Journal of Biotechnology 144 (2009) 204–211

Contents lists available at ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

Impact of yeast systems biology on industrial biotechnology

Dina Petranovic*, Goutham N. Vemuri*

Systems Biology, Department of Chemical and Biological Engineering, Chalmers University of Technology, Kemivägen 10, Göteborg 412 96, Sweden

ARTICLE INFO

Article history:
Received 10 March 2009
Received in revised form 29 June 2009
Accepted 8 July 2009

Keywords:
Systems biology
Saccharomyces cerevisiae
Industrial biotechnology
Metabolic engineering

ABSTRACT

Systems biology is yet an emerging discipline that aims to quantitatively describe and predict the functioning of a biological system. This nascent discipline relies on the recent advances in the analytical technology (such as DNA microarrays, mass spectromety, etc.) to quantify cellular characteristics (such as gene expression, protein and metabolite abundance, etc.) and computational methods to integrate information from these measurements. The model eukaryote, *Saccharomyces cerevisiae*, has played a pivotal role in the development of many of these analytical and computational methods and consequently is the biological system of choice for testing new hypotheses. The knowledge gained from such studies in *S. cerevisiae* is proving to be extremely useful in designing metabolism that is targeted to specific industrial applications. As a result, the portfolio of products that are being produced using this yeast is expanding rapidly. We review the recent developments in yeast systems biology and how they relate to industrial biotechnology.

1. Introduction

Biology aims to understand how genetic or environmental perturbations bring about changes in living systems. In this context, a system could be a set of genes or proteins (that work as a module), the entire cell, the whole organism, or even the population itself. The notion that sequencing the genome of an organism will enable achieving this goal still remains a promise. This is largely because of the immense complexity that is inherent to biological systems. Irrespective of whether we restrict the system to a set of few proteins or consider the entire population as a system, we can expect a high level of complexity which is reflected by the number of components and interactions between them. Systems biology aims to integrate interactions between entities to facilitate quantitative prediction of the phenotype in response to a perturbation. There are many views of systems biology, and all of them convey a unified theme of the whole being greater than the sum of its parts. As previously summarized (Nielsen and Jewett, 2008), systems biology aims "to obtain new insight into the molecular mechanisms occurring in living cells or sub-systems of living cells for predicting the function of biological systems through the combination of mathematical modelling and experimental biology". Here we adopt this view of systems biology.

In any systemic study, the typical approach is to first define the system, quantify its components and then elucidate how the components interact to impart the system its properties. The study of biological systems is still in the infant phase where the systemic components are being identified and quantified, thanks to the rapidly growing technology. The subsequent step of elucidating the interaction between them still remains a challenge. Nevertheless, progress in this direction holds the promise to quantitatively explain biological systems. In this review, we present a brief overview of high-throughput experimentation and data integration techniques and relate the progress in this field to industrial biotechnology. We frequently refer to examples from the unicellular eukaryote, *Saccharomyces cerevisiae*, to highlight the progress made in systems biology as well as to relate the impact of this field on industrial biotechnology.

2. The *-omes* of systems biology

The confluence of systems science with biology was documented 95 years ago, in Walter Cannon's homeostasis theory that described the human body as a dynamic control system (Cannon, 1915). This concept of control of a biological system was popularized by the work of Jacob and Monod to explain the functioning of the lac operon (Jacob and Monod, 1961), the biochemical systems theory (Savageau, 1969) and the metabolic control analysis (Kacser and Burns, 1973). Therefore, the concept of data integration using mathematical models to understand biological principles is not new. What is new is our ability to accurately quantify cellular components, even at a global scale. Technological developments towards this goal are motivated by the availability of genome sequences.

Although the technology to sequence DNA was developed much earlier (Sanger and Coulson, 1975), it is only in the last decade that

* Corresponding authors.
E-mail addresses: dina.petranovic@chalmers.se (D. Petranovic), goutham@chalmers.se (G.N. Vemuri).

0168-1656/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbiotec.2009.07.005

sequencing the complete genomes and annotating their content became routine. The genome of *S. cerevisiae* was one of the first to be completed (Goffeau et al., 1996). As of May 2009, genome sequences of about 3000 organisms are publicly available in GenBank. The genome only defines the phenotypic space within which an organism can operate. It is a static entity, with its content relatively constant with time and environment. The availability of genome sequences created the need to develop the technology to quantify other components of cells such as mRNA, proteins and metabolites to identify context-dependent utilization of the genome. The maturation of quantitative biology naturally expanded the vocabulary for appropriate description of the observations and launched a whole set of "omes" to collectively describe a class of components. Fig. 1 illustrates the commonly used omes and the potential information that can be gleaned from them. These "omes" constitute the toolbox for a systems level understanding.

Unlike the genome, the other "omes" shown in Fig. 1 are context-dependent and vary with genetic and environmental perturbations. The primary response to any perturbation occurs at the level of the transcriptome. The transcript levels are quantified by the microarray technology, which has become increasingly popular (Vemuri and Aristidou, 2005) thanks to the availability of the genome sequences of many model organisms. It is now a routine procedure to simultaneously measure the condition-dependant abundance of mRNA species of every ORF in the genome using microarrays. Microarrays come as two-dye spotted arrays (Schena et al., 1995) or GeneChips (Lockhart et al., 1996) that contain primers or oligonucleotides that uniquely represent each ORF (Dufva, 2009). In addition to quantifying gene expression, microarrays are also useful in predicting the function of previously uncharacterized genes and identify signature patterns in gene expression profiles (Fig. 1), based on the assumption that co-expressed genes are co-regulated.

