Fifteen years of large scale metabolic modeling of yeast: Developments and impacts

Tobias Österlund, Intawat Nookaew, Jens Nielsen

Department of Chemical and Biological Engineering, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden

Abstract

Since the first large-scale reconstruction of the Saccharomyces cerevisiae metabolic network 15 years ago the development of yeast metabolic models has progressed rapidly, resulting in no less than nine different yeast genome-scale metabolic models. Here we review the historical development of large-scale mathematical modeling of yeast metabolism and the growing scope and impact of applications of these models in four different areas: as guide for metabolic engineering and strain improvement, as a tool for biological interpretation and discovery, applications of novel computational framework and for evolutionary studies.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

The yeast Saccharomyces cerevisiae serves as an important cell factory in biotech production of food, beer, wine, nutraceuticals, pharmaceuticals, chemicals and fuels. It is also a very important model organism for eukaryal biology as it has a number of features that are conserved with higher eukaryotes, including humans. Its genome was among the first to be completely sequenced (Cherry et al., 1997; Goffeau et al., 1996) and many functional genomics tools have been pioneered using this yeast as a model organism (Chien et al., 1991; Winzeler et al., 1999; Wodicka et al., 1997). Thus there are comprehensive databases available, including the highly structured Saccharomyces Genome Database (SGD) (www.yeastgenome.org) (Weng et al., 2003). Many different yeast strains have been sequenced with the objective to understand evolution towards different kinds of applications (Borman et al., 2008; Legras et al., 2007; Rainieri et al., 2006), e.g. wine, bread and beer production, and for providing a basis for advancing metabolic engineering (Otero et al., 2011). With the advancement of systems biology, in particular for gaining new insight from high-throughput experimental data, S. cerevisiae has also played an important role (Mustacchi et al., 2006; Nielsen and Jewett, 2008). In this interface between experiments and mathematical modeling the concept of genome-scale metabolic models (Covert et al., 2001a) plays an important role, as it allows for direct integration of high-throughput experimental data with mathematical modeling, and hence advance our
understanding of the metabolism at the genome-scale. The first large-scale metabolic model for *S. cerevisiae* was published in 1995 and the first genome-scale metabolic model was published in 2003. Since then 8 genome-scale metabolic models of *S. cerevisiae* have been published, and these models have been extensively employed as a versatile tool for better understanding yeast physiology as well as advance biotechnological applications. Here we review the development in large-scale metabolic modeling of yeast and provide an overview of the large number of different applications of these models.

2. Framework for reconstructing genome-scale metabolic models

The concept of genome-scale metabolic reconstruction was first developed for *Escherichia coli* (Edwards and Palsson, 2000), and several extensive reviews describing the reconstruction are available (Covert et al., 2001a; Feist et al., 2008; Liu et al., 2010; Oberhardt et al., 2009; Thiele and Palsson, 2010). We will therefore here only summarize the conceptual framework to achieve a high quality genome-scale metabolic network (Fig. 1).

### 2.1. Metabolic network reconstruction

Metabolic network reconstruction is an iterative process where first draft metabolic network is established based on genomic information, i.e. which genes are available in the genome encoding enzymes catalyzing specific biochemical reactions. Following several rounds of curation of the network using a priori knowledge of biochemistry and physiology a comprehensive network compiling information about all possible biochemical conversion processes occurring in the cell will become available. This network can then form the basis for metabolic simulations, e.g. of growth of the cells on different carbon sources and formation of specific metabolic products, and based on comparison with experimental data more information may have to be added, e.g. specific transport reactions or fitting of specific energetic parameters. Following this, intensive refinements and modifications based on simulations and computational tools as well as updated physiological and biochemical knowledge are required to obtain a high-quality genome-scale metabolic network.

The metabolic network is normally built up from the set of metabolic reactions inside the cell and the enzymes catalyzing each reaction. Each reaction has a reference in the literature and a coupling to one or more genes, enzymes and EC-numbers. The metabolic network is converted to a mathematical model which can be analyzed with for example Flux Balance Analysis (FBA) or topological analysis or essentiality analysis. Different methods for gap-filling and debugging exist to improve the model. The model is validated by comparing the results from simulations to experimental data, for example comparing in silico growth predictions for deletion mutants with growth phenotypes for in vivo experiments. The model is continuously improved until a good agreement with experimental data and a high connectivity are obtained.

