Received Date: 12-Oct-2011

Revised Date : 18-Nov-2011

Accepted Date: 19-Nov-2011

Article type : MiniReview

Editor : Jens Nielsen

Corresponding author e-mail id: vassily.hatzimanikatis@epfl.ch

From network models to network responses: Integration of thermodynamic and kinetic properties of yeast genome-scale metabolic networks

Keng Cher Soh[1], Ljubisa Miskovic[1] and Vassily Hatzimanikatis[1*]

[1]Laboratory of Computational Systems Biotechnology, Ecole Polytechnique Fédérale de Lausanne, Station 6, CH-1015, Lausanne, Switzerland

* - Corresponding author

ABSTRACT

Many important problems in cell biology emerge from the dense nonlinear interactions between functional modules. The importance of mathematical modeling and computer simulation in understanding cellular processes is now indisputable and widely appreciated, and genome-scale metabolic models have gained much popularity and utility in helping us to understand and test hypotheses about these complex networks. However, there are some *caveat utilitor* that come with the use and interpretation of different types of metabolic models that we aim to highlight here. We discuss and illustrate how the integration of thermodynamic and kinetic properties of the yeast metabolic networks in network analyses can help in understanding and utilizing this organism more successfully in the areas of metabolic engineering, synthetic biology, and disease treatment.

INTRODUCTION

Yeast has been used for the production of food, beverages and ingredients, fuels, chemicals, and pharmaceutical proteins. The long history of yeast in the development of diverse bioprocesses have led to the accumulation of a wealth of data about its physiology, biochemistry, and regulation (Walker, 1998). These studies have established yeast as a robust industrial organism (Petranovic & Vemuri, 2009) and with the current renaissance in pushing for the production of fuels and chemicals from renewable resources, the interest in yeast has exploded and we expect it will grow stronger. A deeper understanding of its cellular physiology and metabolism can help us to better engineer *S. cerevisiae* to improve the efficiency of production of heterologous products (Bro*, et al.*, 2003, van Maris*, et al.*, 2007, Wattanachaisaereekul*, et al.*, 2007, Wisselink*, et al.*, 2009, Zelle*, et al.*, 2010).

In addition, in the last 25 years, yeast has been used extensively as a model system for the study of cell cycle and its connection to cancer (Hartwell, 2004). The similarities of carbon, energy and lipid metabolism between yeast and humans have also made yeast an excellent model system of choice for the study of the role of metabolism in disease etiology and treatment (Petranovic & Nielsen, 2008, Nielsen, 2009, Bolotin-Fukuhara*, et al.*, 2010).

The emergence of functional genomics and systems biology has opened new perspectives for the analysis and the study of biological organisms, and yeast was one of the first organisms to be studied during the development of these technologies. Nielsen and colleagues have reviewed and discussed the development and uses of these technologies in yeast research and development (Jewett*, et al.*, 2005, Nielsen & Jewett, 2008). One of the key messages from these reviews is the importance of metabolic fluxes, as the final outcome of intricate interactions between the different networks; the networks of transcription, translation, post-translational modification, signal transduction and protein-protein interaction.

Network analysis has been a major effort in the area of biological sciences (Albert & Barabasi, 2002, Newman, 2003, Barabasi & Oltvai, 2004, Papin*, et al.*, 2004, Joyce & Palsson, 2006, Feist*, et al.*, 2009). Technologies that emerged from the progress in genomics have allowed the experimental identification and verification of interactions between genes and their products, from proteins to metabolites to integrated phenotypes, and a wealth of computational methods has been developed, and is continuously developing, for the integration of this information into networks and their analysis. The ultimate goal of these methods is to synthesize the knowledge into predictive

mathematical models that can be used in computational analyses to provide insight and accelerate discovery.

Although it is acknowledged that it is difficult to classify mathematical models in systems biology, two main classes are mainly considered (Nielsen & Jewett, 2008): top-down models, where new biological information is extracted from large-data sets and the analysis used there is mainly inductive (Kell, 2005, Joyce & Palsson, 2006, Ananiadou*, et al.*, 2010), and bottom-up models, which are built on detailed mechanistic knowledge and the analysis is deductive but is limited to small networks (Rieger*, et al.*, 2005). And herein lies the major challenge in systems biology, the ability to build models with the mechanistic quality of the bottom-up models and the scale (i.e., number of components and interactions) of the top-down models (Papin*, et al.*, 2004, Mehra & Hatzimanikatis, 2006).

Nielsen and Jewett observed that although it is difficult to reconcile bottom-up and top-down modeling, the efforts in curating and building metabolic network models has been close to achieving this (Nielsen & Jewett, 2008). The combined knowledge of physiology, biochemistry and metabolism allow the reconstruction of networks, which are further curated using flux balance analysis (FBA) to complete missing parts and derive a functional metabolic network. The continuing integration of knowledge about the networks that regulate the activity of the metabolic activities is the first successful demonstration of bridging the gaps between the functional regulatory networks (Papin*, et al.*, 2004, Joyce & Palsson, 2006, Hyduke & Palsson, 2010, Schellenberger*, et al.*, 2010, Hasunuma*, et al.*, 2011).

This remarkable progress in the area of metabolic modeling is very good news for those working in yeast metabolism, physiology, and bioprocessing. The systems-level, genome-scale understanding of carbon and energy metabolism is critical to enhance our metabolic toolbox for optimizing the production of industrial chemicals and fuels from yeast. Additionally, it can also help in the elucidation of the etiology of many metabolic diseases. Understanding carbon metabolism and energy management is also important for understanding and engineering stress tolerance, which is an unavoidable consequence from the bioprocess conditions (e.g., high titers and low pH) (Nicolaou*, et al.*, 2010), and cellular stress has also been implicated in many diseases (Costa & Moradas-Ferreira, 2001, Sorolla*, et al.*, 2008).

However, many important technological issues limit the full promise of useful applications of metabolic models in yeast. While we know very well the structure of the model, i.e., the biochemistry and many regulatory connections, we do not have enough global-scale data. Rather, we have partial data from proteomic, metabolomics, and physiology studies, and in many cases, uncritical analysis of partial data can lead to erroneous conclusions. Another problem arises from the complexity of biological systems and the large-scale, high-throughput nature of the data, as there are also differences in results that arise from the "same" studies in different laboratories, an important issue that has been recently acknowledged and addressed in various studies (Lin*, et al.*, 2009, Ukibe*, et al.*, 2009, Hong*, et al.*, 2010, Zhao*, et al.*, 2010).

In all the studies of networks in yeast, the information is mainly qualitative, i.e., network interactions are described as on/off properties, with very little information on the strength of these interactions. Hence, while these networks can be used to integrate and interpret quantitative observations, such as fluxes and expression data, they can only

simulate and predict experiments that disrupt the network connectivity, such as gene knock-out, loss-of-function mutations, and mutations of gene regulatory elements. The ultimate objective would be to formulate models that can both describe the steady-state behavior and predict the dynamic responses of yeast metabolic networks in order to guide how the system would behave to knock-in or knock-out of genes. This would allow us to manipulate the metabolic network to achieve our desired objectives. We will discuss here some of the approaches towards this objective and how integration of thermodynamic and kinetic properties can bring us closer to that aim.

SOME CONSIDERATIONS IN THE USE AND APPLICATIONS OF THE GENOME-SCALE METABOLIC MODEL OF YEAST

Genome-scale metabolic models have gained significant popularity as versatile tools in many studies (Feist & Palsson, 2008, Oberhardt*, et al.*, 2009) and they have proven to be valuable assets in guiding metabolic engineering decisions (Bro*, et al.*, 2003, Patil*, et al.*, 2004). With the development of high-throughput and automated reconstruction methods (DeJongh*, et al.*, 2007, Henry*, et al.*, 2010, Radrich*, et al.*, 2010), genome-scale metabolic reconstructions have been increasing with an accelerated pace, even though their number still lags behind that of number of genome sequences being completed.

The first genome-scale metabolic model of the yeast *Saccharomyces cerevisiae*, named *iFF708*, was published in 2003 (Famili*, et al.*, 2003, Forster*, et al.*, 2003). This model was subsequently modified through the inclusion of additional biochemical reactions, genes, regulatory constraints and compartments (Duarte*, et al.*, 2004, Kuepfer*, et al.*, 2005, Herrgard*, et al.*, 2008, Nookaew*, et al.*, 2008). Three laboratories, two of which have collaborated in the development of the original model, have led the main

developments of these models. Recently, a "consensus" model has been developed through a collaborative approach of a community of yeast researchers to serve as a resource for collecting and summarizing the current and growing knowledge of yeast metabolism (Herrgard*, et al.*, 2008).

A main use of the genome-scale models is the study of the physiology of gene deletions. Snitkin *et. al.* (Snitkin*, et al.*, 2008) compared model (*iFF708*) predictions against 465 gene deletion mutants under 16 conditions and found a high fraction of correct predictions (94%) that validated the high predictive capacity of the model and demonstrated how inconsistencies can also be used to drive further hypothesis testing. What is interesting in this study is that, Segre and colleagues used the disagreements between model predictions and experiments to guide *experimental* refinement, which also improved significantly the experimental data. After these refinements in experimental information, they repeated the computational analysis and comparisons with experiments to improve and refine the genome-scale metabolic model.

The first yeast genome-scale model has until now (September 2011), 365 citations, with 79 reviews and 263 research articles, suggesting an important impact in yeast research. However, a few observations can be made regarding the applications of the model. First, a very small number have used the model for discovery of genetic modifications and guidance for metabolic engineering towards the improved strain performance. In the first of the three most notable cases, the model was used to identify and rank a set of gene deletions and insertions for the manipulation of redox metabolism towards increased of ethanol yield (Bro*, et al.*, 2006). Experimental implementation validated the predictions and demonstrated improved ethanol yields even on xylose/glucose mixtures. The second notable case involves the identification of five,

nonobvious gene deletions for the engineering of C1 metabolism (Kennedy, *et al.*, 2009). Finally, in the third case, the yeast genome-scale model was used for the identification of metabolic engineering targets for improving the production of sesquiterpenes (Asadollahi, *et al.*, 2009). The complexity of the pathway and its distance from the central carbon pathway (there are 8 reactions in the mevalonate pathway from acetyl-CoA to farnesyl-diphosphate which is the primary precursor of the various sesquiterpenes) made the use of the genome-scale model indispensible. The resulting metabolic engineering strategy, which involved multiple genetic modifications, demonstrated the value and the validity of the model.

On the other hand, a large number of papers that cited the first yeast genome-scale model, focused on metabolomics analysis. However, they primarily use the model as a high-quality curated database of metabolites and reactions. While this two-dimensional annotation has been one of the objectives in genome-scale modeling (Palsson, 2004, Reed, *et al.*, 2006), it does not contribute immediately into design of strategies for strain improvement of disease treatment.