The proteome describes the set of all the proteins in the system. Analysis of proteins divulges information on protein function and the pathways on which they act (Fig. 1). Moreover, although proteins are the end products of gene transcription, early reports did not find any direct linear correlation between the number of mRNA molecules and the number of corresponding proteins (Griffin et al., 2002; Gygi et al., 2000) so the mere transcriptome analysis may not reflect the functional profile at the protein level. The conventional proteomics workflow comprises separation of proteins using two-dimensional gel electrophoresis (2DGE) and identifying the proteins using mass spectrometer (MS). Although protein separation by 2DGE was introduced over 30 years ago (O'Farrell, 1975), the popularity of its application in proteomics increased only after the development of MS-based high-throughput protein identification technologies (Mann et al., 2001). Using this method, the sub-proteome map of the yeast mitochondria was generated (Ohlmeier et al., 2004). Methods based on 2DGE have inherent disadvantages such as being labor intensive, poor gel-to-gel comparison and inability to detect low-abundance proteins. More recently, proteomics methods that combine liquid chromatography (LC) with MS proved to be a more efficient method (Shi et al., 2004; Washburn et al., 2001). By stable isotope labeling of the amino acids in cell culture (SILAC), it is possible to quantify the proteome to a high precision. This method relies on the incorporation of amino acids substituted with isotopic backbone into proteins. By following the incorporation of "labeled" amino acids into the protein using liquid chromatography coupled with mass spectrometry (de Godoy et al., 2006), it was possible to quantify over 2000 proteins (even those occurring in low abundance) in *S. cerevisiae* proteome. Using the MUDPIT (Multi Dimensional Protein Identification Technology) method (Washburn et al., 2001), more recently over 2400 proteins were quantified in yeast (Usaite et al., 2008). Furthermore, proteomics also demonstrated that yeasts with different ploidy status differ in the levels of the proteins (de Godoy et al., 2008). Besides the MS-based proteomics, protein arrays are rapidly becoming powerful high-throughput tools to identify proteins, monitor their expression, and elucidate their function and interactions in yeast. Despite many challenges, the development of protein microarrays has begun to achieve some recent success (Bertone and Snyder, 2005; Hall et al., 2007). Currently protein arrays are available in two main formats: (1) abundance-based microarrays, which seek to measure the abundance of specific biomolecules using analyte-specific reagents, such as antibodies and (2) function-based microarrays, which examine protein function in high-throughput by printing a collection of target proteins on the array surface and assessing their interactions and biochemical activities. The main obstacle in developing the protein array technology is the availability of sufficient amount of proteins easily. Since protein amplification methods (analogous to the DNA amplification methods such as the PCR) do not exist, they have to be produced recombinantly and purified.

The third "*-ome*" commonly used is the metabolome, which describes the abundances of the metabolites in the system (Nielsen and Oliver, 2005). The cells control the concentrations of their intracellular metabolites very rigidly. There is normally a very low tolerance on the allowable variation in the metabolite concentrations for a given physiological state. Since they are the intermediates of biochemical reactions, metabolites decide the connectivity in the metabolic network. Certain metabolites such as ATP or NADH, which are involved in a large number of reactions in the metabolic network, are capable of bringing about significant changes in large parts of the metabolism (Nielsen, 2003). Similar to the transcriptome and proteome, the metabolome also presents a snapshot of the physiological state of the cell. Therefore, measuring the changes in the concentrations of intracellular metabolites would reveal an aspect of regulation that is not covered by the other "omes". Metabolite profiling is now considered an important part of systems biology, playing a complementary role to genomics and proteomics (Fig. 1). However, this field is still in its infancy, the main obstacle being the enormous diversity in the chemical structure of the metabolites (Jewett et al., 2006; Mapelli et al., 2008; Nielsen and Oliver, 2005; Oldiges et al., 2007). Although the num-

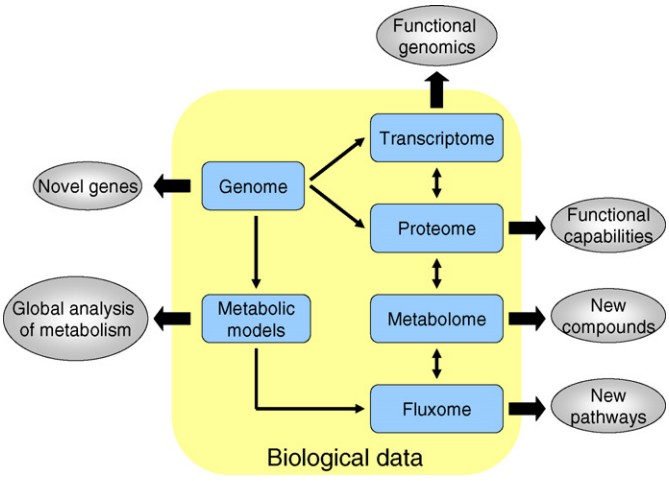

Fig. 1. The diverse set of "omes" that are considered to be part of the systems biology arsenal. Each "ome" provides global information on a cellular component. For example, the transcriptome captures global information on all the mRNA transcripts and the proteome captures information on the abundance level of all the proteins. As described in the text, not one "ome" by itself can divulge complete information about cellular processes. Even if all the cellular components were to be quantified, the challenge still lies in processing the data and extracting information about interactions between the different "omes". Different aspects of high-throughput biological data are shown in the figure and the relevant information they provide that has direct applications in industrial biotechnology.

ber of metabolites in a cell is usually one order lower compared with the number of genes or proteins, it is virtually impossible to simultaneously determine the complete metabolome with current technology. Mass spectrometry (MS) and nuclear magnetic resonance (NMR) are the most frequently used methods of detection in the analysis of the metabolome (Dunn, 2008; Powers, 2007). NMR is very useful in determining the structure of unknown compounds, but comes with the drawback of expensive instrumentation. However, it is non-destructive to samples and provides rich information on the structures of molecules in complex mixtures. On the other hand, MS is considerably more sensitive and facilitates the identification of unknown and unexpected compounds. The combination of separating the metabolites using a GC or LC coupled with MS is transpiring to be the most promising technique for metabolite profiling.