**Fig. 1.** A framework for the construction of a high-quality genome-scale metabolic model. The first step is to generate a draft metabolic network. For this multiple data types are used, i.e. genomic information, data from physiological characterization, biochemical knowledge etc. The metabolic network consists of metabolites connected through metabolic reactions. Each reaction has a reference in the literature and a coupling to one or more genes, enzymes and EC-numbers. The metabolic network is converted to a mathematical model which can be analyzed with for example Flux Balance Analysis (FBA) or topological analysis or essentiality analysis. Different methods for gap-filling and debugging exist to improve the model. The model is validated by comparing the results from simulations to experimental data, for example comparing in silico growth predictions for deletion mutants with growth phenotypes for in vivo experiments. The model is continuously improved until a good agreement with experimental data and a high connectivity are obtained.
reaction over participated metabolites. The information about the metabolite pathways and reactions are collected from multiple types of data sources. The initiation step for finding the set of species specific metabolic reactions is directly derived from the functional annotations of open reading frames (ORFs) through genomic approaches. For S. cerevisiae, high quality genome annotations are represented in organism-specific databases, such as the Saccharomyces genome database (SGD) (Weng et al., 2003) and the Comprehensive yeast genome database (CYGD) (Guldener et al., 2005). Biochemical evidence including functional characterization of enzymes through extensive studies is another important source. For model organisms such as S. cerevisiae, there is extensive biochemical information available and the metabolism is well-characterized. Legacy data is also collected (bibliometrics) to find more information about the biochemical related evidences that happen inside the cell. A high-quality metabolic reconstruction can be viewed as an extensive, manually curated database or collection of metabolic reactions where most of the reactions have some support by evidences from biochemical or physiological studies in the literature. Each reaction is linked to an enzyme commission (EC) number and has also a gene–protein–reaction (GPR) association describing which genes or gene complexes those are responsible for each reaction.

The complexity of organelle organization of eukaryotic organisms is an obstacle for metabolic reconstruction. However, the information about cellular localization of proteins provided either by an in vivo approach e.g. Yeast protein localization database (YPL) (Kals et al., 2005) or in silico approach e.g. PSORT (Nakai and Horton, 1999) and CELLO (Yu et al., 2006) is also important in order to be able to describe multiple compartments (organelles) in the model within agreement with yeast's nature. Moreover, transport reactions, i.e. exchange of metabolites between the different compartments, are also needed to increase the degree of connectivity within the reconstructed genome-scale metabolic network.

2.2. Mathematical formulation and debugging

Once a high quality metabolic network has been reconstructed detailed adjustments of the network are needed including elemental and charge balancing for all participated reactions as well as estimation of directionality of the reactions that are lacking biochemical evidence by thermodynamic approaches (Jankowski et al., 2008). Then, transformation of the qualitative network to the mathematical space is the next procedure. The set of reactions and their participated metabolites are converted to a stoichiometric matrix S with the rows representing the metabolites and the columns representing the reactions by metabolite balancing under steady state assumption. Simulation of the metabolic network using the Flux Balance Analysis (FBA) approach (Price et al., 2004; Varma and Palsson, 1994b) is a successful method for description and prediction of the cellular metabolism as well as metabolic phenotypes. Therefore, collection of constraints and objective functions is required in the mathematical formulation. The objective function can be formulated following different biological hypotheses that can infer the properties of the reconstructed network such as specific metabolite formation, energy production, nutrient assimilation and biomass formation, which has been accepted to be an excellent objective function to describe the yeast habitat (Famili et al., 2003; Nookaew et al., 2008). The formulation of biomass formation can be achieved from experimental measurements of macromolecules such as nucleotides, amino acids, carbohydrates, lipids, etc. Other simulation methods, e.g. based on network topology, such as extreme pathways and elementary modes have also been used to find the set of steady state solutions that can describe all possible properties of reconstructed networks from the stoichiometric matrix (Schilling et al., 2000; Schuster et al., 2000). All these mathematical frameworks can also be used to infer missing reactions and genes through gap filling in order to improve the connectivity of the network. The two types of missing information that can exist in the draft metabolic network are so-called “dead ends”, where a reaction consuming a specific metabolite is missing and “orphans” where the reaction is known to exist, but the gene coding for the enzyme catalyzing the reaction is unknown. Different methods and algorithms for gap filling in metabolic networks have recently been reviewed (Orth and Palsson, 2010).