The work by Patil and Nielsen (Patil & Nielsen, 2005) has been an enabling breakthrough because it allowed the integration of gene expression and metabolomic data into the genome-scale metabolic model for the identification of network patterns that follow a common transcriptional response. The algorithm they developed identifies *reporter metabolites* and a set of connected genes with significant coordinated changes to genetic and environmental perturbations. This method allows now to use the genome-scale model, together with other genomic technologies, such as transcription factor enrichment, for the identification of important regulatory proteins and their associated

regulatory networks (Cakir, *et al.*, 2006, Raghevendran, *et al.*, 2006, Fazio, *et al.*, 2008, Cimini, *et al.*, 2009).

Finally, the integration of proteomics information within the context of genome-scale modeling is a recent exciting development (Costenoble, *et al.*, 2011). While this study focused in the study of metabolic adaptation to changes in nutritional conditions, it demonstrated the feasibility to use targeted proteomics for the quantification of almost all the enzymes in central carbon and amino-acid pathways. The synergistic application of these technologies and methodologies with genome-scale model analysis will be a major progress for metabolic engineering.

APPROACHES TO ADDRESS SOME ISSUES IN THE FLUX BALANCE ANALYSIS OF METABOLIC MODELS

The discussion above highlights a surprisingly limited use of genome scale models for metabolic engineering. It appears that the community working in this field has been more active with generating new and larger models and less so with actually using the models. As Uwe Sauer observed (personal communication), it is the latter that matters, but in every nascent field it is a bit like that because it is easier to develop tools than to reach new scientific discoveries by applying them, and metabolomics or fluxomics are no different in this respect.

The limited uses of genome scale models are due to many challenges and issues, which make it hard for somebody without a good experience in modeling and computation to use them in a productive fashion. One of the key challenges in FBA of genome-scale models is the possibility of multiple solutions resulting from the underdetermined nature of the problem. The number of alternate solutions scale exponentially with the size of the

network (Mahadevan & Schilling, 2003). Even though there are methods that aim to characterize the different flux modes in order to systematically analyze the possibilities, such as Elementary Flux Modes (EFM), Extreme Pathway (EPs) and other variants, most of these methods still do not perform well as the size of the model increases, and their applicability and usefulness are restricted. Therefore, given the limited amount of information about certain fluxes or enzyme activities, the main challenge is how we can derive a representative or characteristic flux distribution that can explain the observed phenotype at steady state. Such representative flux state(s) could also be a combination of more elementary flux states that should be further identified and characterized (Hoffmann*, et al.*, 2006, Barrett*, et al.*, 2009, Llaneras & Pico, 2010).

Flux balance models of metabolism are routinely used in the fitting of labeling experiments for the quantification of metabolic fluxes. However, all of these studies, with one notable exception (Blank*, et al.*, 2005), employ small scale, reduced models of yeast metabolism and they derive additional constraints for determining unique flux profiles. The concept of core models is not new; in fact historically genome-scale models have evolved from reduced "core" stoichiometric models by including increasingly details. The scale of these models made them more manageable and facilitated analysis. The issue of manageability is illustrated by the number of possible flux modes that the network can have, e.g. for a small yeast network comprising 53 reactions, there can be up to 6,741 EFMs depending on the carbon source (Dunn*, et al.*, 1994) whereas for an *E. coli* model with 112 reactions, the number of EFMs calculated was 2,450,787 (Perko, 1986). Hence, even though there are methods that can allow the almost complete enumeration and characterization of the EFMs/EPs, the scale of the resulting number of

possibilities will remain a huge obstacle in analysis and we must make some drastic assumptions to reduce the possibilities.

Another driver in the use of reduced models has been the objective to understand central metabolism well enough, before attempting to understand and make predictions at the genome-scale. Actually, in most of the problems in metabolic engineering the desired outcome has been the manipulation of central metabolism for redirecting the carbon flux towards desired pathways.

However, reduced models that are used to perform analyses of experimental data are often incompatible with each other as the set of reactions, components and degree of detail (e.g. proton-balancing and balancing of cofactors, etc.) differ significantly. Moreover, there is not an explicit list of the assumptions that would allow consistency checks of the model. For example, the assumptions about the presence or absence of alternate pathways in the determination of flux ratios for labeling experiments will affect the variability of the flux distribution. This can lead to different conclusions arising from the same set of data and difficulty in cross-utilization of datasets across laboratories that could have helped to further the complete characterization of the network.

Typically, reduced or core models in the past have been built in a bottom up approach. We believe that we need a top-down approach that can take advantage of all the knowledge in the genome-scale models. One of the main objectives of such approach will be to recover the simplicity and clarity of these earlier core models without losing the annotation details and the curated knowledge that has been amassed into the genome-scale models. With the increasing addition of details in these genome-scale models, it is necessary and important to add new knowledge consistently and modularly in order to keep track of changes, for example, with different releases of the *S. cerevisiae*

reconstructions. Moreover, a computational method will allow for a systematic and unambiguous model reduction, and it will facilitate consistency and communication between different laboratories.

Thermodynamic analysis of metabolic networks has been also shown to be important in reducing the flux space and eliminating thermodynamically infeasible pathways (Henry, *et al.*, 2007, Boghigian, *et al.*, 2010, Soh & Hatzimanikatis, 2010). Thermodynamics can also help to eliminate the need for *ad hoc* assignment of reaction directionality that can unwittingly preclude possible flux distributions that might be of interest. An example is the phosphoenoylpyruvate carboxykinase (PEPCK) reaction that is often assumed to be operating in the ATP-utilization direction. However, as shown both experimentally and computationally, this reaction can operate in the reverse direction under certain conditions (Deok, *et al.*, 2006, Gorsich, *et al.*, 2006, Singh, *et al.*, 2011). Hence by assuming certain fixed directionalities in the model, we might eliminate prematurely the true state of the network prior to analysis. Therefore, thermodynamics must be used to improve the curation of the models, as they provide additional control over the decision between the assumed, in literature or based on generalized arguments, reaction directionality vs. the possible reaction directionality, based on the estimated Gibbs free energy *and* the possible range of metabolite concentrations in the cell or metabolomics data.

Besides reducing the flux space effectively, thermodynamics offer another approach for integrating and overlaying additional layers of information in the form of thermodynamic displacement and metabolite concentration information. It has been shown (Henry, *et al.*, 2007, Soh & Hatzimanikatis, 2010) that if we include additional information in the form of metabolomics and fluxomics data, we can reduce the possible flux ranges of the

network to help us in better characterization of the flux distribution. Network thermodynamics can also be used to check for consistency of the metabolomics data with flux data, since we would expect that the directionality of the reactions, as determined by the full set of metabolites measured in an experiment, is not in conflict with the directionality of the reactions as determined from the labeling data.

As FBA models are only snapshots of the network at a point in time and they do not allow us to extrapolate the dynamic response of the network. Although approaches based on FBA (Mahadevan*, et al.*, 2002) attempt to overcome this limitation, these methods often use a highly reduced model, and they cannot simulate or predict the response of the metabolite levels since they do not integrate kinetic information. The biggest limitation of FBA methods is their inability to predict response to changes in enzyme activities. In most case in metabolic engineering we are interested to identify enzymes as targets for overexpression and/or downregulation, since gene knock out can have a detrimental effect on the physiology of the strain. A recent, very interesting study investigated the effects of single nucleotide polymorphisms (SNPs) on the phenotypic differences between two different yeast strains (Canelas*, et al.*, 2010). The investigators found SNPs in 20% of the metabolic genes and based on these differences they hypothesized physiological differences, which they confirmed experimentally. Based on further transcriptomic analysis, the authors hypothesized that SNPs can be responsible for change in enzyme concentration and/or function, such as kinetic properties. Such hypotheses, as well as identification of targets for gene overexpression and protein engineering, cannot be analyzed without the use of kinetic models of metabolic networks.

SOME CONSIDERATIONS ON THE DEVELOPMENT OF KINETIC MODELS OF YEAST METABOLISM

One of the more widely used yeast kinetic models (Teusink*, et al.*, 2000) for analysis and also further model development has about 257 citations to date (September 2011). However, very little has been done in the original development of large-scale kinetic models in yeast. With the exception of one case (Wang & Hatzimanikatis, 2006), almost all models of yeast central carbon catabolism do not distinguish the mitochondrial reactions from the cytosolic reactions. The main issue in the development of kinetic models of metabolic networks is the limited available information and the uncertainty associated with this information. We have previously studied and classified the uncertainty in the study of metabolic pathways in two types: *structural* and *quantitative* (Miskovic & Hatzimanikatis, 2010). *Structural uncertainty* concerns the limited knowledge in the stoichiometry and in the kinetic laws of the enzymes in the pathways. While the stoichiometry of the pathways in yeast is well characterized, there still exist gaps in some pathways and the kinetics of their enzymes are completely unknown (DeJongh*, et al.*, 2007, Feist*, et al.*, 2009, Henry*, et al.*, 2009, Kumar & Maranas, 2009, Stanley*, et al.*, 2010). Also the kinetic parameters of most of enzymes are not available and when they are available they are usually known as "apparent K_m's" but not as parameters in detailed kinetic mechanisms. There is also an important concern on how the parameters of the enzymes quantified *in vitro* will change in the crowded intracellular environment (Savageau, 1995, Schnell & Turner, 2004).

Flux distributions, thermodynamic information, metabolite concentration and kinetic parameters are subject to *quantitative uncertainty*. Despite the advances of methods for the quantification of metabolic fluxes, they still carry some error. The thermodynamic

properties of most of the reactions are estimated using group contribution methods, and therefore they contain estimation errors and the error of the experiments used in the estimation process (Jankowski*, et al.*, 2008). The highest uncertainty is in the metabolite measurements, and in addition only a relatively small number of metabolites can be measured compared to the entire metabolome of the organism.

The uncertainty in building mathematical models is very large even for systems that are well studied, such as *E. coli* and *S. cerevisiae*. Therefore, when we consider the analysis and engineering of novel pathways, we should expect much higher qualitative and quantitative uncertainty in the information about these systems (Tyo*, et al.*, 2007, Alper & Stephanopoulos, 2009).