The fluxome quantifies the physiological state of the cell, and is the fourth "ome" that is commonly used in industrial biotechnology. In contrast to the "snapshots" of cellular state given by the transcriptome, proteome or the metabolome, the fluxes are time-dependent (Sauer, 2006). The metabolic fluxes can be *estimated* using stoichiometric mass balance models using appropriate objective functions. However, since the estimate is very much tuned to the user-defined objective function, which may not be the biological reality, discrepancies in the actual metabolic flux and the estimated flux may arise. To experimentally *measure* metabolic fluxes, stable isotope tracers such as ^{13}C-labeled substrates are used (Christensen et al., 2002; Wiechert et al., 2001). Since the turnover time of the primary metabolites does not permit their accurate quantification, the labeling patterns in the proteinogenic amino acids, which have slower turnover time as well as preserved carbon backbones from the primary metabolites, are quantified instead. Metabolic fluxes can be inferred from ^{13}C data using either NMR or MS. Both approaches use a data-fitting algorithm where the extracellular fluxes and biomass synthesis are simultaneously integrated with a metabolic network model to iteratively identify a flux distribution pattern where the difference between the simulated isotope pattern and the measured isotope pattern is minimized (Sauer, 2006). Currently, this approach is limited to only about 50 reactions of the central metabolism.

In order to get an estimate of the carbon flow pattern in the entire metabolic pathways, genome-scale metabolic models have been developed for *S. cerevisiae* (Duarte et al., 2004; Forster et al., 2003; Herrgard et al., 2008; Kuepfer et al., 2005; Nookaew et al., 2008) that aim to encompass all the metabolic reactions. Based on the stoichiometry of the metabolites involved in the different reactions and experimentally observed constraints on some reactions, the carbon flow distribution is estimated using linear programming. The constraints that are imposed to reduce the solution space range from physico-chemical and environmental to regulatory constraints. The real utility of genome-scale metabolic models lies in analyzing the feasible solution space after imposing the constraints. For example, this can be used to identify metabolic capabilities such as flux variability, effect of gene deletions and additions and alternate optima (Price et al., 2004). The experimentally determined fluxes using isotope tracers could be used to further constrain the solution space for improved accuracy of the predictions. Additionally, genome-scale metabolic models serve as excellent databases for organisms as they involve extensive compiling of information from genome sequences, enzyme and pathway databases, physiological data, transcription interactions and other high-throughput approaches (Feist et al., 2009).

In addition to the stoichiometric models that are currently popular, dynamic models based on mechanistic properties of the system also enable understanding the systemic properties. The kinetic models that have been traditionally developed to study metabolism (Gadgil et al., 1996; Kompala et al., 1984), dynamic models have been successfully used also to understand the regulation. Particularly, the galactose regulon in *S. cerevisiae* has served as a paradigm for this approach. The genes of the galactose pathway (Leloir pathway) are relatively inert in the absence of galactose and are highly expressed when galactose is the only carbon source (Yarger et al., 1984). Temporal modeling of the *GAL* regulon at different concentrations of galactose revealed the ultrasensitive nature of the switch of the galactose genes depending on the localization of Gal80 and its interaction with transcription activator, Gal4 (Verma et al., 2003). More recently, dynamic models were used to integrate aspects of cell signaling and regulation of the osmotic shock response in *S. cerevisiae* (Klipp et al., 2005). This comprehensive model of the *HOG* pathway (mitogen-activated protein kinase in yeast) captured the dynamics of gene expression under different stress conditions and also predicted a new feedback regulation, triggered by osmolyte accumulation in the cell. Besides contributing to understanding the specific regulatory pathways (which involve one or more regulatory proteins) mathematical models can be used for studies of cell cycle (Alberghina et al., 2009; Barberis and Klipp, 2007). Mathematical models have also been used in the aging studies such as the study of selective benefits of asymmetrical cytokinesis of *S. cerevisiae* in which it was shown that systems dividing asymmetrically (size-wise) or displaying damage segregation can withstand higher degrees of damage before entering clonal senescence (Erjavec et al., 2008).

3. Two views the system: top-down and bottom-up systems biology

Progress in molecular biology demanded biologists to think at the molecular level and relate biological phenomena to a few genes or proteins. Such reductionist approaches involve describing a basal sub-set of the system, such as a few genes or proteins, in great detail. These elements are reconstituted to form larger subsystems until the top-level system is comprehensively defined (Fig. 2). This approach is more commonly known as the bottom-up approach to process biological information. The generation of high-throughput quantitative data on gene expression and protein abundance initiated a top-down approach to studying biological systems. The top-down approach establishes our knowledge of the system and disassembles it to understand the interconnectivity between different functional modules.

The bottom-up approach heavily relies on the accumulated knowledge base and captures the functional properties that emerge by integrating subsystems that have been characterized to a high level of detail, usually by series of ordinary differential equations (Fig. 2). Since the subsystems are studied and modeled in isolation, the bottom-up approach runs the inherent risk of the complete system not performing optimally when integrated with other subsystems. Therefore the hypotheses and inferences related to this approach are best applicable at the level of genes or pathways that form an independent functional module. Consequently, the experiments required for evaluating the modules are not necessarily high-throughput, but rather focussed. Bottom-up systems biology is based on existing detailed knowledge, which is then translated into a mathematical formulation used to simulate the behaviour of the system. Models of osmotic response in yeast to predict osmolyte accumulation and feedback control (Klipp et al., 2005), pheromone response to discover the presence of a feedback loop to increase the fidelity of information transmission (Yu et al., 2008) or the hypersensitivity of the galactose genes to the nucleocytoplasmic shuttling of Gal80 (Verma et al., 2003) are examples of bottom-up systems biology approaches.