2.3. Validation with experimental data

The mathematical model can be validated in several ways by comparing in silico simulation results with data from in vivo experiments. For example, the growth predicted by the model under different growth conditions can be compared to the experimental growth phenotypes from the same conditions. Another common way to validate the model is to perform both in silico and in vivo gene deletions and investigate if the mutant can grow (Duarte et al., 2004a; Förster et al., 2003b; Nookaew et al., 2008; Zomorrodi and Maranas, 2010). A recent study provides the in vivo growth phenotypes for about 5.4 million double gene knockouts in yeast (Costanzo et al., 2010) which can be used for validation of the ability of the genome-scale model to predict the phenotypes of double gene deletions. If the simulation results are not in agreement with the experimental results the model has to be improved using the tools for debugging and simulation described in Section 2.2. The results from simulations can also be compared to different types of omics data including transcriptomics, metabolomics and proteomics to check the consistency of the results (Herrgard et al., 2006).

3. Development of yeast metabolic models

Fig. 2 provides a historical summary of large-scale yeast metabolic modeling since 1995. The first comprehensive model of the S. cerevisiae metabolism was published 16 years ago (Van Gulik and Heijnen, 1995) and this model contained 88 metabolic reactions. Shortly following this model two other comprehensive metabolic models were published (Nissen et al., 1997; Vanrolleghem et al., 1996). The major applications of these earlier models were to perform metabolic flux analysis (MFA), i.e. to quantify the activity of the different pathways in the metabolic network. A major advancement in metabolic modeling of yeast occurred in 2003 when the first genome-scale metabolic model of yeast was published (Förster et al., 2003a), and following this several other genome-scale metabolic models have been published. In the following we will review this historical development in metabolic modeling of yeast.

3.1. Models of central carbon metabolism

Understanding the metabolism of baker's yeast by mathematical modeling started in the late 1990s. In that time, the species specific metabolic network of S. cerevisiae was difficult to obtain. The central carbon metabolism for E. coli (Varma et al., 1993) and Corynebacterium glutamicum (Vallino and Stephanopoulos, 1994) was constructed and formulated as a stoichiometric model for flux balance analysis modeling. Van Gulik and Heijnen (1995) then reconstructed central carbon metabolic networks in the same manner for two different yeast species, S. cerevisiae and Candida utilis. The networks were constructed relying on the current available information about the biochemical pathways derived from literature evidences in that time. For S. cerevisiae, six different models were constructed using different growth conditions, e.g. anaerobic growth on glucose and aerobic growth on glucose, ethanol and acetate. This strategy resulted in a model of 88 reactions and 81 metabolites that allowed description of growth on a mixed substrate of glucose and ethanol. Two different parameters namely the P/O ratio and a growth-related maintenance coefficient were introduced into the models for simulation purposes. In a later study (Vanrolleghem et al., 1996) further investigation of experimental data was used to validate the network. The model was
also improved to be able to describe the growth on a mixed substrate of glucose and ethanol. Another metabolic network model, primarily aimed for simulation of anaerobic growth on glucose, was reconstructed by Nissen et al. (1997). This model contained 37 reactions and 27 intracellular metabolites. The model was employed for metabolic flux analysis to estimate the carbon channeling by the intracellular reactions.

The yeast prototype metabolic models were used and adapted for many different types of studies during the coming years through stoichiometric modeling approaches. Jin et al. (1997) applied metabolic modeling for heterologous protein production by a recombinant yeast strain, and these models were also used for identification of targets for improving ethanol production through metabolic engineering (Nissen et al., 2000). Metabolic networks were also used together with 13C-labeling experiments as a successful tool to gain insight into the yeast metabolism (Gombert et al., 2001; Maaheimo et al., 2001). Carlson et al. (2002) reconstructed a metabolic model for yeast and used topological analysis to calculate the elementary flux modes (EFMs) of the system for poly-beta-hydroxybutyrate (PHB) production, and Förster et al. (2002) combined metabolomic data with metabolic network modeling to infer the function of some unannotated genes (orphans).

The development of prototype metabolic networks with focus on the central carbon metabolism of yeast and their applicability provides a solid basis for development of comprehensive metabolic models following the availability of the yeast genome sequence.