Uncertainty is a problem common in many areas of physical and chemical sciences and engineering. Within these fields there exist a large number of methods and approaches that allow for the modeling and quantification of uncertainty. These methods have been used in the analysis of metabolic networks and they have provided some insight into the properties of the networks, and guidance for metabolic engineering (Wang*, et al.*, 2004, Wang & Hatzimanikatis, 2006, Wang & Hatzimanikatis, 2006, Kiparissides*, et al.*, 2009). However, any significant effort in this area faces challenges in the modeling and simulation of uncertainty. When we consider uncertainty modeling and analysis of kinetic models of chemical and biochemical systems, we must ensure sufficiency in the sampling of the kinetic parameters, calculate the properties of a *population* of the system, solve large systems of nonlinear equations, and perform a statistical analysis to characterize the properties of the population of the system. This leads to many computational challenges: (i) the ranges of the parameter values are not known or they are very large; (ii) the size and nonlinearities introduce computational difficulties; and (iii) reliable statistics can

require a computationally prohibiting number of samples. We have recently developed an uncertainty analysis framework, tailored to metabolic systems, and we have made significant progress in addressing these issues (Miskovic & Hatzimanikatis, 2010).

PREDICTING NETWORK RESPONSES WITH LIMITED INFORMATION

Optimization and Risk Analysis of Complex Living Entities (ORACLE) is a modeling and computational framework we have recently introduced for the study of metabolic networks under uncertainty (Wang*, et al.*, 2004, Wang & Hatzimanikatis, 2006, Wang & Hatzimanikatis, 2006, Miskovic & Hatzimanikatis, 2010). It uses uncertainty and risk analysis methods, and it circumvents most of the limitations mentioned above. In its current stage ORACLE is used for Metabolic Control Analysis (MCA) and it allows the quantification of the flux control coefficients and concentration control coefficients. These coefficients quantify the fold-change in metabolic fluxes and metabolite concentrations for a fold change in enzyme activities or in any environmental parameter. There are other similar algorithms for analysis of kinetic metabolic models (Steuer*, et al.*, 2006, Tran*, et al.*, 2008) and a recent paper (Miskovic & Hatzimanikatis, 2011) explains the differences between ORACLE and these approaches. Main advantages of the ORACLE framework are: (i) the capability to consistently integrate thermodynamics and physico-chemical constraints into kinetic models, (ii) the capability to integrate omics information (transcriptomics, proteomics, metabolomics, and fluxomics) and (iii) its scalability that enables to predict kinetic responses of metabolism even for genome-scale metabolic models, which is not feasible with any of the other approaches.

The pivotal point in the development of ORACLE is the recognition that control coefficients depend on the degree of enzyme saturation, also known as enzyme elasticities, which in turn can be estimated through the distribution of the enzyme between the different mechanistic enzyme states. This observation led us to the reconsideration of the uncertainties in the *enzyme state space* instead of the *kinetic parameter space*. This reformulation gives the major advantage that we can derive the degree of saturation, or elasticities, by sampling the *enzyme state space*, which, unlike the parameter space, is very well bounded between 0 and 1. These bounds can be further constrained if the kinetic parameters of an enzyme are approximately known.

The ORACLE framework involves a set of computational procedures, which integrate the available information into a mathematical structure, and through Monte Carlo sampling for retrofitting missing information they generate the population of all possible control coefficients. Conceptually, ORACLE involves the following steps (Figure 1):

Step 1. Integration of available information. We start by defining the *stoichiometry*, based on the information from the genome scale model. We proceed further by integrating the estimated *flux profiles* based on information from fluxomics analysis or on hypotheses about desirable flux distributions in an engineered pathway. Finally, we estimate the *standard free energy of reactions* based on the available experimental information, or using group contribution methods.

Step 2. Exploring the space of metabolite concentrations. The levels of concentrations for some of the metabolites in the system might be available, or can be estimated from experiments under similar physiological conditions. For the metabolites whose levels are missing, we can use sampling under thermodynamic constraints in order to preserve the observed flux directionality.

Step 3. Exploring the space of the kinetic properties (elasticities). Sampling of either the enzyme states (Miskovic & Hatzimanikatis, 2011) or the degree of saturation of the enzymes active site (Wang*, et al.*, 2004) is very efficient and it can also integrate partial knowledge of the enzyme kinetic parameters.

Step 4. Consistency checks and pruning. Partial knowledge of experimentally observed response of a metabolic flux to the changes in the activity of an enzyme is used to reject inconsistent samples.

Step 5. Calculation and statistical analysis, **data mining**, **and visualization of control coefficients.** The populations of control coefficients are subsequently analyzed using non-parametric statistics and data mining in order to asses and rank the importance of the enzymes with respect to their impact on the specified objectives (Silverman, 1986, Conover, 1998, Chen & Lonardi, 2009).

Ultimately, the results from ORACLE are not *predictions* but *statistical expectations* of success of the metabolic engineering targets they identify. ORACLE provides a set of alternative solutions, evaluated with respect to their uncertainty, which can be given back to the experts for evaluation. This 'expert opinion' is the ultimate integration of information that is almost impossible to take into account during the formulation of the model. Overall, ORACLE employs modeling and analysis in a new way, which have been successfully used in other disciplines.

THERMODYNAMIC AND KINETIC ANALYSIS OF A REDUCED, CORE METABOLIC MODEL OF YEAST

In this section, we will discuss some of our recent unpublished work to illustrate how we can approach some of the problems discussed earlier in the paper. Our work is based on a core yeast metabolic model. We have developed a computational algorithm that allows the reduction of genome scale models into core metabolic models. This method allows the unambiguous reduction of genome scale models and it is also "reversible", in the sense that the results from the analysis of the core model can be compared *exactly* with the genome scale model. The reduction was based on the *iMM904* model (Mo*, et al.*, 2009) and this is the first such core model for yeast and it consists of 89 reactions and 88 metabolites across 2 compartments (cytosol and mitochondrial) as shown in Figure 2. All the reactions are proton-balanced as it has been shown to be important in affecting the overall solution (Fox*, et al.*, 2007). Modeling the reactions around, across, and inside compartments is very important for understanding the *in vivo* redox and energy balance (Karbowicz & Smith, 1984), but unfortunately is often neglected in most reduced models. In addition, we performed thermodynamic curation and we have been able to include thermodynamic constraints in our reduced model.

We first used this model to perform some basic flux balance analysis, and we used reference experimental data from recent work from the Sauer Lab to compare our results. Initially we used only information about the carbon source uptake rate and product fluxes (Wang*, et al.*, 2011) and without assuming any reaction directionalities *a priori,* we performed FBA and flux variability analysis without any thermodynamic constraints. As expected, we found that the system is under-constrained and it is able to generate biomass

from CO_2 recycling reactions and ATP recycling. However, after adding thermodynamic constraints, we observe that the maximum biomass flux drops to close to the measured value as many of the CO_2 recycling reactions are automatically constrained in the proper direction under normal physiological concentration ranges predicted by the thermodynamic constraints (Figure 3). On the other hand, when we fix the reaction directionalities in the direction most commonly assumed in genome scale models, we find that the flux variability is significantly reduced (Figure 4). By specifying the reaction directionality *a priori,* we can overly constraint the model in two ways. First, as discussed earlier, reactions that can be reversible under certain conditions, e.g. in the case of PEPCK which was found to be able to operate in the ATP-generating direction in *E. coli* (Deok*, et al.*, 2006, Singh*, et al.*, 2011) and in *S. cerevisiae* (Gorsich*, et al.*, 2006) under high CO_2 concentrations. Hence by setting the reaction directionality a priori we would have eliminated this possibility and we could not explain the observed physiology using the model. Second, by assigning *a priori* directionalities, we introduce in the system *ad hoc* inflexibility and tight constraints, as we observe that, even with thermodynamic constraints, the flux ranges are quite large as compared to those with specified reaction directionalities. Although in metabolomics and fluxomics studies, we would like the model to have few degrees of freedom, in order to have smaller uncertainties in the estimation of the flux values, we should not contaminate our analysis with artifacts from arbitrary assumptions about reaction directionality.

All these results and conclusions from FBA of the core model have been found to hold when we used the corresponding genome scale model used to derive the core model. Therefore, researchers who are familiar with FBA on small stoichiometric models, but they are not experienced working with genome scale models, can use and analyze this

reduced core model much easier, and their results and conclusions can then be used for genome-scale analysis.

After obtaining a representative flux profile from the thermodynamics-based flux balance analysis (TFBA), we sampled feasible metabolite concentrations and computed the corresponding reaction displacement from thermodynamic equilibrium. We observed (Figure 2) that the displacement of approximately half of the reactions could be either near or far from equilibrium, whereas the other reactions could assume a wider range of displacements (Table 1).

We also used ORACLE to investigate how changes in the activities of the enzymes in the network would affect the flux distribution and the levels of the metabolites. We investigated the response of the splitting ratio of the glycolytic fluxes, quantified by the ratio of the flux through fructose-biphosphate aldolase (FBA), over the flux through glucose-6-phosphate-1-dehydrogenase (ZWF). We found that the primary positive control over this ratio lies in ATP maintenance and pyruvate decarboxylase (PDC), whereas the negative control lies in ammonia (NH4t) and oxaloacetate (OAt) transport (Figure 5). Interestingly, even though an enhancement of hexose transporters (HXT) or hexokinase (HXK) activity has, in average, negative impact on this ratio error bars suggest that there exist physiological states where its effect could be positive.

The ATP/ADP ratio and the redox potential (NADH/NAD) are important factors in metabolic engineering as adenylate cofactors and pyridine nucleotides are involved in many reactions. Our analysis suggests that the control over these quantities is distributed differently depending on the compartment of the cell. More specifically, we observe that a group of enzymes, i.e. HXT, PDC, external NADH dehydrogenase (NDH), ATP synthase (ASN) and CO2 transport, have positive control, and ATPM, ADP/ATP carrier

protein (AAC) have negative control over ATP/ADP ratio in the mitochondria, (Figure 6A). On the other hand, in the cytosol the positive control over ATP/ADP ratio is primarily from HXT, whereas the negative control is shifted to PDC, pyruvate dehydrogenase (PDA) and NH4t (Figure 6B). Similarly, we observe that the major positive control over redox potential in mitochondria (Figure 6C) is in glucose-6-phosphate isomerase (PGI), and the negative control is distributed between ZWF, CO2 and 6-Phospho-D-glucono-1,5-lactone lactonohydrolase (GND1). In contrast, in the cytosol the biggest positive control coefficients of energy charge are those with respect to PGI nad ATPM, whereas HXT and HXK appear to have the most important negative control. This interesting connection of the redox potential in the mitochondria and the activities in the upper glycolysis and pentose phosphate can be identified and explained only through the application and use of ORACLE.

We also have analyzed how the control is distributed over the ethanol yield (with ethanol being as one of most important industrial products) from glucose. Though ATPM and HXT have major positive and negative effect respectively (Figure 7), we observe that the control coefficients are very small in magnitude, and even for the most significant enzymes the mean value is not bigger than 0.1 suggesting that activities of multiple enzymes should be altered to effectively increase ethanol yield.