In contrast to the bottom-up approach, the holistic top-down approach which is data-driven or inductive, establishes our

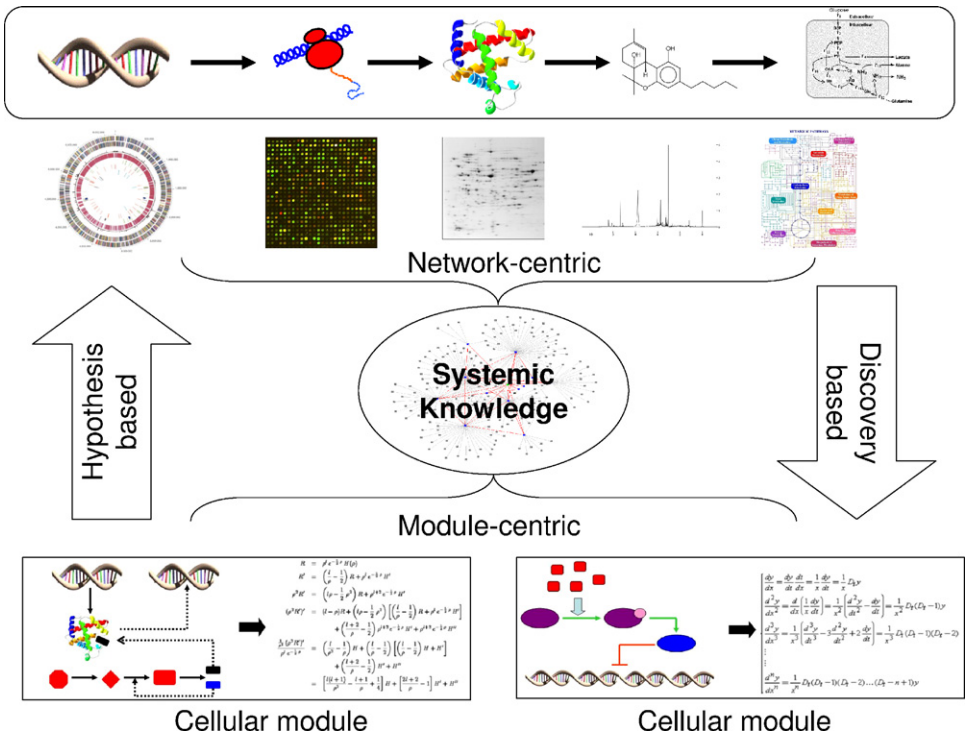

Fig. 2. The two ways to understanding a system: a holistic network analysis or top-down approach; and a deterministic modular analysis or bottom-up approach. The philosophy and the experimental design that drive these two approaches are described in the text. Despite some differences, the two ways communicate the common theme of system complexity.

knowledge of the system and interconnectivity between different functional modules without an *a priori* hypothesis. The basic philosophy of this approach is that the systemic properties emerge from the collaborative behaviour of its components the whole is greater than the sum of its parts. Unlike the bottom-up approach where the goal is to exhaustively characterize the biochemistry of the pathways, the top-down approach aims to characterize networks on a global scale. The primary goal of this approach is to discover new interactions between the subsystems that involve several iterations of experimentation, analysis, hypothesis generation/modification and further experimentation to validate the hypotheses. The top-down approach uses global approaches with transcriptome, proteome and metabolome data, and is therefore largely phenomenological and studies the dynamics of the entire network (Fig. 2). The approach to model the metabolism using genome-scale models (Feist et al., 2009), quantifying regulation using transcriptional regulatory networks (Liao et al., 2003), approaches to understand the structure and function of proteins (de Godoy et al., 2008; Washburn et al., 2001), etc. are examples of top-down approaches to systems biology. Despite these differences, the two complementary approaches acknowledge the complexity of biological systems and aim at the better quantitative understanding.

There are different mathematical models that can be applied in either top-down or bottom-up systems biology such as models based on ordinary differential equations, stochastic models, stoichiometric models and graph models. Ideker and Lauffenburger (2003) classified mathematical models used in systems biology as high-level models and low-level models, where the high-level models refer to a top-down approach where structures, interactions and their strengths are extracted from global data and low-level models refer to a bottom-up approach where the system is reconstructed from quantifying all the interactions (Ideker and Lauffenburger, 2003). Most bottom-up driven models only describe a sub-set of the complete biological system, but they can still be extremely useful like the metabolic network models that are based on collecting the stoichiometry for all metabolic reactions and through the use of flux balance analysis, and linear programming and stoichiometric models can be used for simulation of growth and product formation (Famili et al., 2003; Forster et al., 2003; Price et al., 2004). More recently, a large-scale community-based jamboree was devoted to developing and updating the genome-scale metabolic model of yeast that can be used for such simulation (Herrgard et al., 2008). The models used in top-down systems biology can be the so-called "soft models" such as neural networks, graphs or even statistical models.

4. Systems biology strategy for metabolic engineering

Exploitation of biological systems holds the promise to provide cure for diseases, fuel from renewable resources, and enable sustainable production of bulk chemicals as well as drugs, pharmaceuticals and nutraceuticals. Naturally occurring biological systems are not capable of realizing these imposed, ambitious goals. Given the complexity of the goals, it is not immediately obvious how to change them to be able to fulfill these roles. Metabolic engineering is the rational development of cell factories using directed genetic engineering for industrial applications (Nielsen, 2001). The development process involves improvement of existing cell factories or designing new ones. Traditional metabolic engineering was limited to overexpressing or introducing pathways into a host organism to produce a specific compound and/or deleting those pathways that contribute to by-product formation. While this approach enjoyed moderate success, more complex phenotypes such as expanding the range of substrates and enhancing the stress tolerance could not be achieved. The second generation strategy of metabolic engineering focussed on engineering the systems involved in regulation of metabolism. The rationale behind this strategy is that, since regulatory proteins control a large number of metabolic genes, a more pronounced phenotype can be achieved by manipulating the action