### 3.2. Genome-scale metabolic models

The first comprehensive reconstruction of the yeast metabolism, which also is the first genome-scale model for a eukaryotic organism, was a joint effort from the groups of Nielsen and Palsson (Förster et al., 2003a). The model consists of 708 metabolic genes associated with 1175 reactions and 733 metabolites. Three cellular compartments, namely cytosol, mitochondria and the extracellular space were included. The physiological validations and predictions by the model using flux balance analysis (FBA) were shown to have good agreements with many experimental datasets (Famili et al., 2003). Furthermore, large-scale in silico gene deletion analysis of the model showed high accuracy in predicting gene essentiality when comparing to in vivo knock-out phenotypes (Förster et al., 2003b). The first genome-scale model was named iFF708 where FF stands for Förster and Famili, which are the names of the model creators and 708 is the number of ORFs included in the model. After iFF708 had been released...
8 other genome-scale metabolic models of *S. cerevisiae*, which all differ in scope and objective, were developed originating from the iFF708 model. The historical development of genome-scale metabolic models for yeast and their relationship is summarized as blue boxes in Fig. 2.

The iND750 model (Duarte et al., 2004a) includes 750 ORFs and the number of compartments included in the model was expanded to 8 (also including nucleus, endoplasmic reticulum, golgi, vacuole and peroxisome). All reactions in the iND750 model were also charge balanced. The biggest difference between the iND750 model and the iFF708 model is the number of unique reactions that increased to 1149 compared to 842 in iFF708 due to the introduction of several more compartments. This resulted in a model that could describe the current biological knowledge in a considerable way, however, the ability of correctly predicting growth phenotypes for single mutants was reduced when compared to the iFF708 model. In addition, the iND750 model was also used for so called phenotype phase-plane analysis where the metabolic states can be projected on different levels of glucose and oxygen uptake rates (Duarte et al., 2004b). A modification of the iND750 model called iHF732 reducing the number of compartments to two (cytoplasm and extracellular space) was used for dynamic FBA simulations by Hjersted and Henson (2009).

Since there are some states of metabolism that cannot be described only by metabolic network simulations the regulation mechanisms of the metabolic genes could also be taken into consideration in order to fully describe the physiological behavior of the cell in these conditions. In the study by Herrgård et al. (2006), the iND750 model was overlaid with a regulatory network derived from 55 key transcription factor genes related with carbon and nitrogen metabolism. The reconstruction was done by generation of mathematical rules that can explain the transcription factor’s activation or repression of metabolic genes with evidence from the literature. Following Boolean formalism, transcriptional regulation events can be calculated simultaneously with the genome-scale metabolic model as additional constraints. The resulting model called iMH805/775 demonstrated the capability to predict the changes in gene expression of some transcription factor deletion strains using the regulated flux balance approach (rFBA) (Covert et al., 2001b).

The first genome-scale models contained several reactions that were not connected to the network as well as dead-end reactions. Therefore Sauer et al. (Blank et al., 2005; Kuepfer et al., 2005) presented a refined version of iFF708 called iLL672. The resulting model consists of 636 metabolites and 1038 reactions associated with 672 metabolic genes (less than the iFF708 model). The model was demonstrated to have a higher ability to correctly predict growth phenotypes for single deletion mutants due to the removal of unlinked reactions from the network. One of the most complex part of the metabolism, lipid metabolism, which is an important part of the various biochemical processes going on in yeast (Nielsen, 2009) was poorly described in the original genome-scale metabolic models. Therefore Nookaew et al. (2008) published a new model by expanding the iFF708 model and included 143 new reactions associated with 65 ORFs involved in lipid metabolism. The model named iMM904 also had an improved biomass equation, involving details in the formation of fatty acids and lipids for carbon- and nitrogen-limited growth conditions. Following this, another model called iMM904 was constructed using the iND750 model as a template, consequently containing 8 different compartments (Mo et al., 2009). The model was used to correlate the concentrations of metabolites, which were measured by metabolomic approaches, with the intracellular fluxes by random sampling of the metabolic network (Schellenberger and Palsson, 2009). Recently, Zomorrodi and Maranas (2010) proposed a revised version of the iMM904 model called iA2900. They proposed 120 revisions to the model based on synthetic lethality/gene essentiality analysis, in order to increase the prediction power.