Table 1: Distribution of reactions' displacement from thermodynamic equilibrium in the network

	NE (I)	Between (II)	FA(III)	I+II	II+III	I+II+III
Number of rxns	9	0	27	11	6	22
Percentage	12%	0	36%	15%	8%	29%

CONCLUSIONS

The integration of regulatory constraints will be the next major advancement in the area of metabolic modeling in yeast. Past and ongoing work in the Palsson and Nielsen labs is advancing rapidly developments in this area. An interesting approach could come from the combination of concepts from the work of Patil and Nielsen (Patil & Nielsen, 2005) and the work by Price (Hasunuma*, et al.*, 2011). Such approach will provide important missing links for the development of kinetic models.

Ultimately a kinetic, nonlinear model is the goal. While there exist a number of publications, which claim such models, they all face a lot of limitations, which has not been adequately addressed. We should always keep in mind the proverbial quote from Manfred Eigen: "A theory has only the alternative of being right or wrong. A model has a third possibility: it may be right, but irrelevant." The relevance of the mathematical models in yeast will be evaluated from their contribution to the advancement in our understanding of disease and to the accelerated development of industrial strains. While there is a lot of evidence from the research fronts in these areas, successful resolution of some of the issues discussed in this article will enhance and broaden the impact of mathematical modeling in yeast research.

ACKNOWLEDGEMENTS

The authors acknowledge the help of Christen Stefan and Prof Uwe Sauer with providing details on their published data. KCS was supported by the Swiss National Science Foundation. LM and VH were supported by funding from Ecole Polytechnique Fédérale de Lausanne (EPFL) and NEMO for Bioethanol, EU FP7 Programme. Support has been also provided from SystemsX.ch, The Swiss Initiative in Systems Biology, through the project MetaNetX.

REFERENCES

Albert R & Barabasi AL (2002) Statistical mechanics of complex networks. *Reviews of Modern Physics* **74**: 47-97.

Alper H & Stephanopoulos G (2009) Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? *Nature Reviews Microbiology* **7**: 715-723.

Ananiadou S, Pyysalo S, Tsujii J & Kell DB (2010) Event extraction for systems biology by text mining the literature. *Trends in Biotechnology* **28**: 381-390.

Asadollahi MA, Maury J, Patil KR, Schalk M, Clark A & Nielsen J (2009) Enhancing sesquiterpene production in Saccharomyces cerevisiae through in silico driven metabolic engineering. *Metabolic Engineering* **11**: 328-334.

Barabasi AL & Oltvai ZN (2004) Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics* **5**: 101-U115.

Barrett C, Herrgard M & Palsson B (2009) Decomposing complex reaction networks using random sampling, principal component analysis and basis rotation. *BMC Syst Biol* **3**: 30.

Blank LM, Kuepfer L & Sauer U (2005) Large-scale C-13-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biology* **6**.

Boghigian BA, Shi H, Lee K & Pfeifer BA (2010) Utilizing elementary mode analysis, pathway thermodynamics, and a genetic algorithm for metabolic flux determination and optimal metabolic network design. *Bmc Systems Biology* **4**: 49.

Bolotin-Fukuhara M, Dumas B, Gaillardin C & (Eds) (2010) Special Issue: Yeasts as a model for human diseases. *FEMS Yeast Research* **10**.

Bro C, Regenberg B, Forster J & Nielsen J (2003) Metabolic engineering in Saccharomyces cerevisiae through the use of a reconstructed genome-scale metabolic network leads to improved ethanol production. *Yeast* **20**: S284-S284.

Bro C, Regenberg B, Forster J & Nielsen J (2006) In silico aided metabolic engineering of Saccharomyces cerevisiae for improved bioethanol production. *Metabolic Engineering* **8**: 102-111.

Cakir T, Patil KR, Onsan ZI, Ulgen KO, Kirdar B & Nielsen J (2006) Integration of metabolome data with metabolic networks reveals reporter reactions. *Molecular Systems Biology* **2**.
Canelas AB, Harrison N, Fazio A*, et al.* (2010) Integrated multilaboratory systems biology reveals differences in protein metabolism between two reference yeast strains. *Nature communications* **1**: 145.
Chen JY & Lonardi S (2009) *Biological Data Mining*. Chapman and Hall/CRC.
Cimini D, Patil KR, Schiraldi C & Nielsen J (2009) Global transcriptional response of Saccharomyces cerevisiae to the deletion of SDH3. *Bmc Systems Biology* **3**.
Conover WJ (1998) *Practical Nonparametric Statistics*. Wiley.
Costa V & Moradas-Ferreira P (2001) Oxidative stress and signal transduction in Saccharomyces cerevisiae: insights into ageing, apoptosis and diseases. *Molecular aspects of medicine* **22**: 217-246.
Costenoble R, Picotti P, Reiter L, Stallmach R, Heinemann M, Sauer U & Aebersold R (2011) Comprehensive quantitative analysis of central carbon and amino-acid metabolism in Saccharomyces cerevisiae under multiple conditions by targeted proteomics. *Molecular Systems Biology* **7**.
DeJongh M, Formsma K, Boillot P, Gould J, Rycenga M & Best A (2007) Toward the automated generation of genome-scale metabolic networks in the SEED. *BMC Bioinformatics* **8**.
Deok KY, Lee SY & Kim P (2006) Influence of gluconeogenic phosphoenolpyruvate carboxykinase (PCK) expression on succinic acid fermentation in Escherichia coli under high bicarbonate condition. *Journal of Microbiology and Biotechnology* **16**: 1448-1452.
Duarte NC, Herrgard MJ & Palsson BO (2004) Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome Research* **14**: 1298-1309.
Dunn DA, Grover WD & Macgregor MH (1994) Comparison of K-Shortest Paths and Maximum Flow Routing for Network Facility Restoration. *Ieee Journal on Selected Areas in Communications* **12**: 88-99.
Famili I, Forster J, Nielson J & Palsson BO (2003) Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 13134-13139.
Fazio A, Jewett MC, Daran-Lapujade P*, et al.* (2008) Transcription factor control of growth rate dependent genes in Saccharomyces cerevisiae: A three factor design. *Bmc Genomics* **9**.
Feist AM & Palsson BØ (2008) The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. Vol. 26 ed.^eds.), p.^pp. 659-667.
Feist AM, Herrgard MJ, Thiele I, Reed JL & Palsson BO (2009) Reconstruction of biochemical networks in microorganisms. *Nature Reviews Microbiology* **7**: 129-143.
Forster J, Famili I, Fu P, Palsson BO & Nielsen J (2003) Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. *Genome Research* **13**: 244 - 253.
Fox Z & Antunes F & Davey R*, et al.* (2007) Predictors of CD4 count change over 8 months of follow up in HIV-1-infected patients with a CD4 count >= 300 cells/mu L who were assigned to 7.5 MIU interleukin-2. *Hiv Medicine* **8**: 112-123.
Gorsich SW, Dien BS, Nichols NN, Slininger PJ, Liu ZL & Skory CD (2006) Tolerance to furfural-induced stress is associated with pentose phosphate pathway genes ZWF1,

GND1, RPE1, and TKL1 in Saccharomyces cerevisiae. *Applied Microbiology and Biotechnology* **71**: 339-349.
Hartwell LH (2004) Nobel 2001 lecture: Yeast and cancer. *Bioscience Reports* **24**: 523-544.
Hasunuma T, Sanda T, Yamada R, Yoshimura K, Ishii J & Kondo A (2011) Metabolic pathway engineering based on metabolomics confers acetic and formic acid tolerance to a recombinant xylose-fermenting strain of Saccharomyces cerevisiae. *Microbial cell factories* **10**: 2.
Henry CS, Broadbelt LJ & Hatzimanikatis V (2007) Thermodynamics-based metabolic flux analysis. *Biophysical Journal* **92**: 1792-1805.
Henry CS, Zinner JF, Cohoon MP & Stevens RL (2009) iBsu1103: a new genome-scale metabolic model of Bacillus subtilis based on SEED annotations. *Genome Biology* **10**.
Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B & Stevens RL (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature Biotechnology* **28**: 977-U922.
Herrgard MJ, Swainston N, Dobson P*, et al.* (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nature Biotechnology* **26**: 1155-1160.
Hoffmann S, Hoppe A & Holzhutter HG (2006) Composition of metabolic flux distributions by functionally interpretable minimal flux modes (MinModes). *Genome informatics. International Conference on Genome Informatics* **17**: 195-207.
Hong ME, Lee KS, Yu BJ*, et al.* (2010) Identification of gene targets eliciting improved alcohol tolerance in Saccharomyces cerevisiae through inverse metabolic engineering. *Journal of Biotechnology* **149**: 52-59.
Hyduke DR & Palsson BO (2010) Towards genome-scale signalling-network reconstructions. *Nature Reviews Genetics* **11**: 297-307.
Jankowski MD, Henry CS, Broadbelt LJ & Hatzimanikatis V (2008) Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophysical Journal* **95**: 1487-1499.
Jewett MC, Oliveira AP, Patil KR & Nielsen J (2005) The role of high-throughput transcriptome analysis in metabolic engineering. *Biotechnology and Bioprocess Engineering* **10**: 385-399.
Joyce AR & Palsson BO (2006) The model organism as a system: integrating 'omics' data sets. *Nature Reviews Molecular Cell Biology* **7**: 198-210.
Karbowicz CJ & Smith JM (1984) A K-Shortest Paths Routing Heuristic for Stochastic Network Evacuation Models. *Engineering Optimization* **7**: 253-280.
Kell DB (2005) Metabolomics, machine learning and modelling: towards an understanding of the language of cells. *Biochemical Society Transactions* **33**: 520-524.
Kennedy CJ, Boyle PM, Waks Z & Silver PA (2009) Systems-Level Engineering of Nonfermentative Metabolism in Yeast. *Genetics* **183**: 385-397.
Kiparissides A, Kucherenko SS, Mantalaris A & Pistikopoulos EN (2009) Global Sensitivity Analysis Challenges in Biological Systems Modeling. *Industrial & Engineering Chemistry Research* **48**: 7168-7180.
Kuepfer L, Sauer U & Blank LM (2005) Metabolic functions of duplicate genes in Saccharomyces cerevisiae. *Genome Research* **15**: 1421-1430.
Kumar VS & Maranas CD (2009) GrowMatch: An Automated Method for Reconciling In Silico/In Vivo Growth Predictions. *Plos Computational Biology* **5**.

