White Paper

Yeast Systems Biology Network

Stefan Hohmann, Jens Nielsen, Hiroaki Kitano

(see for further contributers at end of text)

Göteborg, Lyngby, Tokyo, May 2004

Executive Summary

Systems Biology is an emerging field expected to have major impact on the future of biological and medical research. It aims at system-level understanding of biological processes employing mathematical analysis and computational tools to integrate the information content obtained in experimental biology. The Yeast Systems Biology Network proposes the yeast *Saccharomyces cerevisiae* as a model system to develop the field and to advance our understanding of the rules and principles of the dynamic operation of cellular systems.

The Network will serve the integration of activities in yeast systems biology. It is proposed to be based on a web resource, which will allow the research community to contribute data as well as mathematical models and to simulate biological processes. The community will then through further experimentation contribute to the recursive refinements of the models to advance our understanding of the biological process. The vision is to generate reference models of budding yeast, eventually at the level of the whole cell.

The Network will facilitate the cooperation between experimental yeast researchers and theoreticians. The Network will stimulate the generation of quantitative data generated in time courses and containing information on the spatial distribution of the individual components of the yeast cell. Development of relevant experimental as well computational tools will be an important goal of the Network.

The Network will have highest scientific standards, excellence and visibility.

Funding will be required for the establishment and maintenance of the web resource, the underlying infrastructure, development of experimental tools and theoretical approaches as well as training. The Alliance will make yeast systems biology visible and support raising of funds for the research area at the national and regional level.

Systems Biology

For the purpose of this White Paper and the Yeast Systems Biology Network we define “Systems Biology” as having the following characteristics:

- Multidisciplinary research with a close interaction between experimental research (biology, chemistry, physics…) and mathematical modeling and analysis (mathematics, bioinformatics, engineering…).
- Generation and iterative improvement of mathematical models, which describe realistically cellular networks or dynamic biological processes. Models have the ability to elucidate systems properties and to predict the outcome of perturbations.
- Where possible use of existing and generation of novel data, in particular quantitative data, time courses and spatial information of high definition.

Objectives

21st century biology will gradually move from studying individual cellular components to addressing how individual modules interact to make the cellular system function as a whole. Thus, there will be a trend towards a *systems* view of cellular function based on an understanding of the molecular mechanisms of individual processes.

In order to generate accurate descriptions of the interaction and operation of different components and to understand at a quantitative level the relationship among genotype and phenotype it will be necessary to develop a research environment in which experimental and theoretical scientists work together. Hence, we will see an integration of experimental research with efforts to generate and optimize computer models of cellular networks and processes. This integration represents a real challenge for researchers in the future. In particular, in order to

describe the *dynamic operation* of biological systems, it requires the generation of quantitative data, time courses and spatial information. In fact, generating the data to answer questions such as “Which proportion of protein X is located in compartment Y at time Z” or “How many molecules of metabolite A are present in compartment B” constitutes a true experimental challenge. Therefore, Systems Biology will encompass the development of tools and approaches to generate such data. This type of information will help to better understand diseases and hence Systems Biology will become an integral part of drug target identification and drug design.

In the development towards quantitative biology – or Systems Biology – the study of suitable model systems will be pivotal. The yeast *Saccharomyces cerevisiae* represents an obvious model system for a concentrated research effort in this area to significantly advance biological sciences, just as it has been for genomics and functional genomics.

For these reasons the Yeast Systems Biology Network phrases its objectives as follows:

> *To generate the environment for a close interaction of yeast researchers and mathematicians/computer scientists (and other disciplines such as bioinformatics, engineering, physics and chemistry) interested in developing* ***Systems Biology of yeast.***
>
> *To advance strategies for model generation and to phrase protocols that allow different models generated in the Alliance to be compared and eventually connected, with a vision to progressively build reference models up to the level of the whole cell.*
>
> *To make available existing experimental data and use those for generating cellular network models as well as mathematical descriptions of defined dynamic cellular modules/processes, such as signaling, metabolism and the cell cycle as starting points.*
>
> *To generate and/or advertise tools to produce quantitative data of cellular processes and to present the formats in which the research community can contribute novel data to test and iteratively improve such models.*
>
> *To build and maintain, in collaboration with other resources such as SGD, integrated databases of quantitative data of different types, that will serve as a major source of information for model building.*
>
> *To generate the computational environment that allows researchers (and the interested public) to run and test animated models on their computer screen.*
>
> *To increase the visibility of the research area and thereby help attracting financial support for yeast systems biology.*
>
> *To establish an infrastructure at the regional (European, American, Asian…) and global level for training of students and researchers in Quantitative and Systems Biology.*
>
> *The raise the awareness for Systems Biology in the general public and at the level of school education.*

The integrating platform for the Alliance will be a web-based functionality that will enable evaluation of the role and operation of individual components in yeast cells, and thereby use *in silico* techniques for experimental design. In this sense the web resource will help generating a global virtual research community.