So far all the different metabolic network reconstructions had slightly different naming of the metabolites and reactions because the development of the different reconstructions was performed independently of each other, leading to inconsistent representations of some parts of the metabolism. This disagreement made the direct comparison and mapping between the models difficult. To solve this problem, standardization of the components in the metabolic network was needed and had to be organized in a consensus manner. Therefore several research groups in the field of Systems Biology gathered for a jamboree to construct a consensus genome-scale metabolic network of yeast (Herrgard et al., 2008). The resulting metabolic network is here referred to as Yeast 1.0 and contains 857 genes, 1857 reactions and 1168 metabolites. Each metabolite in the network was annotated by established nomenclature conventions including the international chemical identifier (InChi) codes and KEGG identifiers. This standardization enables the linking of metabolic networks with metabolite data from mass spectrometry that is very useful for metabolomic studies. A consistent representation of metabolic models by Systems Biology Markup Language (SBML) format was suggested following the MIRIAM standards (minimum information requested in the annotation of biochemical models) (Hucka et al., 2003; Le Novere et al., 2005). This standardization is very helpful for model exchanges within the scientific community. The consensus metabolic network was not suitable for computational analysis because of missing information about reversibility of some reactions and also parts of the network that were not connected. To overcome this weak point Dobson et al. (2010) presented an updated model based on the consensus model called Yeast 4.0. The lipid metabolism which was weakly covered in Yeast 1.0 was improved following the basis of the iIN800 model. This model had also improved connectivity and provided a mathematical formulation for constraint-based simulations. The biomass equation used for simulations by the Yeast 4.0 model was adopted from iIN800. In total the Yeast 4.0 includes 932 genes associated with 1865 reactions and 1319 metabolites and is bigger in scope than any of the previous reconstructions. The ultimate goal of genome-scale metabolic network reconstruction in the future is to have a well-annotated network including all parts of the metabolism without any missing reactions or gaps; however it is not yet possible due to incomplete knowledge of the yeast metabolism.

4. Applications

There are many different applications and uses for genome-scale metabolic models that can be found in the literature. Here we classify the applications of genome-scale metabolic models into four major application categories as shown in Fig. 3: i) guidance for metabolic engineering and strain improvement, ii) biological interpretation and discovery, iii) applications of novel computational framework, and iv) evolutionary elucidation. An overview of 55 publications applying yeast genome-scale metabolic models in one or more of these four categories is shown in Fig. 4.

4.1. Guidance for metabolic engineering and strain improvement

Genome-scale metabolic models can be used as a versatile tool for in silico metabolic engineering (Nielsen and Jewett, 2008; Patil et al., 2004). Computational methods have been applied to analyze the genome-scale metabolic model for evaluation of the metabolite production potential and for providing the maximum theoretical yield for a specific metabolite under different growth conditions. Moreover, the approach can also suggest targets for over-expression or deletion to improve the ability of a strain to produce the desired metabolites. The most applied computational method for this purpose is FBA and for microorganisms such as *S. cerevisiae* the production of biomass is a reasonable objective function (Famili et al., 2003). Nevertheless, to obtain a knockout strategy for improved production of a specific product, the MOMA algorithm (minimization of metabolic adjustment) was developed (Segre et al., 2002), where an engineering strategy that is coupled to maximum growth of the cell is identified under the
assumption that the knock-out strain should have a minimal change in flux distribution compared to the wild type strain.

In *S. cerevisiae*, metabolic engineering strategies aided by genome-scale metabolic modeling have lead to improved production of for example ethanol (Bro et al., 2006; Dikicioglu et al., 2008; Hjersted and Henson, 2009), purine (Burgard et al., 2004), proline/pyrimidines (Bundy et al., 2007), vanillin (Brochado et al., 2010) and sesquiterpenes (Asadollahi et al., 2009). Wang and Hatzimanikatis (2006) studied the biomass production in batch and chemostat fermentations. Yeast can be used as a cell factory both for production of native yeast compounds but also heterologous products by the introduction of recombinant pathways. The genome-scale metabolic model can be a useful tool to predict the performance of novel pathways, and below we discuss a few selected studies that utilize genome-scale metabolic models for strain improvement and metabolic engineering.

Bro et al. (2006) used in silico simulations with the iFF708 model as aid for increasing ethanol production and at the same time decrease the amount of glycerol produced by the cell under anaerobic growth conditions. The best strategy suggested by the simulations was to introduce a heterologous enzyme, gapN, and this strategy was used to construct an in vivo strain by genetic engineering. The engineered strain had a decreased glycerol production of 40% and an increased ethanol yield of 3%. Another strategy for ethanol production improvement was found in the study by Hjersted and Henson (2009) who used a different approach, also aided by genome-scale metabolic models, where ethanol production in batch culture is modeled both with steady-state FBA and dynamic FBA. For steady state simulations the prediction power was compared for three different models, iFF708, INDB750 and a decompartmentalized version of INDB750 used in this study called iHH732. They also used the dynamic FBA approach (Mahadevan et al., 2002) to model combined aerobic and anaerobic growth and the shift between the two phases. Dynamic FBA can account for dynamic changes in for example batch fermentations. The concentrations of the internal metabolites in the model are assumed to be in pseudo-steady state for each time point, and the external fluxes such as the glucose and oxygen uptake rates change over time.