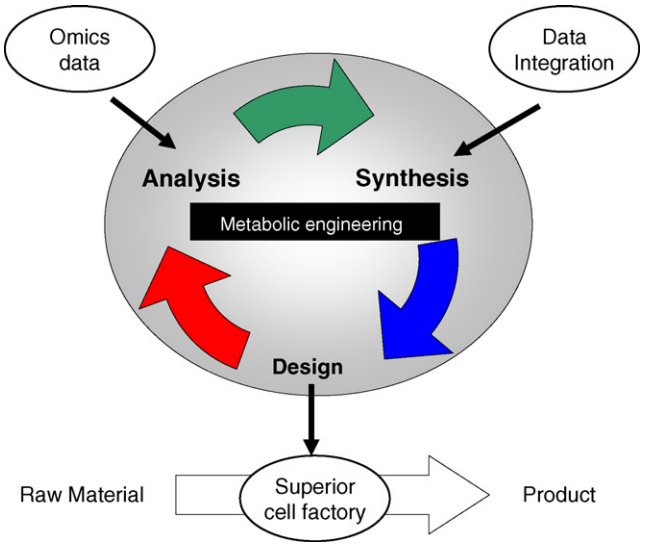

Fig. 3. Applying the principles of systems biology to metabolic engineering. The fundamental engineering principles of analysis, synthesis and design of metabolic networks will be greatly benefited by superposing high-throughput data on metabolic pathways. The omics data provide the much needed information on the factors that control metabolic pathways leading to a greater understanding of the system. This will facilitate effective design of superior cell factories that are tailored for a specific application.

of regulatory proteins (Vemuri and Aristidou, 2005). Besides applications in medical research, metabolic engineering is the other important applied outcome of systems biology. In the context of metabolic engineering, systems biology is a powerful tool that incorporates high-throughput data, leading to the synthesis and design of a cell factory that can efficiently and effectively convert raw materials into the desired products (Fig. 3).

Metabolic engineering using systems biology relies on the availability of tools shown in Fig. 1. For any given product, the most important decision is the choice of the organism. The characteristics of an ideal organism include high product titers, minimal by-product formation, easy cultivation techniques using cheap raw materials, amenability to genetic modifications and safe to work with. Given these basic requirements, the genome sequence can be determined and other high-throughput tools such as DNA microarrays and protein chips can be fabricated. It is quite common that many products we desire to produce using bioprocesses are found in microorganisms in nature that do not meet many of the basic requirements mentioned above. Consequently, it is common to introduce the pathways for these products in well-established model systems such as *S. cerevisiae* by cloning the appropriate genes (Nevoigt, 2008). Conventionally, *S. cerevisiae* is used to produce recombinant proteins and bulk compounds such as ethanol. However, the portfolio of compounds that can be produced in *S. cerevisiae* expanded impressively (Table 1). To a great extent, research in the area of systems biology is also dependent on the market demand. For example, in the present times, when the demand for the production of alternate energy sources is high, the capability of *S. cerevisiae* to produce higher alcohols such as butanol and methyl butanol is investigated (Schoondermark-Stolk et al., 2006; Steen et al., 2008).

We argue that once the base strain is constructed, systems biology can improve the product titers to achieve the near-maximum yields. Two main hurdles that impede product formation are product toxicity and regulation-based inhibition. When using *S. cerevisiae*, extensive information on the regulatory networks and metabolic pathways is already available. By coupling this information with large-scale genome-wide analyses, an integrated bioprocess can be developed. We illustrate this concept using some case studies of successful metabolic engineering strategies using systems biology approaches.

4.1. Case study 1: increasing tolerance to ethanol stress

Ethanol production by *S. cerevisiae* is arguably the most well-studied application in biotechnology. Although near-maximal titers of ethanol are naturally produced by *S. cerevisiae*, this process is severely hampered by the stress caused by high concentration of ethanol. Complex functions, such as stress response, involve multiple genes and conventional approaches of metabolic engineering cannot be employed to engineer these phenomena. The conventional approach of random mutagenesis and screening has been used to instil complex phenotypes. The greatest disadvantage with this method is that the nature of genetic changes is not known and hence, cannot be transferred into other strains. Global transcription machinery engineering aims to modify the transcription factor behaviour (global transcription machinery engineering) to regulate the expression of these genes (Alper and Stephanopoulos, 2007). The global influence of the transcription factors makes them an attractive tool for strain construction. In this study, random mutations were introduced in Spt15 TATA binding protein to induce changes in global gene expression profile (Alper et al., 2006). The exact molecular changes that bring about the changes in gene expression are not known by this method, but the net outcome can be easily determined by quantifying gene expression. In contrast to the conventional evolutionary engineering applications where the exact locus of the mutation is now known, this method allows to transfer the mutation to other strains.

4.2. Case study 2: reducing overflow metabolism

Applications related to biomass applications such as recombinant protein production that require elimination of overflow metabolism. Ethanol production is redox-neutral with glucose. But the biomass that is produced concomitantly generates NADH. The additional NADH generated by biomass synthesis is reoxidized by producing glycerol (Overkamp et al., 2000). Since aerobic production of ethanol is believed to be due to inadequate respiratory capacity in *S. cerevisiae* (Sonnleitner and Kappeli, 1986), increasing the capacity to oxidize NADH would decrease ethanol overflow. When cytosolic NADH oxidase was overexpressed, glycerol production was completely eliminated with only a marginal decrease in ethanol production (Vemuri et al., 2007). By affecting NADH levels in the mitochondria using alternative oxidase, it was shown that ethanol production, which occurs in the cytosol, could be reduced. The study concluded that transcriptional regulation of the genes of the TCA cycle and respiration and allosteric enzyme inhibition contributed to a coordinated decrease in respiration. These affects are brought about by glucose repression, which is prevalent during exponential growth of *S. cerevisiae* on glucose. Mitochondrial redox state was subsequently demonstrated to increase the efficiency of oxidative phosphorylation (measured by the *P*/*O* ratio) and play a crucial role in decreasing ethanol production (Hou et al., 2009). In an independent study, ethanol overflow was eliminated by replacing all the hexose transporters with only a chimeric transporter. This transporter was a hybrid of Hxt1 and Hxt7 and when expressed in *S. cerevisiae* also decreased overflow metabolism (Otterstedt et al., 2004). These two independent strategies to overcome ethanol production suggest the presence of an underlying mechanism centered around the redox and energy metabolism that eventually controls the gene expression and metabolic flux channeling to ethanol from glucose.