The vision of the Network as part of the yeast research community is to work towards a comprehensive understanding of the function of the yeast cell, which will serve as a paradigm for all eukaryotic cells.

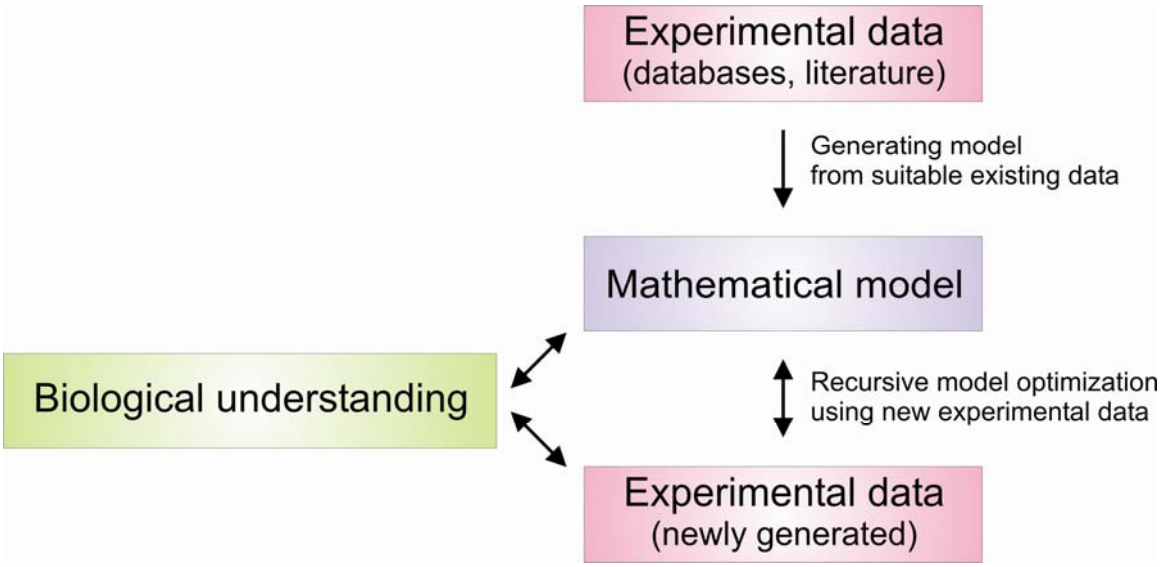
Importance

Why is Systems Biology important and why do we need a network to advance the field?

Functional genomics, i.e. global approaches on gene and protein expression, metabolomics as well as interaction of cellular components, is generating huge amounts of data at increasing pace.

Bioinformatics is required to evaluate and organize the data and make them available to researchers. The yeast community is realizing that it studies an incredibly complex network of interconnecting pathways and processes. In order to interpret the maltitude of data in terms of biological functionality, i.e. networks and processes (systems), mathematical models are required. Such models should be realistic descriptions or "replicas" of the biological system. The models should be able to elucidate system properties and assist phrasing of hypotheses for experimental studies and predict the outcome of (even subtle) perturbations, such as mutations or drug effects. Therefore, mathematical models are tools for data interpretation and experimental planning as well as being a vehicle for exploring insights on fundamental principles behind biological systems.

In order for the models to describe *dynamics* of cellular processes realistically, such as flow of metabolites or information through pathways, quantitative data, time courses and spatial information is required. Assessing the information available on even the best studied cellular system, *S. cerevisiae*, most data are of the "yes/no" character, i.e. qualitative. Those have been useful to define pathways and describe functions but they are largely inadequate to describe processes. To obtain quantitative data and to capture subtle changes is trivial in some instances (such as relative levels of certain metabolites, proteins, mRNA) but in many cases a technological challenge (quantification of absolute levels of biomolecules, for instance). However, *quantitative* differences, i.e. if a protein kinase is 30% or 60% "on", may make a major difference in cell behavior and drug effects. Ultimately, we would like to be able to measure (or predict) AIMS-4D (Amount, Interactions, Modifications, Spatial movements at each XYZ+T coordinate) for every component. For these reasons we need to advance technology to assess quantitative data.



The use of mathematical models is part of an important transition that is already occurring in biological research: moving from description to understanding. Much of genetic and molecular data *describe* a system but do not *explain* why, for instance, a specific feedback loop is implemented or why the system is robust against a certain mutation. Modeling helps providing a logical *explanation* and in many instances is the only way to discover such regulatory features in the first place.

Understanding biological systems is crucial for understanding diseases. Hence, one of the major promises of Systems Biology is to help understanding how diseases alter biological processes, identify the appropriate drug targets, and circumvent feedback control and robustness in drug development and to design the appropriate treatment for the individual patient.