Lin FM, Qiao B & Yuan YJ (2009) Comparative proteomic analysis of tolerance and adaptation of ethanologenic Saccharomyces cerevisiae to furfural, a lignocellulosic inhibitory compound. *Applied and Environmental Microbiology* **75**: 3765-3776.
Llaneras F & Pico J (2010) Which metabolic pathways generate and characterize the flux space? A comparison among elementary modes, extreme pathways and minimal generators. *Journal of biomedicine & biotechnology* **2010**: 753904.
Mahadevan R & Schilling CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Vol. 5 ed.^eds.), p.^pp. 264-276.
Mahadevan R, Edwards JS & Doyle FJ (2002) Dynamic flux balance analysis of diauxic growth in Escherichia coli. *Biophysical Journal* **83**: 1331-1340.
Mehra A & Hatzimanikatis V (2006) An algorithmic framework for genome-wide modeling and analysis of translation networks. *Biophysical Journal* **90**: 1136-1146.
Miskovic L & Hatzimanikatis V (2010) Production of biofuels and biochemicals: in need of an ORACLE. *Trends in Biotechnology* **28**: 391-397.
Miskovic L & Hatzimanikatis V (2011) Modeling of uncertainties in biochemical reactions. *Biotechnology and Bioengineering* **108**: 413-423.
Mo ML, Palsson BO & Herrgard MJ (2009) Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst Biol* **3**: 37.
Newman MEJ (2003) The structure and function of complex networks. *Siam Review* **45**: 167-256.
Nicolaou SA, Gaida SM & Papoutsakis ET (2010) A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: From biofuels and chemicals, to biocatalysis and bioremediation. *Metabolic Engineering* **12**: 307-331.
Nielsen J (2009) Systems biology of lipid metabolism: From yeast to human. *Febs Letters* **583**: 3905-3913.
Nielsen J & Jewett MC (2008) Impact of systems biology on metabolic engineering of Saccharomyces cerevisiae. *Fems Yeast Research* **8**: 122-131.
Nookaew I, Jewett MC, Meechai A*, et al.* (2008) The genome-scale metabolic model iIN800 of Saccharomyces cerevisiae and its validation: a scaffold to query lipid metabolism. *Bmc Systems Biology* **2**.
Oberhardt MA, Palsson BØ & Papin JA (2009) Applications of genome-scale metabolic reconstructions. Vol. 5 ed.^eds.), p.^pp. 1-15.
Palsson B (2004) Two-dimensional annotation of genomes. *Nature Biotechnology* **22**: 1218-1219.
Papin JA, Reed JL & Palsson BO (2004) Hierarchical thinking in network biology: the unbiased modularization of biochemical networks. *Trends in Biochemical Sciences* **29**: 641-647.
Patil KR & Nielsen J (2005) Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 2685-2689.
Patil KR, Akesson M & Nielsen J (2004) Use of genome-scale microbial models for metabolic engineering. Vol. 15 ed.^eds.), p.^pp. 64-69.
Perko A (1986) Implementation of Algorithms for K-Shortest Loopless Paths. *Networks* **16**: 149-160.
Petranovic D & Nielsen J (2008) Can yeast systems biology contribute to the understanding of human disease? *Trends in Biotechnology* **26**: 584-590.

Petranovic D & Vemuri GN (2009) Impact of yeast systems biology on industrial biotechnology. *Journal of Biotechnology* **144**: 204-211.
Radrich K, Tsuruoka Y, Dobson P, Gevorgyan A, Swainston N, Baart G & Schwartz JM (2010) Integration of metabolic databases for the reconstruction of genome-scale metabolic networks. *Bmc Systems Biology* **4**: -.
Raghevendran V, Patil KR, Olsson L & Nielsen J (2006) Hap4 is not essential for activation of respiration at low specific growth rates in Saccharomyces cerevisiae. *Journal of Biological Chemistry* **281**: 12308-12314.
Reed JL, Famili I, Thiele I & Palsson BO (2006) Towards multidimensional genome annotation. *Nature Reviews Genetics* **7**: 130-141.
Rieger TR, Morimoto RI & Hatzimanikatis V (2005) Mathematical modeling of the eukaryotic heat-shock response: Dynamics of the hsp70 promoter. *Biophysical Journal* **88**: 1646-1658.
Savageau MA (1995) Michaelis-Menten mechanism reconsidered - Implications of fractal kinetics. *Journal of Theoretical Biology* **176**: 115-124.
Schellenberger J, Park JO, Conrad TM & Palsson BO (2010) BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *Bmc Bioinformatics* **11**.
Schnell S & Turner TE (2004) Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws. *Progress in Biophysics & Molecular Biology* **85**: 235-260.
Silverman BW (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall/CRC.
Singh A, Cher Soh K, Hatzimanikatis V & Gill RT (2011) Manipulating redox and ATP balancing for improved production of succinate in E. coli. *Metab Eng* **13**: 76-81.
Snitkin ES, Dudley AM, Janse DM, Wong K, Church GM & Segre D (2008) Model-driven analysis of experimentally determined growth phenotypes for 465 yeast gene deletion mutants under 16 different conditions. *Genome Biology* **9**: R140.
Soh KC & Hatzimanikatis V (2010) Network thermodynamics in the post-genomic era. *Current Opinion in Microbiology* **13**: 350-357.
Sorolla MA, Reverter-Branchat G, Tamarit J, Ferrer I, Ros J & Cabiscol E (2008) Proteomic and oxidative stress analysis in human brain samples of Huntington disease. *Free radical biology & medicine* **45**: 667-678.
Stanley D, Chambers PJ, Stanley GA, Borneman A & Fraser S (2010) Transcriptional changes associated with ethanol tolerance in Saccharomyces cerevisiae. *Applied Microbiology and Biotechnology* **88**: 231-239.
Steuer R, Gross T, Selbig J & Blasius B (2006) Structural kinetic modeling of metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 11868-11873.
Teusink B, Passarge J, Reijenga CA*, et al.* (2000) Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. *Eur J Biochem* **267**: 5313-5329.
Tran LM, Rizk ML & Liao JC (2008) Ensemble modeling of metabolic networks. *Biophysical Journal* **95**: 5606-5617.
Tyo KE, Alper HS & Stephanopoulos GN (2007) Expanding the metabolic engineering toolbox: more options to engineer cells. *Trends in Biotechnology* **25**: 132-137.

Ukibe K, Hashida K, Yoshida N & Takagi H (2009) Metabolic engineering of Saccharomyces cerevisiae for astaxanthin production and oxidative stress tolerance. *Applied and Environmental Microbiology* **75**: 7205-7211.
van Maris AJ, Winkler AA, Kuyper M, de Laat WT, van Dijken JP & Pronk JT (2007) Development of efficient xylose fermentation in Saccharomyces cerevisiae: xylose isomerase as a key component. *Adv Biochem Eng Biotechnol* **108**: 179-204.
Walker GM (1998) *Yeast Physiology and Biotechnology*. Wiley.
Wang LQ & Hatzimanikatis V (2006) Metabolic engineering under uncertainty - II: Analysis of yeast metabolism. *Metabolic Engineering* **8**: 142-159.
Wang LQ & Hatzimanikatis V (2006) Metabolic engineering under uncertainty. I: Framework development. *Metabolic Engineering* **8**: 133-141.
Wang LQ, Birol I & Hatzimanikatis V (2004) Metabolic control analysis under uncertainty: Framework development and case studies. *Biophysical Journal* **87**: 3750-3763.
Wang X, Miller EN, Yomano LP, Zhang X, Shanmugam KT & Ingram LO (2011) Increased furfural tolerance due to overexpression of NADH-dependent oxidoreductase FucO in Escherichia coli strains engineered for the production of ethanol and lactate. *Applied and Environmental Microbiology* **77**: 5132-5140.
Wattanachaisaereekul S, Lantz AE, Nielsen ML, Andresson OS & Nielsen J (2007) Optimization of heterologous production of the polyketide 6-MSA in Saccharomyces cerevisiae. *Biotechnol Bioeng* **97**: 893-900.
Wisselink HW, Toirkens MJ, Wu Q, Pronk JT & van Maris AJA (2009) Novel evolutionary engineering approach for accelerated utilization of glucose, xylose, and arabinose mixtures by engineered Saccharomyces cerevisiae strains. *Appl Environ Microbiol* **75**: 907-914.
Zelle RM, de Hulster E, Kloezen W, Pronk JT & van Maris AJ (2010) Key process conditions for production of C(4) dicarboxylic acids in bioreactor batch cultures of an engineered Saccharomyces cerevisiae strain. *Appl Environ Microbiol* **76**: 744-750.
Zhao X, Jiang R, Li N, Yang Q & Bai F (2010) [Improving ethanol tolerance of Saccharomyces cerevisiae industrial strain by directed evolution of SPT3]. *Sheng wu gong cheng xue bao = Chinese journal of biotechnology* **26**: 159-164.

Figure captions

Figure 1. Flowchart of computational procedures for uncertainty analysis of metabolic networks within the ORACLE framework. Available information from different sources is integrated in the model through the successive application of the ORACLE procedures. The resulting models are consistent with the thermodynamics and experimentally observed data, while preserving all physical and chemical constraints of the underlying metabolic network.

Figure 2. Core metabolic network of *S. cerevisiae* showing the thermodynamic displacement of reactions in the network. The abbreviations for the pathway steps and enzymes can be found in the Appendix.

Figure 3. Effect of thermodynamic constraints on biomass flux

Figure 4. Flux Variability Analysis of central metabolic network of *S. cerevisiae* without and with thermodynamic constraints

Figure 5. Distribution of the control coefficients of the splitting ratio between the fluxes through fructose-biphosphate aldolase (FBA) and glucose-6-phosphate-1-dehydrogenase (ZWF) with respect to enzymes having the most of control over this ratio. The bars represent the mean values of the control coefficients, while the errorbars correspond to the 25 and 75 percentiles of the control coefficients with respect to their mean values. The abbreviations for the enzymes are given in Fig. 1.

Figure 6. Distribution of the control coefficients of: (A) ATP/ADP ratio in mitochondria (*ATPm/ADPm*), (B) ATP/ADP ratio in cytosol (*ATPc/ADPc*), (C) redox potential in mitochondria (*NADHm/NADm*) and (D) redox potential in cytosol (*NADHc/NADc*), with respect to enzymes having the most of control over these quantities.

Figure 7. Distribution of the control coefficients of ethanol yield from glucose with respect to enzymes having the most of control over the yield.