In another study (Brochado et al., 2010) a recombinant yeast strain producing vanillin was constructed by introducing a heterologous enzyme from Arabidopsis thaliana. This strain was modeled in silico to find deletion targets that could improve the vanillin yield on glucose. The iFF708 model was used with MOMA simulations to suggest two different deletion targets. Successfully, when introducing the gene deletions in vivo, a 5-fold increase of vanillin yield was observed as compared to previously reported vanillin production in *S. cerevisiae*.

It is clear that in silico modeling using genome-scale metabolic models can accelerate the process of metabolic engineering by suggesting rational targets for over-expression or deletion for improved production of a certain metabolite, or computationally test different strategies against each other to find a candidate strategy that can be tested experimentally.

### 4.2. Biological interpretation and discovery

One prominent application for genome-scale metabolic models is to serve as a tool for biological interpretation and discovery through integrated analysis. A genome-scale metabolic model is a comprehensive and well-annotated knowledge-base of the metabolism and biochemical pathways going on inside the cell. This enrichment enables data integration and examination of physiology in various perturbations of *S. cerevisiae*. There are several examples where yeast genome-scale metabolic models have been used as a tool for biological interpretation.

Many studies employ genome-scale metabolic networks as a tool to perform integrated analysis of multilevel omics data, for example...
transcriptome and metabolome. The analysis enables to obtain better insight into the physiological adaptations of metabolism that are related to different perturbations or environmental conditions (Chechik et al., 2008; Cimini et al., 2009; Fu, 2009; Nacher et al., 2006; Papini et al., 2010; Rossouw et al., 2009; Ruenwai et al., 2011; Usaitė et al., 2006). Patil and Nielsen (2005) developed an algorithm called “Reporter metabolites” which integrates transcript data into metabolic models using the topological structure by identifying the metabolites in the metabolic network with the biggest transcriptional response among its neighboring enzymes. Oliveira et al. (2008) extended the reporter algorithm to also include other biological networks such as transcription factor interactions. Another study (Snitkin et al., 2008) investigated the growth phenotypes of 465 deletion mutants and integrated the experimental results into three yeast genome-scale metabolic models, iFF708, iND750 and iLL672 by FBA and MOMA simulations. The results suggested that simulations with genome-scale metabolic models can be used both as a tool for quality control of the experimental data, and for hypothesis testing. Another phenomenon that is studied in yeast is respiration versus fermentation. Yeast naturally has the capability to produce ethanol in anaerobic conditions or with excess of glucose in aerobic conditions, but this behavior is normally not captured by a FBA solution. In a recent study (Simeonidis et al., 2010) an extra energy cost was introduced, representing the synthesis of mitochondria during respiration, to improve the agreement of simulation results for respiration and fermentation as compared to experimental observations. Another study on a similar topic (Teh and Lutz, 2010) used thermodynamics to study the behavior of yeast fermentation in anaerobic conditions and ethanol production. They found that the fermentation process is very efficient; the thermodynamic efficiency of ethanol production in yeast under anaerobic conditions is around 90%.

Genome-scale metabolic models can also serve as a platform for how different components in yeast interact. Several studies use the connectivity between metabolic genes and simulations with genome-scale metabolic models to find co-expression patterns between genes (Kharchenko et al., 2005), gene essentiality (Del Rio et al., 2009), functional similarities of metabolic genes, (Rokhlenko et al., 2007) nutrient–gene interactions (Diamant et al., 2009), protein–protein interactions (Aho et al., 2010), epistatic interactions (Segre et al., 2005) and condition-dependent gene annotations (Shlomi et al., 2007).

The use of metabolic networks and models as a platform for data integration for novel discovery is growing in scope and also as the genome-scale metabolic models become more comprehensive, including more mechanistic details of components in the genome-scale metabolic model such as metabolites, reactions and pathways. The enriched information combined with computational methods enables the application of data driven discovery via an integrated analysis framework of genome-scale models for a better understanding of the complexity of life.