Systems Biology operates as collaboration between experimentalists from different disciplines such as molecular biology, chemistry and physics on the one hand and bioinformaticians, computer scientists, mathematicians and engineers on the other to generate and improve models iteratively, and to explore fundamental principles. To achieve such collaboration, interdisciplinary projects, networks and even institutes are being formed all over the world. The yeast community has the expertise, excellence and the capacity to form an interdisciplinary network of Systems Biology. The Network is needed to instrumentalize such an effort and provide the platform to integrate data acquisition, data generation, modeling and recursive model optimization.

Yeast as a Model System

One of the major objectives of Systems Biology is to be able to develop general concepts for quantitative description of living cells. This requires model systems in which very detailed studies can be performed rapidly, accurately and reproducibly. The yeast *Saccharomyces cerevisiae* is the best and most intensely studied eukaryotic cell and serves as an excellent model system:

- *S. cerevisiae* has the best studied eukaryotic genome, with now even several closely related genome sequences being available.
- Yeast has always been at the forefront of the development of large-scale genomic approaches. Essentially all high-throughput functional genomics techniques and the underlying bioinformatics for analysis of the transcriptome, the proteome, the metabolome, and the interactome were originally developed using yeast as model system. This has resulted in large databases, which represent a valuable though not sufficient source for systems biology.
- Genetic manipulations can be performed with greater ease than in any other cell, and human (and other eukaryotic) genes can easily be expressed and studied in *S. cerevisiae*.
- Yeast constitutes the largest eukaryotic toolbox with totally unique resources. A complete set of *S. cerevisiae* deletion mutants for all 6,000 open reading-frames is available. Moreover, genome-wide sets of ORFs tagged for analysis of protein localization, protein levels and protein complex purification are also available.
- *S. cerevisiae* can be cultivated at well controlled conditions, allowing for studies of high reproducibility that are difficult to perform with any other model system.
- Yeast is a paradigm for eukaryotic cells. Most fundamental cellular processes are conserved from *S. cerevisiae* to humans and have first been discovered in yeast. Moreover, even though *S. cerevisiae* is only distantly related to human cells, many disease-related genes have been identified in this yeast.
- The yeast research community is large (more than 1,000 laboratories in Europe, USA, Canada and Japan plus many more elsewhere) and involves many world leading biological researchers, including Nobel laureates. The research community is well organized with large annual conferences; it is open for exchange of data and material and has a demonstrated record in joint achievements.
- Due to the ease of handling and the highly developed genetics, functional genomics and cell biology yeast is an excellent model organism for training of candidates in cell biology, bioinformatics and Systems Biology.

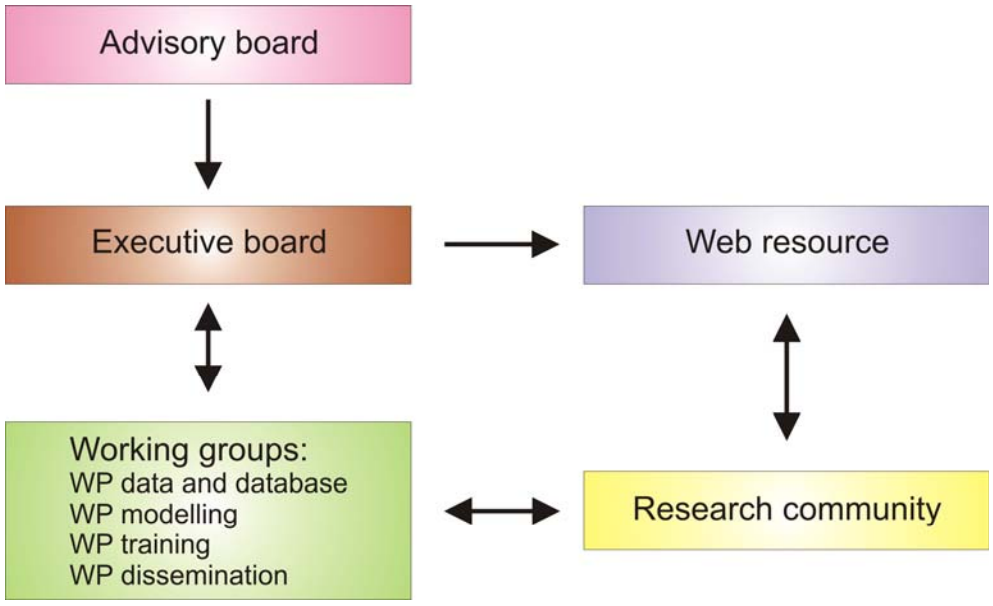
Only with the yeast *S. cerevisiae* as a model organism will it be possible to apply a full-scale integrated analysis of a given organism. Although at this stage defined cellular modules are approached by Systems Biology the “whole cell vision” is an important element of the field as it will lead to development of the necessary knowledge base, tools and resources needed to decipher the function of genes in higher eukaryotes and enable quantitative mapping of all interactions between the individual components in the system.