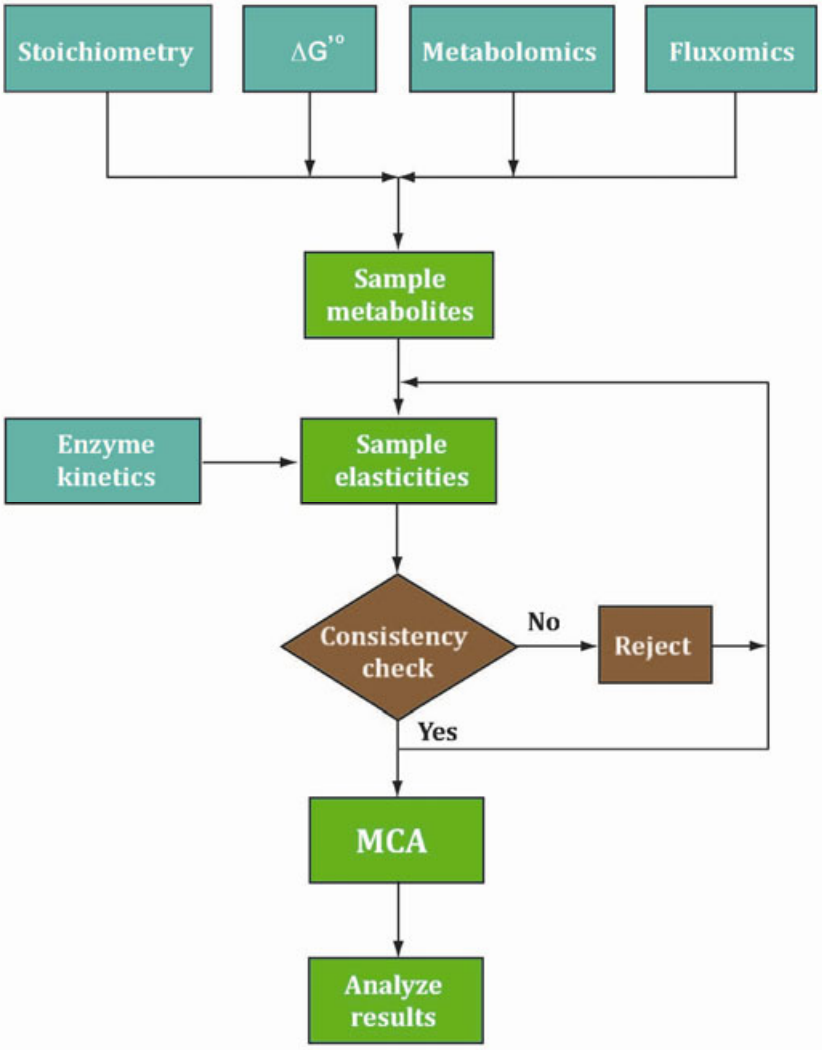
APPENDIX

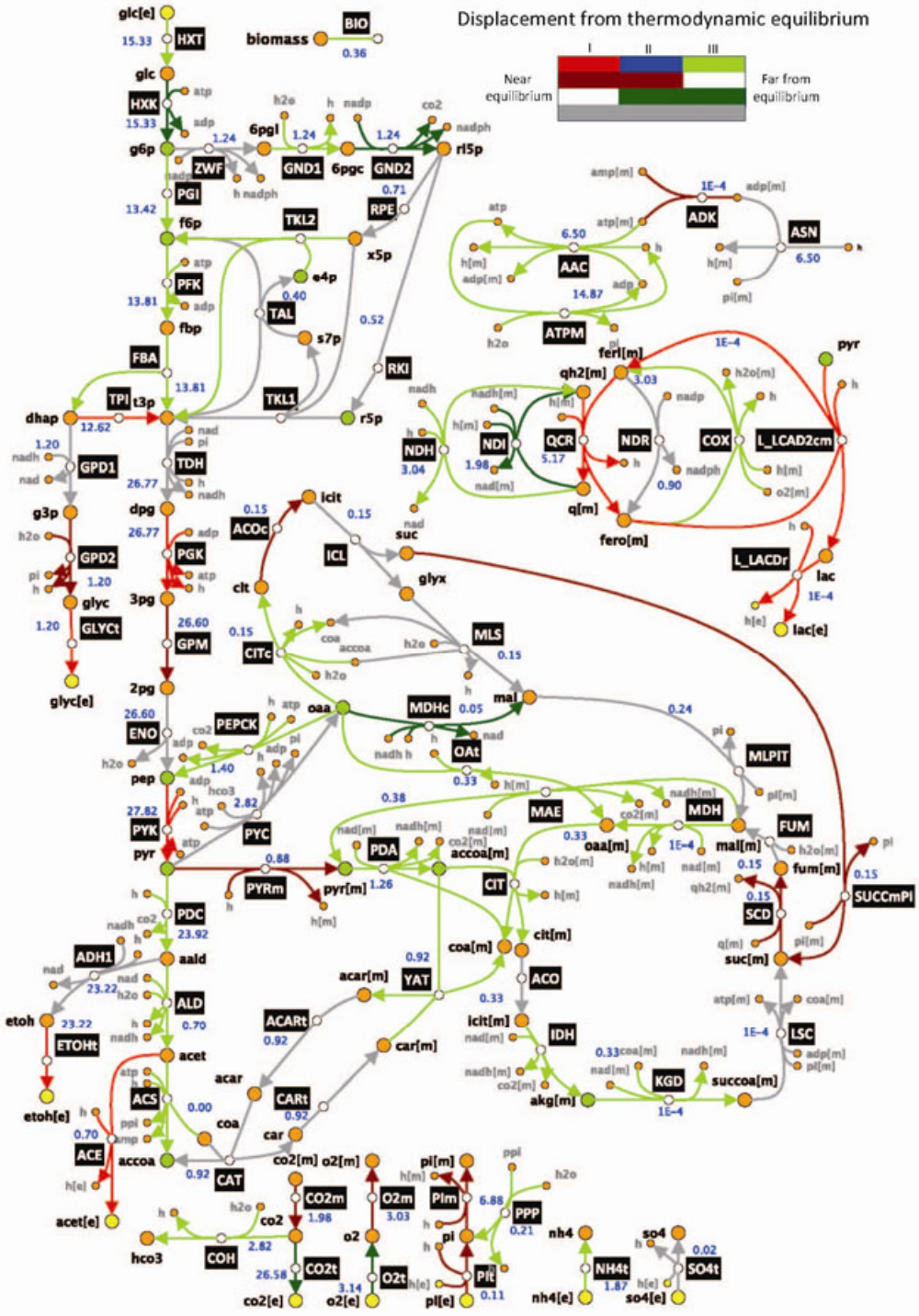
Names of metabolites and reactions of network in Figure 2:

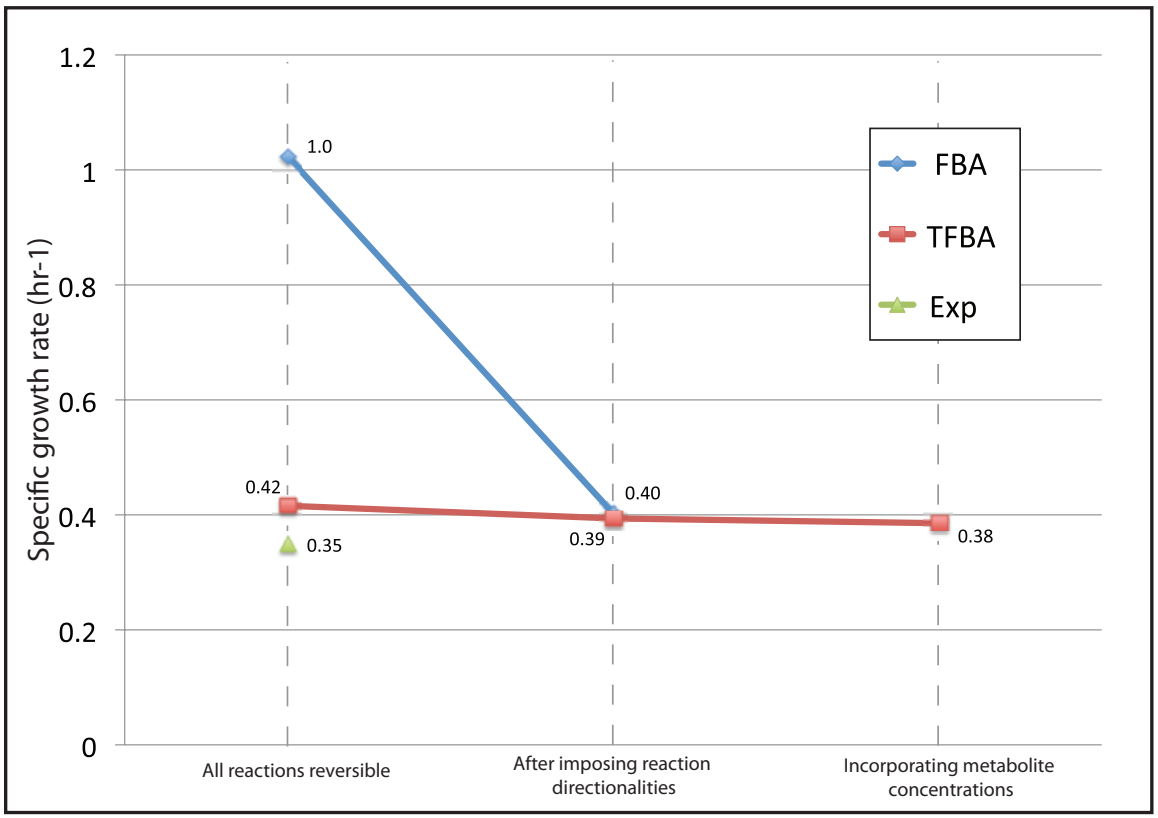
HXT, hexose transporters for glucose; *HXK*, hexokinase; *PGI*, glucose-6-phosphate isomerase; *PFK*, phosphofructokinase; *FBA*, fructose-biphosphate aldolase; *TPI*, triose phosphate isomerase; *TDH*, glyceraldehyde-3-phosphate dehydrogenase; *PGK*, phosphoglycerate kinase; *GPM*, phosphoglycerate mutase; *ENO*, enolase; *PYK*, pyruvate kinase; *ZWF*, glucose-6-phosphate-1-dehydrogenase; *RKI*, ribose-5-phosphate isomerase; *RPE*, ribulose-5-phosphate 3-epimerase; *TKL1*, transketolase; *TKL2*, transketolase; *TAL*, transaldolase; *PDC*, pyruvate decarboxylase; *ALD*, aldehyde dehydrogenase; *ACS*, Acetyl-CoA synthase; *CAT*, carnitine o-acetyltransferase; *ACARtrans*, acetylcarnitine diffusion; *YAT*, carnitine o-acetyltransferase; *CARtrans*, carnitine diffusion; *PYRtrans*, pyruvate carrier; *PDA*, pyruvate dehydrogenase; *PYC*, pyruvate carboxylase; *PCK*, phosphoenolpyruvate carboxylkinase; *OAtrans*, oxaloacetate carrier; *MAE*, malic enzyme; *CIT*, citrate synthase; *ACO*, aconitase; *IDH*, isocitrate dehydrogenase; *KGD*, a-ketoglutarate dehydrogenase; *LSC*, succinate-CoA ligase; *SDH*, succinate dehydrogenase; *FUM*, fumaratase; *MDH*, malate dehydrogenase; *NDH*, external NADH dehydrogenase; *NDI*, NADH dehydrogenase; *NDR*, NADPH reductase; *QCR*, ubiquinol cytochrome C reductase; *COX*, cytochrome C oxidase; *ASN*, ATP synthase; *AAC*, ADP/ATP carrier protein; *ADK*, adenylates kinase; *ATPmt*, ATP maintenance; *ADH*, cytosolic alcohol dehydrogenase; *SCD,* succinate dehydrogenase (ubiquinone-6), mitochondrial; *ACET*, acetate diffusion; *COH*, carbonic acid hydro-lyase; *PPP,* Pyrophosphate phosphohydrolase; *MLPIT,* malate transport, mitochondrial; *ICL,* Isocitrate