### 4.3. Applications of novel computational framework

The development of genome-scale metabolic models has also accelerated the number of published computational methods and framework for in silico analysis of metabolic systems. Many of them have been applied on the genome-scale metabolic models for S. cerevisiae. Here some examples will be described to show how reconstructed genome-scale metabolic networks have been widely applied for computational analysis and development of novel algorithms.

The analysis of metabolic fluxes can be performed with flux balance analysis where the optimal solution for a given objective function is obtained using linear programming methods (Price et al., 2004). Several methods for identification of flux distributions inside the cell that are based on FBA has been presented e.g. MOMA (Segre et al., 2002), ROOM (regulatory on/off minimization of metabolic flux) (Shlomi et al., 2005) and regulatory FBA (rFBA) (Covert et al., 2001b). Lee et al. (2008) present a FBA-based method for integrating signaling and regulatory networks into the metabolic network and two other studies (Acencio and Lemke, 2009; Whelan and King, 2008) developed new modeling frameworks to predict the growth of yeast and gene essentiality. Based on topological analysis Kaleta et al. (2009) developed a method to decompose the genome-scale model by a new algorithm called elementary flux patterns. Furthermore Patil et al. (2005) developed an evolutionary algorithm for finding a set of gene knock-outs that can improve production of a specific metabolite. Another study (Bordel et al., 2010) applied random sampling in the space of possible steady state solutions, defined by fixing some experimental metabolic fluxes to obtain a set of condition dependent fluxes. The estimated changes in fluxes for each reaction for 4 different conditions in yeast and 5 deletion mutants were compared to the changes in expression derived from microarray experiments. This framework can be used to suggest if changes in reaction rates between different conditions are due to transcriptional regulation, metabolic regulation or post-transcriptional regulation. Burgard et al. (2004) presented a method for identification of flux coupling in metabolic networks both for detecting blocked reactions and suggesting targets for genetic manipulations.
Genome-scale metabolic reconstructions do not normally incorporate information about kinetics, but are still very useful for steady-state simulations and constraint-based modeling. Nevertheless, parameterized kinetic modeling would represent the enzymatic metabolic reactions going on in the cell in a more detailed way. A recent study (Smallbone et al., 2010) is an example of an attempt to build a genome-scale kinetic model for yeast based on genome-scale metabolic reconstruction of yeast. The model is built from the Yeast 1.0 consensus network and used simplified linlog kinetics and parameters from kinetic model repositories. This can be seen as a starting point for expansion of static metabolic simulation to dynamic simulation.

Several methods have been developed for integration of different data types into genome-scale metabolic models, e.g. transcriptome data (Oliveira et al., 2008; Patil and Nielsen, 2005), metabolome data (Cakir et al., 2006) and diverse types of data (Cvijovic et al., 2010). Several methods have also been developed for helping the reconstruction process, for instance identification of reactions with stoichiometric inconsistencies (Georgygan et al., 2008) automatic assignment of genes to metabolic pathways based on expression data (Popescu and Yona, 2005) and network-based prediction of the localization of metabolic enzymes (Mintz-Oron et al., 2009). Several computational tools have been developed for analysis and assessment of genome-scale metabolic models. One important area is the development of methods for data integration and correlation between experiments and simulations.

4.4. Evolutionary elucidation

One common way to study the evolutionary relationship between different species is to investigate the degree of conservation at the gene or protein level based on sequence similarity. The metabolic models and other biological networks which also contain genomic information can be used together with evolutionary studies to gain even more insights into the relationships between species, e.g. by identification of whole pathways that are conserved or missing.

The functions of duplicate genes, i.e. genes with the same function due to gene duplication events (paralogs) have been studied in yeast using genome-scale metabolic models and flux balance analysis. Kuepf et al. (2005) used their model iLL672 to classify families of duplicate genes in S. cerevisiae. They found, using simulations, that the presence of redundant enzymes is not more common for essential genes than for non-essential genes. Mahadevan and Lovley (2008) compared FBA simulations of yeast with simulations of three bacteria and one archaea to elucidate the redundancy of the gene functions in different species is to investigate the degree of conservation at the gene or protein level based on sequence similarity. The metabolic models and other biological networks which also contain genomic information can be used together with evolutionary studies to gain even more insights into the relationships between species, e.g. by identification of whole pathways that are conserved or missing.