Table 1
Summary of the products from yeast.

Hormones	Production of insulin and insulin precursors. Through engineering of leader sequences the productivity of protein production has been increased substantially	
Vaccines	Production of hepatitis vaccines. Through expression of a virus surface protein in yeast an efficient vaccine has been developed	
Organic acids	Production of lactic acid. Through expression of a heterologous lactic acid dehydrogenase it was shown possible to produce lactic acid in yeast	(Ishida et al., 2006; Porro et al., 1995)
Sesquiterpenes	Many different sesquiterpenes have been produced in yeast through heterologous expression of plant genes. This include the anti-malarial drug precursor artemisinic acid	(Ro and Douglas, 2004)
Carotenoids	By expression of bacterial genes β-carotene and lycopene was produced in yeast	(Yamano et al., 1994)
Diterpenoids	Through expression of 10 plant genes in yeast it was shown possible to reconstruct a major part of the biosynthetic route towards taxol	(Dejong et al., 2006)
Polyketides	Through combined expression of a polyketide synthetase encoding gene together with a an activating enzyme it has been shown possible to produce the simple polyketide antibiotic 6-MSA in high titers in yeast	(Mutka et al., 2006; Wattanachaisaereekul et al., 2008)

4.3. Case study 3: increasing galactose uptake

An important aspect of metabolic engineering is to link the phenotype with the genes. It is quite common to characterize the phenotype in response to the deletion/insertion of genes. With the advent of functional genomics, the inverse approach of identifying genes that can generate the desired phenotype should be possible. This method is called "inverse metabolic engineering" and requires extensive mapping of traits to genes before the common functionalities of genes could be identified. An example of this approach was to increase galactose uptake rate in *S. cerevisiae*. The regulation of galactose metabolism in *S. cerevisiae* has served as a paradigm for understanding eukaryotic gene expression (Sellick et al., 2008). Despite the structural similarity to glucose, galactose has a very different effect on the metabolism in *S. cerevisiae*. The key transcription factors that regulate the expression of the genes in the galactose pathway are Gap80 (repressor), Gal4 (activator), Mig1 (repressor) and Gap6 (post-transcriptional repressor). By deleting the three repressors (Gal80, Gal6 and Mig1), the uptake rate of galactose was improved by 40% (Ostergaard et al., 2000). Global gene expression analysis did not identify any genes of the galactose pathway whose expression differed in the triple deletion mutant or the Gap4 overexpressed strain, relative to the wild-type (Bro et al., 2005). However, the expression of the major isozyme of phosphoglucomutase (Pgm2) was increased in the strains with higher galactose uptake rate. Based on this discovery, the inverse metabolic engineering approach of overexpressing *PGM2* alone in the wild-type strain resulted in a 70% increase in the galactose uptake rate (Bro et al., 2005).

5. Challenges in systems biology

As in any new and developing technology, systems biology has many obstacles that prevent its widespread utility and commercial success in the area of industrial biotechnology. The obstacles can be grouped in at least three categories: (i) limitations in experimental techniques needed to gather large amount of data, (ii) gathering and processing of the data and (iii) interpretation of the data. Systems biology has the potential not only to improve industrial biotechnology of yeast and other organisms, or help us understand the functioning of biological systems in more general and fundamental terms. The above mentioned obstacles have to be cleared to pave the path forward. The main limitations in experimental techniques include the problems in measuring subtle changes in transcription, protein levels, metabolite levels or flux levels and problems in identifying relevant point mutations through whole genome sequencing. Limitations in gathering and handling data include at least problems related to data diversity and insufficient information about interactions between cellular components. Data that are generated in systems biology experiments are highly diverse, both in the context of what the data contain as well as how they are collected, processed and stored. For data to be useful outside of the laboratory in which they are generated, they must be standardized. Such standards, which require a minimum information amount of any ***e***xperiment such as microarray experiment (MIAME), proteome experiment (MIAPE), annotation of biochemical models (MIRIAM), genome sequence specification (MIGS), flow cytometry experiment, etc (Brazma et al., 2001; Field et al., 2008; Le Novere et al., 2005; Lee et al., 2008; Taylor et al., 2007). The lack of standards is prevalent when presenting networks, such as the interactions between proteins or between proteins and DNA and to an extent the genome-scale metabolic models.

The third type of challenge lies in the data interpretation and use of mathematical and bioinformatical tools. Currently, the development of computational tools for integrative biological analysis remains difficult to use by external users, even when the developers are willing to release it, because supporting documentation is often quite poor. The tools are highly localized and are suited to solve issues related to a specific research community. Consequently, there is a scarcity in predictive models that are validated and have diverse applications. The issue of model sensitivity and specificity is the major setback in developing dynamic models to describe the biological systems.

The introduction of high-throughput experimental techniques has clearly enabled faster progress in strain development and phenotypic characterization. In the future, with the continuous development of mathematical models and bioinformatics algorithms specifically suited for metabolic engineering, the value of using high-throughput experimental techniques for mapping detailed phenotypes is expected to increase. This is exemplified by the introduction of an algorithm for identification of reporter metabolites, which enables rapid identification of hot-spots in the metabolism based on transcriptome data (Patil and Nielsen, 2005). It is expected that mathematical models will be used more extensively in the design of metabolic engineering strategies. In order to exploit yeast systems biology for the field of industrial biotechnology, it will be important that metabolic models are extended to include regulation.

In present times, it seems likely that metabolic engineering benefits more from the top-down systems biology approach but in the future it will become desirable to have detailed kinetic models as this could enable the identification of new metabolic engineering strategies that also include fine-tuning of the activities and regulations of specific pathways.