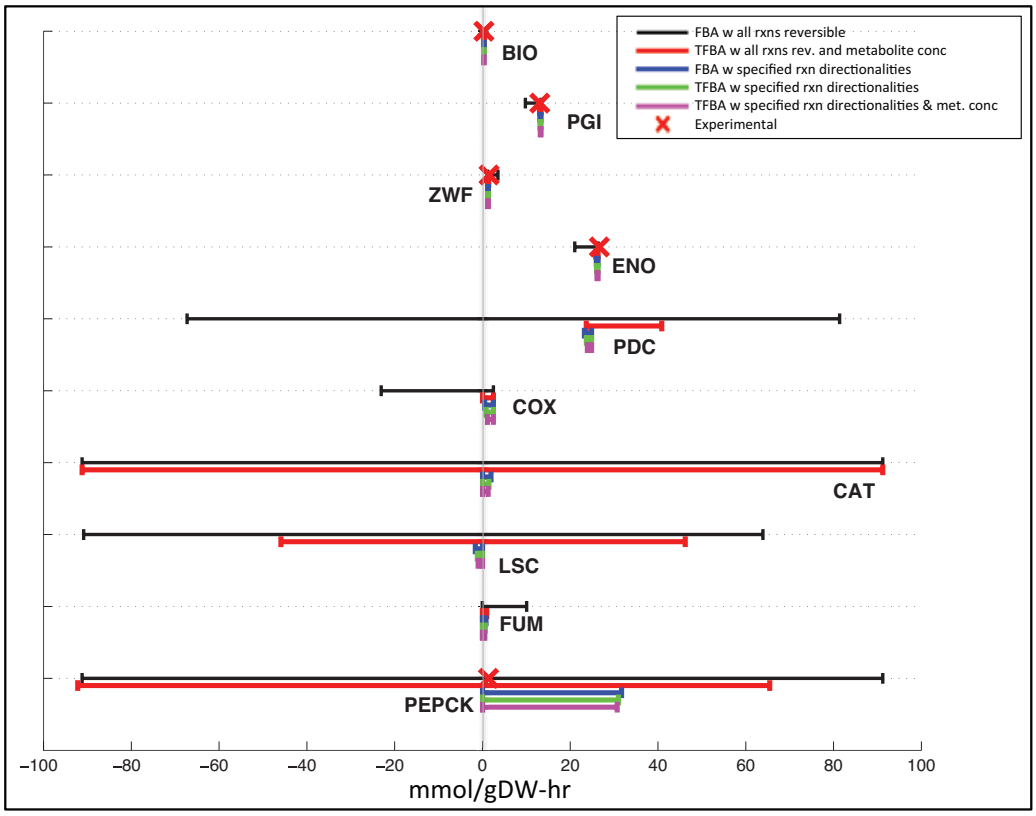
glyoxylate-lyase; *MLS,* L-Malate glyoxylate-lyase (CoA-acetylating); *MDHc,* (S)-malate:NAD+ oxidoreductase; *CITc,* Citrate oxaloacetate-lyase cytosolic; *ACOc*, citrate hydro-lyase cytosolic; *LACm2r,* D-lactate transport, mitochondrial; *CITt2m,* citrate transport, mitochondrial; *LDH,* (R)-Lactate:ferricytochrome-c 2-oxidoreductase; *O2m,* O2 transport (diffusion); *CO2m,* CO2 transport (diffusion), mitochondrial; *PIm,* phosphate transporter, mitochondrial; *CO2t,* CO2 transport via diffusion; *GLYCt,* glycerol transport in/out via diffusion reversible; *PYRst,* Pyruvate transport via proton symport; *SO4t,* sulfate transport via proton symport; *Pit,* phosphate transport via proton symport; *O2t,* O2 transport via diffusion; *NH4t,* Ammonia transport via diffusion; *LACt2r,* D-lactate transport via proton symport; *SUCCt2r,* succinate transporter in/out via proton symport; *MALt2r,* L-malate transport in via proton symport; *GND1,* 6-Phospho-D-glucono-1,5-lactone lactonohydrolase; *GND2,* 6-Phospho-D-gluconate:NADP+ 2-oxidoreductase (decarboxylating); *GPD1,* Glycerol-3-phosphate:NAD+ 2-oxidoreductase; *GPD2,* Glycerol-3-phosphate phosphohydrolase. The abbreviations for the chemical species: XL, xylose; XLT, xylitol; XYLL, xylulose; GLC, glucose; G6P, glucose-6-phosphate; F6P, fructose-6-phosphate; FBP, fructose 1,6-diphosphate; T3P, glyceraldehydes-3-phosphate; DHAP, glycerone phosphate; DPG, bisphosphoglycerate; 3PG, 3-phosphoglycerate; 2PG, 2-phosphoglycerate; PEP, phosphoenolpyruvate; PYR, pyruvate; 6PGL, glucono-1,5-lactone 6-phosphate; RL5P, ribulose 5-phosphate; R5P, ribose 5-phosphate; X5P, xylose-5-phosphate; E4P, erythrose 4-phosphate; S7P, sedoheptulose 7-phosphate; AALD, acetaldehyde; ACET, acetate; ACCOA, acetyl-CoA; CAR, carnitine; ACAR, acetylcarnitine; OAA,

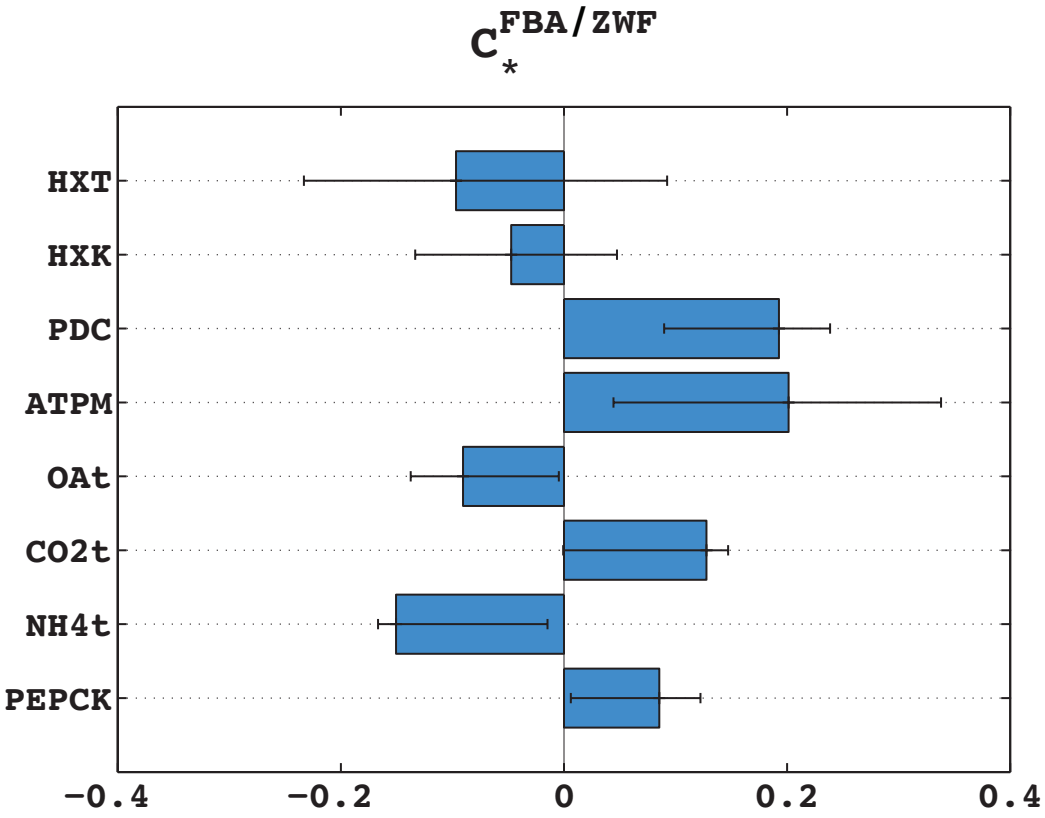
oxaloacetate; CIT, citrate; ICIT, isocitrate; AKG, 2-oxoglutarate; SUCCOA, succinyl-CoA; SUCC, succinate; FUM, fumarate; MAL, malate; GL, glycerol; ETH, ethanol; CO2; HCO3-; O2; PPi, Pyrophosphate; Pi, Phosphate; NH4; SO4; 6PGC, 6-Phospho-D-gluconate; GLYC3P, glycerol-3-phosphate; ACET, acetate; LAC, D-lactate; GLYX, glyoxylate.