The functions of duplicate genes, i.e. genes with the same function due to gene duplication events (paralogs) have been studied in yeast using genome-scale metabolic models and flux balance analysis. Kuepf et al. (2005) used their model iLL672 to classify families of duplicate genes in S. cerevisiae. They found, using simulations, that the presence of redundant enzymes is not more common for essential genes than for non-essential genes. Mahadevan and Lovley (2008) compared FBA simulations of yeast with simulations of three bacteria and one archaea to elucidate the redundancy of the gene functions in different species is to investigate the degree of conservation at the gene or protein level based on sequence similarity. The metabolic models and other biological networks which also contain genomic information can be used together with evolutionary studies to gain even more insights into the relationships between species, e.g. by identification of whole pathways that are conserved or missing.

The functions of duplicate genes, i.e. genes with the same function due to gene duplication events (paralogs) have been studied in yeast using genome-scale metabolic models and flux balance analysis. Kuepf et al. (2005) used their model iLL672 to classify families of duplicate genes in S. cerevisiae. They found, using simulations, that the presence of redundant enzymes is not more common for essential genes than for non-essential genes. Mahadevan and Lovley (2008) compared FBA simulations of yeast with simulations of three bacteria and one archaea to elucidate the redundancy of the gene functions in different species is to investigate the degree of conservation at the gene or protein level based on sequence similarity. The metabolic models and other biological networks which also contain genomic information can be used together with evolutionary studies to gain even more insights into the relationships between species, e.g. by identification of whole pathways that are conserved or missing.

The functions of duplicate genes, i.e. genes with the same function due to gene duplication events (paralogs) have been studied in yeast using genome-scale metabolic models and flux balance analysis. Kuepf et al. (2005) used their model iLL672 to classify families of duplicate genes in S. cerevisiae. They found, using simulations, that the presence of redundant enzymes is not more common for essential genes than for non-essential genes. Mahadevan and Lovley (2008) compared FBA simulations of yeast with simulations of three bacteria and one archaea to elucidate the redundancy of the gene functions in different species is to investigate the degree of conservation at the gene or protein level based on sequence similarity. The metabolic models and other biological networks which also contain genomic information can be used together with evolutionary studies to gain even more insights into the relationships between species, e.g. by identification of whole pathways that are conserved or missing.

The functions of duplicate genes, i.e. genes with the same function due to gene duplication events (paralogs) have been studied in yeast using genome-scale metabolic models and flux balance analysis. Kuepf et al. (2005) used their model iLL672 to classify families of duplicate genes in S. cerevisiae. They found, using simulations, that the presence of redundant enzymes is not more common for essential genes than for non-essential genes. Mahadevan and Lovley (2008) compared FBA simulations of yeast with simulations of three bacteria and one archaea to elucidate the redundancy of the gene functions in different species is to investigate the degree of conservation at the gene or protein level based on sequence similarity. The metabolic models and other biological networks which also contain genomic information can be used together with evolutionary studies to gain even more insights into the relationships between species, e.g. by identification of whole pathways that are conserved or missing.

The functions of duplicate genes, i.e. genes with the same function due to gene duplication events (paralogs) have been studied in yeast using genome-scale metabolic models and flux balance analysis. Kuepf et al. (2005) used their model iLL672 to classify families of duplicate genes in S. cerevisiae. They found, using simulations, that the presence of redundant enzymes is not more common for essential genes than for non-essential genes. Mahadevan and Lovley (2008) compared FBA simulations of yeast with simulations of three bacteria and one archaea to elucidate the redundancy of the gene functions in different species is to investigate the degree of conservation at the gene or protein level based on sequence similarity. The metabolic models and other biological networks which also contain genomic information can be used together with evolutionary studies to gain even more insights into the relationships between species, e.g. by identification of whole pathways that are conserved or missing.

The functions of duplicate genes, i.e. genes with the same function due to gene duplication events (paralogs) have been studied in yeast using genome-scale metabolic models and flux balance analysis. Kuepf et al. (2005) used their model iLL672 to classify families of duplicate genes in S. cerevisiae. They found, using simulations, that the presence of redundant enzymes is not more common for essential genes than for non-essential genes. Mahadevan and Lovley (2008) compared FBA simulations of yeast with simulations of three bacteria and one archaea to elucidate the redundancy of the gene functions in different species is to investigate the degree of conservation at the gene or protein level based on sequence similarity. The metabolic models and other biological networks which also contain genomic information can be used together with evolutionary studies to gain even more insights into the relationships between species, e.g. by identification of whole pathways that are conserved or missing.