Acknowledgments

We are thankful to Jens Nielsen for insightful discussions and to EU-funded coordination actions YSBN (FP6), and SYSINBIO (FP7) for financing the efforts in the field.

References

Alberghina, L., et al., 2009. Systems biology of the cell cycle of *Saccharomyces cerevisiae*: From network mining to system-level properties. Biotechnol. Adv..

Alper, H., et al., 2006. Engineering yeast transcription machinery for improved ethanol tolerance and production. Science 314, 1565–1568.

Alper, H., Stephanopoulos, G., 2007. Global transcription machinery engineering: a new approach for improving cellular phenotype. Metab. Eng. 9, 258–267.

Barberis, M., Klipp, E., 2007. Insights into the network controlling the G1/S transition in budding yeast. Genome Inform. 18, 85–99.

Bertone, P., Snyder, M., 2005. Advances in functional protein microarray technology. FEBS J. 272, 5400–5411.

Brazma, A., et al., 2001. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat. Genet. 29, 365–371.

Bro, C., et al., 2005. Improvement of galactose uptake in *Saccharomyces cerevisiae* through overexpression of phosphoglucomutase: example of transcript analysis as a tool in inverse metabolic engineering. Appl. Environ. Microbiol. 71, 6465–6472.

Cannon, W.B., 1915. Bodily Changes in Pain, Hunger, Fear and Rage, An Account of Recent Researches into the Function of Emotional Excitement. D Appleton and Co., New York and London.

Christensen, B., et al., 2002. Analysis of flux estimates based on (13)C-labelling experiments. Eur. J. Biochem. 269, 2795–2800.

de Godoy, L.M., et al., 2008. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature 455, 1251–1254.

de Godoy, L.M., et al., 2006. Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. Genome Biol. 7, R50.

Dejong, J.M., et al., 2006. Genetic engineering of taxol biosynthetic genes in *Saccharomyces cerevisiae*. Biotechnol. Bioeng. 93, 212–224.

Duarte, N.C., et al., 2004. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. Genome Res. 14, 1298–1309.

Dufva, M., 2009. Fabrication of DNA microarray. Methods Mol. Biol. 529, 63–79.

Dunn, W.B., 2008. Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. Phys. Biol. 5, 11001.

Erjavec, N., et al., 2008. Selective benefits of damage partitioning in unicellular systems and its effects on aging. Proc. Natl. Acad. Sci. U.S.A. 105, 18764–18769.

Famili, I., et al., 2003. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. Proc. Natl. Acad. Sci. U.S.A. 100, 13134–13139.

Feist, A.M., et al., 2009. Reconstruction of biochemical networks in microorganisms. Nat. Rev. Microbiol. 7, 129–143.

Field, D., et al., 2008. The minimum information about a genome sequence (MIGS) specification. Nat. Biotechnol. 26, 541–547.

Forster, J., et al., 2003. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. Genome Res. 13, 244–253.

Gadgil, C.J., et al., 1996. Cybernetic model for the growth of *Saccharomyces cerevisiae* on melibiose. Biotechnol. Prog. 12, 744–750.

Goffeau, A., et al., 1996. Life with 6000 genes. Science 274, 546,563-546,567.

Griffin, T.J., et al., 2002. Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. Mol. Cell. Proteomics 1, 323–333.

Gygi, S.P., et al., 2000. Measuring gene expression by quantitative proteome analysis. Curr. Opin. Biotechnol. 11, 396–401.

Hall, D.A., et al., 2007. Protein microarray technology. Mech. Ageing Dev. 128, 161–167.

Herrgard, M.J., et al., 2008. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nat. Biotechnol. 26, 1155–1160.

Hou, J., et al., 2009. Metabolic impact of redox cofactor perturbations in *Saccharomyces cerevisiae*. Metab. Eng..

Ideker, T., Lauffenburger, D., 2003. Building with a scaffold: emerging strategies for high- to low-level cellular modeling. Trends Biotechnol. 21, 255–262.

Ishida, N., et al., 2006. Metabolic engineering of *Saccharomyces cerevisiae* for efficient production of pure L-(+)-lactic acid. Appl. Biochem. Biotechnol. 129-132, 795–807.

Jacob, F., Monod, J., 1961. Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 3, 318–356.

Jewett, M.C., et al., 2006. Fungal metabolite analysis in genomics and phenomics. Curr. Opin. Biotechnol. 17, 191–197.

Kacser, H., Burns, J.A., 1973. The control of flux. Symp. Soc. Exp. Biol. 27, 65–104.

Klipp, E., et al., 2005. Integrative model of the response of yeast to osmotic shock. Nat. Biotechnol. 23, 975–982.

Kompala, D.S., et al., 1984. Cybernetic modeling of microbial growth on multiple substrates. Biotechnol. Bioeng. 26, 1272–1281.

Kuepfer, L., et al., 2005. Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. Genome Res. 15, 1421–1430.

Le Novere, N., et al., 2005. Minimum information requested in the annotation of biochemical models (MIRIAM). Nat. Biotechnol. 23, 1509–1515.

Lee, J.A., et al., 2008. MIFlowCyt: the minimum information about a Flow Cytometry Experiment. Cytometry A 73, 926–930.

Liao, J.C., et al., 2003. Network component analysis: reconstruction of regulatory signals in biological systems. Proc. Natl. Acad. Sci. U.S.A. 100, 15522–15527.

Lockhart, D.J., et al., 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat. Biotechnol. 14, 1675–1680.

Mann, M., et al., 2001. Analysis of proteins and proteomes by mass spectrometry. Annu. Rev. Biochem. 70, 437–473.

Mapelli, V., et al., 2008. Metabolic footprinting in microbiology: methods and applications in functional genomics and biotechnology. Trends Biotechnol. 26, 490–497.

Mutka, S.C., et al., 2006. Metabolic pathway engineering for complex polyketide biosynthesis in *Saccharomyces cerevisiae*. FEMS Yeast Res. 6, 40–47.