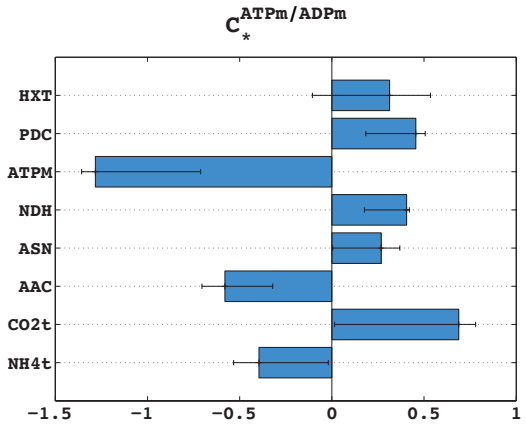

Stoichiometry
ΔG'°
Metabolomics
Fluxomics
Sample metabolites
Enzyme kinetics
Sample elasticities
Consistency check
No
Reject
Yes
MCA
Analyze results


Displacement from thermodynamic equilibrium
I
II
III
Near equilibrium
Far from equilibrium
glc[e]
HXT
biomass
BIO
glc
HXK
g6p
6pgl
6pgc
rl5p
GND1
GND2
ZWF
PGI
f6p
RPE
TKL2
x5p
e4p
TAL
s7p
RKI
r5p
PFK
fbp
FBA
TPI
t3p
TKL1
dhap
GPD1
TDH
dpg
PGK
3pg
GPM
2pg
ENO
pep
PEPCK
PYK
pyr
PYC
PYRm
pyr[m]
PDC
aald
ADH1
ALD
acet
ACS
accoa
ACE
acet[e]
ETOHt
etoh
etoh[e]
ACARt
acar
acar[m]
YAT
car
car[m]
CARt
coa
CAT
co2[m]
o2[m]
CO2m
O2m
co2
o2
COH
hco3
CO2t
O2t
co2[e]
o2[e]
Pim
pi[m]
pi
Pit
pi[e]
PPP
ppi
nh4
NH4t
nh4[e]
so4
SO4t
so4[e]
g3p
GPD2
glyc
GLYCt
glyc[e]
oaa
MDHc
mal
OAt
MLPIT
mal[m]
MAE
MDH
oaa[m]
PDA
accoa[m]
CIT
cit[m]
ACO
icit[m]
IDH
akg[m]
KGD
succoa[m]
LSC
suc[m]
SCD
fum[m]
FUM
SUCCmPI
ACOc
icit
CITc
clt
ICL
suc
glyx
MLS
AMP[m]
ADK
adp[m]
ATP[m]
AAC
ASN
ATPM
NDH
NDI
QCR
NDR
COX
L_LCAD2cm
qh2[m]
ferl[m]
fero[m]
q[m]
L_LACDr
lac
lac[e]
pyr
BIO
0.36
HXT
15.33
HXK
15.33
6pgl
1.24
6pgc
1.24
GND2
1.24
ZWF
GND1
PGI
13.42
RPE
0.71
TKL2
PFK
13.81
TAL
0.40
TKL1
0.52
FBA
13.81
TPI
dhap
12.62
GPD1
1.20
TDH
26.77
GPD2
1.20
GLYCt
1.20
PGK
26.77
GPM
26.60
ENO
26.60
PEPCK
1.40
PYK
27.82
PYC
2.82
PYRm
0.88
pyr[m]
1.26
PDC
23.92
ADH1
23.22
ALD
0.70
ETOHt
23.22
ACS
0.00
ACE
0.70
CAT
0.92
ACARt
0.92
YAT
0.92
CARt
0.92
CO2m
1.98
O2m
3.03
CO2t
26.58
O2t
3.14
Pim
6.88
Pit
0.11
PPP
0.21
NH4t
1.87
SO4t
0.02
ACOc
0.15
CITc
0.15
ICL
0.15
MLS
0.15
MDHc
0.05
OAt
0.33
MAE
0.38
PDA
0.33
CIT
0.33
ACO
0.33
IDH
0.33
KGD
0.33
LSC
1E-4
SCD
0.15
SUCCmPI
0.15
MLPIT
0.24
ADK
1E-4
AAC
6.50
ATPM
14.87
ASN
6.50
NDH
3.04
NDI
1.98
QCR
5.17
NDR
0.90
COX
3.03
L_LCAD2cm
1E-4
L_LACDr
1E-4
MDH
1E-4
FUM
0.15
SCD
0.15
LSC
1E-4
KGD
1E-4

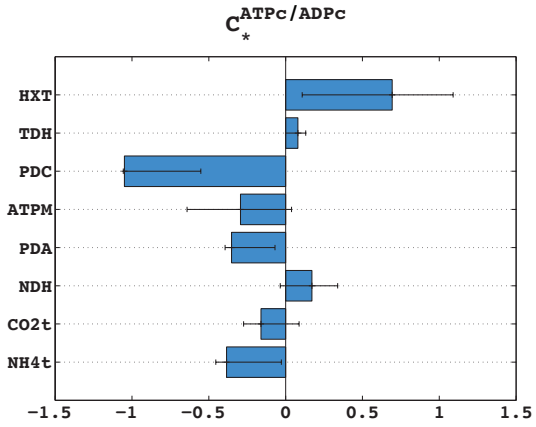

1.2
1
0.8
0.6
0.4
0.2
0
Specific growth rate (hr-1)
1.0
0.42
0.35
0.40
0.39
0.38
FBA
TFBA
Exp
All reactions reversible
After imposing reaction directionalities
Incorporating metabolite concentrations


FBA w all rxns reversible
TFBA w all rxns rev. and metabolite conc
FBA w specified rxn directionalities
TFBA w specified rxn directionalities
TFBA w specified rxn directionalities & met. conc
Experimental
BIO
PGI
ZWF
ENO
PDC
COX
CAT
LSC
FUM
PEPCK
−100
−80
−60
−40
−20
0
20
40
60
80
100
mmol/gDW-hr

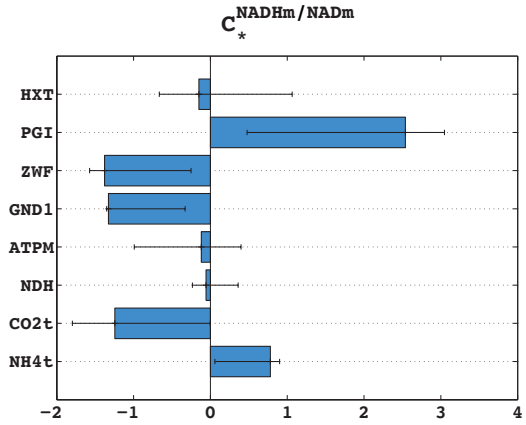

$C_*^{FBA/ZWF}$
HXT
HXK
PDC
ATPM
OAt
CO2t
NH4t
PEPCK
−0.4
−0.2
0
0.2
0.4


$C_*^{ATPm/ADPm}$
HXT
PDC
ATPM
NDH
ASN
AAC
CO2t
NH4t
−1.5
−1
−0.5
0
0.5
1

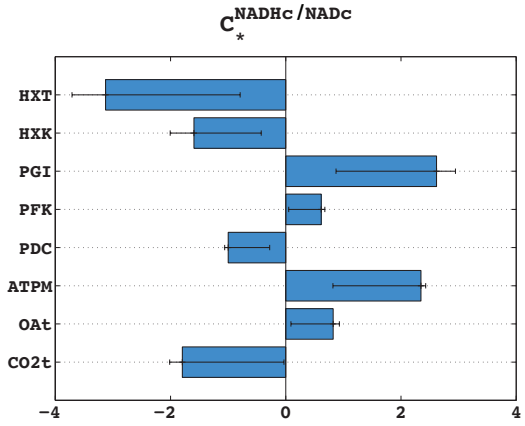
(A)


$C_*^{ATPc/ADPc}$
HXT
TDH
PDC
ATPM
PDA
NDH
CO2t
NH4t
−1.5
−1
−0.5
0
0.5
1
1.5

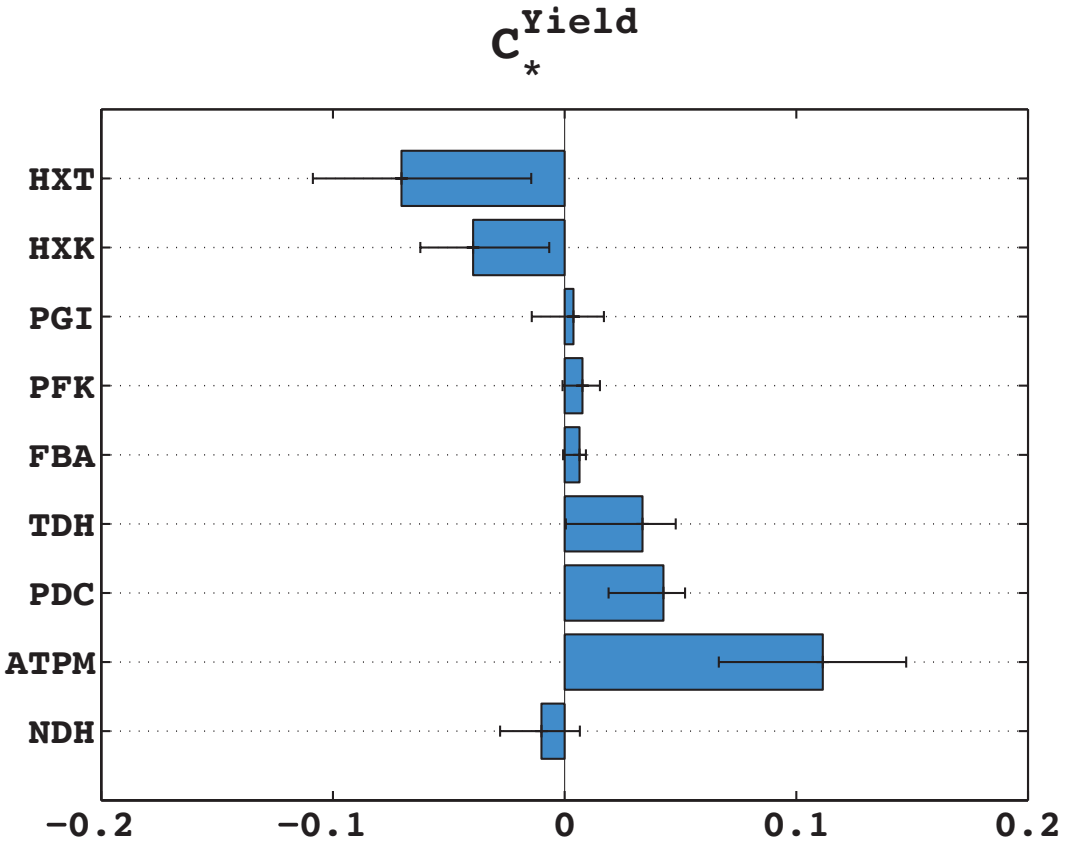
(B)


$C_*^{NADHm/NADm}$
HXT
PGI
ZWF
GND1
ATPM
NDH
CO2t
NH4t
−2
−1
0
1
2
3
4

(C)


$C_*^{NADHc/NADc}$
HXT
HXK
PGI
PFK
PDC
ATPM
OAt
CO2t
−4
−2
0
2
4

(D)


C^{Yield}_{*}
HXT
HXK
PGI
PFK
FBA
TDH
PDC
ATPM
NDH
−0.2
−0.1
0
0.1
0.2