A systems biology initiative on yeast will have substantial impact on society as concepts and methods will migrate to other research fields studying plants, animals and humans.

Possible organization

The organization of the Network is proposed to consists of the following elements

- The web resource as an interface between the yeast research community (data acquisition), the community of bioinformaticians, mathematicians, computer scientists and engineers (modeling) and back to the yeast research community (use and improvement of models). This web resource (forming a "virtual research community") could be closely linked to but probably be distinct from the Saccharomyces Genome Database SGD, the Yeast Resource Center at Washington University and the Munich Institute for Protein Sequences MIPS. The web resource will store and make available relevant data in collaboration with SGD. The models that are made available on this web site shall be described based on widely accepted standard representation schemes. Building and maintaining the resource (coordinator, curators) will require dedicated funding. It does not need to be physically located at one place. http://www.YSBN.org/ will be used as a gateway for the resources of this initiative.
- The executive board, which overseas the web resource and coordinates all activities on a day-by-day basis. It will consist of chair persons of the working groups and a small number of coordinators.
- The advisory board, which consists of highly respected and influential members of the relevant research communities. It assists the executive board and ensures high visibility of the field to open up opportunities for funding of community research.
- Working groups, which consist of high level researchers in relevant areas of biology and bioinformatics, computer sciences, physics and mathematics. The working groups suggest the content of the web resource, future activities and work out the relevant detailed strategies such as standards for data presentation and modeling.
- The contributing and profiting research communities, who will provide data and use, test and help improving models. It will apply for own funding in quantitative/systems biology using the Alliance to support their grant applications.



Issues for getting the Network into practice during 2004

The web resource should be started in 2004 and in the first place illustrate success stories of Systems Biology, present tools for quantitative analysis (or links to such tools) and link to already existing websites where computer simulations are run. To establish this basic “advertising place” should be done from existing resources.

During the course of 2004 the Network will have to establish its organisatory structure, apply for the funds needed to run the web resource and get the working groups started.

An important aspect is to ensure proper standardization of how experimental as well as modeling procedures are to be documented. This standardization needs to take into account the diversity of yeast strains and experimental regimes and is required to define proper standardization procedures for documentation of experimental results. MIAME, which is a definition on minimal information to be documented in connection with storing transcriptome analysis data, is an initiative in this direction. The type of definition for data presentation and documentation will encourage researchers to produce time courses, quantitative data and to assess spatial information and hence to contribute to the goals of the Alliance. In terms of modeling, the standards and documentation procedures should allow new models to be integrated into existing models. SBML standard (http://www.sbml.org/), which is widely accepted de facto standard for model representation will be a basis of model definition, but need to be augmented to incorporate rich features that may be required for the alliance’s purposes. Thus, the alliance shall maintain close communication with the SBML group, as well as other standard formation efforts. All the above aspects should be discussed and put in place by working groups.

A major issue is how to organize databases such that data of different types can be integrated. To a large extent, different pre-existing databases will have to be made to work together. It should be possible to easily access various data types (protein localization, complex formation, mRNA turnover etc.) that may be available. Initiatives at regional levels are underway to address integration of databases and analysis tools.

In its initial phases the Network will focus on cellular modules where Systems Biology is already being applied such as glycolytic metabolism (strong European tradition in metabolic control analysis, see e.g. http://jjj.biochem.sun.ac.za/); galactose metabolism (http://www.systemsbiology.org), signaling (EC-funded projects as well as different projects in the US and Japan) and the cell cycle (work by Novak and Tyson, EC-funded project). Further modules are added as they emerge in the community or suggested from internal discussions in the Network. The initial models will be derived from existing data from global approaches and the literature, with subsequent additions from new data produced by the community.

Time line

Visibility and PR: ongoing, to be increased in 2004

Website with success stories: Mid 2004

Next meeting to put in place the organisatory structure: May 14-16 in (near) Copenhagen

First projects integrated: Summer 2004

Information and advertising: Seattle yeast conference August 2004

Full launch: October 2004 at Heidelberg conference

Contributers

The following individuals have seen this draft so far and their comments have been taken into account. This does not mean that any of these individuals fully supports the initiative as a whole or any/all of its elements:

Steven Oliver, Edda Klipp, Per Sunnerhagen, Anders Blomberg, Peter Philipssen, Marja Makarow, Mark Johnston, Lilia Alberghina, Karl Kuchler, Søren Brunak, Michael Hucka, Michael Cherry, Michel Aigre, Tau-Mu Yi, Erin O'Shea.

The draft has been slightly altered after the workshop in Copenhagen (May 14-16). See separate report on list of participants and results.