so that experimentalists begin to work together to share protocols and standards, and also to produce accurate data on a large scale. We also need grants to pay PROFESSIONAL software engineers to develop the adequate piece of softwares, rather than tiny grants to pay post-docs who spend most of their time learning basic computational techniques.	Basically, people are there and so are computer power and many of the methods. The real problem is a funding environment that brings the right people together, otherwise systems biology will not fly in Europe. There is of course a funding need for expensive quantitative experimental analyses, but also for people that take care of standardisation (experimental and computational), databases, and transfer, for which it is diffcult to obtain money otherwise.	Collaborations, ideas, tools, software,…..	People willing to go for SB; enthusiastic about collaborating with others, experimentalists and theoreticians alike; perhaps even more so people that do both expiriments and computations. Hardware: some supercomputing for the silicon cell; interfaces for computer driven/robotic experimentation. Samples bases. Mind you paradoxically perhaps: what is needed much even for purely Computational SB is good experimental SB!!!	People with good ideas! Mathematicians, physicists, engineers - more than computer scientists. No special computer power is required.Our collaborators need the expensive kit to generate the data.