Nevoigt, E., 2008. Progress in metabolic engineering of *Saccharomyces cerevisiae*. Microbiol. Mol. Biol. Rev. 72, 379–412.

Nielsen, J., 2001. Metabolic engineering. Appl. Microbiol. Biotechnol. 55, 263–283.

Nielsen, J., 2003. It is all about metabolic fluxes. J. Bacteriol. 185, 7031–7035.

Nielsen, J., Jewett, M.C., 2008. Impact of systems biology on metabolic engineering of *Saccharomyces cerevisiae*. FEMS Yeast Res. 8, 122–131.

Nielsen, J., Oliver, S., 2005. The next wave in metabolome analysis. Trends Biotechnol. 23, 544–546.

Nookaew, I., et al., 2008. The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. BMC Syst. Biol. 2, 71.

O'Farrell, P.H., 1975. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. 250, 4007–4021.

Ohlmeier, S., et al., 2004. The yeast mitochondrial proteome, a study of fermentative and respiratory growth. J. Biol. Chem. 279, 3956–3979.

Oldiges, M., et al., 2007. Metabolomics: current state and evolving methodologies and tools. Appl. Microbiol. Biotechnol. 76, 495–511.

Ostergaard, S., et al., 2000. Increasing galactose consumption by *Saccharomyces cerevisiae* through metabolic engineering of the GAL gene regulatory network. Nat. Biotechnol. 18, 1283–1286.

Otterstedt, K., et al., 2004. Switching the mode of metabolism in the yeast *Saccharomyces cerevisiae*. EMBO Rep. 5, 532–537.

Overkamp, K.M., et al., 2000. In vivo analysis of the mechanisms for oxidation of cytosolic NADH by *Saccharomyces cerevisiae* mitochondria. J. Bacteriol. 182, 2823–2830.

Patil, K.R., Nielsen, J., 2005. Uncovering transcriptional regulation of metabolism by using metabolic network topology. Proc. Natl. Acad. Sci. U.S.A. 102, 2685–2689.

Porro, D., et al., 1995. Development of metabolically engineered *Saccharomyces cerevisiae* cells for the production of lactic acid. Biotechnol. Prog. 11, 294–298.

Powers, R., 2007. Functional genomics and NMR spectroscopy. Comb. Chem. High Throughput Screen 10, 676–697.

Price, N.D., et al., 2004. Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat. Rev. Microbiol. 2, 886–897.

Ro, D.K., Douglas, C.J., 2004. Reconstitution of the entry point of plant phenylpropanoid metabolism in yeast (*Saccharomyces cerevisiae*): implications for control of metabolic flux into the phenylpropanoid pathway. J. Biol. Chem. 279, 2600–2607.

Sanger, F., Coulson, A.R., 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J. Mol. Biol. 94, 441–448.

Sauer, U., 2006. Metabolic networks in motion: 13C-based flux analysis. Mol. Syst. Biol. 2, 62.

Savageau, M.A., 1969. Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions. J. Theor. Biol. 25, 365–369.

Schena, M., et al., 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270, 467–470.

Schoondermark-Stolk, S.A., et al., 2006. Rapid identification of target genes for 3-methyl-1-butanol production in *Saccharomyces cerevisiae*. Appl. Microbiol. Biotechnol. 70, 237–246.

Sellick, C.A., et al., 2008. Galactose metabolism in yeast-structure and regulation of the leloir pathway enzymes and the genes encoding them. Int. Rev. Cell. Mol. Biol. 269, 111–150.

Shi, Y., et al., 2004. The role of liquid chromatography in proteomics. J. Chromatogr. A 1053, 27–36.

Sonnleitner, B., Kappeli, O., 1986. Growth of *Saccharomyces cerevisiae* is controlled by its limited respiratory capacity: formulation and verification of a hypothesis. Biotechnol. Bioeng. 28, 927–937.

Steen, E.J., et al., 2008. Metabolic engineering of *Saccharomyces cerevisiae* for the production of n-butanol. Microb. Cell Fact. 7, 36.

Taylor, C.F., et al., 2007. The minimum information about a proteomics experiment (MIAPE). Nat. Biotechnol. 25, 887–893.

Usaite, R., et al., 2008. Characterization of global yeast quantitative proteome data generated from the wild-type and glucose repression *saccharomyces cerevisiae* strains: the comparison of two quantitative methods. J. Proteome Res. 7, 266–275.

Vemuri, G.N., Aristidou, A.A., 2005. Metabolic engineering in the -omics era: elucidating and modulating regulatory networks. Microbiol. Mol. Biol. Rev. 69, 197–216.

Vemuri, G.N., et al., 2007. Increasing NADH oxidation reduces overflow metabolism in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. U.S.A. 104, 2402–2407.

Verma, M., et al., 2003. Quantitative analysis of GAL genetic switch of *Saccharomyces cerevisiae* reveals that nucleocytoplasmic shuttling of Gal80p results in a highly sensitive response to galactose. J. Biol. Chem. 278, 48764–48769.

Washburn, M.P., et al., 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. 19, 242–247.

Wattanachaisaereekul, S., et al., 2008. Production of the polyketide 6-MSA in yeast engineered for increased malonyl-CoA supply. Metab. Eng. 10, 246–254.

Wiechert, W., et al., 2001. A universal framework for 13C metabolic flux analysis. Metab. Eng. 3, 265–283.
Yamano, S., et al., 1994. Metabolic engineering for production of beta-carotene and lycopene in *Saccharomyces cerevisiae*. Biosci. Biotechnol. Biochem. 58, 1112–1114.
Yarger, J.G., et al., 1984. Regulation of galactokinase (GAL1) enzyme accumulation in *Saccharomyces cerevisiae*. Mol. Cell. Biochem. 61, 173–182.
Yu, R.C., et al., 2008. Negative feedback that improves information transmission in yeast signalling. Nature 456, 755–761.