Thesis for the degree Doctor of Philosophy

Reconstruction of Biological Networks for Integrative analysis

Reconstruction and Analysis of the Metabolic, Protein Secretion and Regulatory Network in Yeast

Tobias Österlund

Systems and Synthetic Biology
Department of Chemical and Biological Engineering
Chalmers University of Technology
Gothenburg, Sweden 2014

Reconstruction of biological networks for integrative analysis
Reconstruction and Analysis of the Metabolic, Protein Secretion and Regulatory Network in Yeast

Tobias Österlund

ISBN 978-91-7385-960-8

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr: 3641
ISSN 0346-718X

Department of Chemical and Biological Engineering
Chalmers University of Technology
SE-41296 Gothenburg
Sweden
Telephone +46 31 7721000

Printed by Chalmers Reproservice
Gothenburg, Sweden 2014

Reconstruction of biological networks for integrative analysis
Reconstruction and Analysis of the Metabolic, Protein Secretion and Regulatory Network in Yeast

Tobias Österlund
Systems and Synthetic Biology
Department of Chemical and Biological Engineering
Chalmers University of Technology, Sweden

Abstract

Biological systems can be very complex and consist of several thousand components that interact with each other in the cell. One of the goals of systems biology is to study biological systems from a systemic viewpoint in order to get an increased understanding of the behavior of the cell. Biological network reconstructions are important tools in systems biology in order to model the behavior of different biological systems. The biological networks can also be used as a scaffold for integrative analysis where high-throughput data from different conditions or different strains are integrated into the biological network to reduce the dimension of the data and to group the response between conditions or strains into biological pathways or key metabolites etc. The biological interpretation and discovery using integrative analysis can be facilitated by constructing more comprehensive and diverse biological networks.

In this thesis I expanded current biological network reconstructions for the yeast *Saccharomyces cereveisae* in three steps and used them as a scaffold for biological interpretation and discovery. First I constructed an up-to-date yeast genome-scale metabolic model. The model is a comprehensive description of yeast metabolism and contains more genes, reactions and metabolites than previous models. The model performs well in simulating the metabolism under different conditions. Second, I studied the transcriptional regulatory network of yeast in terms of topology and structure of the network and compared it to transcriptional regulation in *E. coli,* human and mouse. I also used high-throughput data from many different conditions to study the condition-dependent response of the yeast transcriptional regulatory network. Third, I was involved in reconstruction of models of the protein secretion machinery in *S. cerevisiae* and for the high protein producer *Aspergillus oryzae*, describing protein folding, post-translational modifications and protein transport etc. High-throughput data from several different strains producing α-amylase were integrated into the models in order to get an insight in the mechanisms and bottlenecks of protein secretion in these organisms.

The biological networks presented here were also used for data integration and the results and interpretation of the cellular behavior under different conditions can give us a deeper understanding and insight in for example condition-specific transcriptional regulation and protein production.

Keywords: Biological networks, integrative analysis, genome-scale metabolic model, transcriptional regulation, protein secretion

List of publications

The thesis is based on the following publications

I. **Österlund T**, Nookaew I, Nielsen J. (2012) Fifteen years of large scale metabolic modeling of yeast: developments and impacts. Biotechnol Adv, 30(5):979-988.

II. **Österlund T**, Nookaew I, Bordel S and Nielsen J. (2013) Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. BMC Systems Biology, 7(1):36

III. **Österlund T**, Bordel S, Nielsen J (2013) Controllability analysis of transcriptional regulatory networks reveals circular control patterns among transcription factors. (Manuscript)

IV. Liu Z, **Österlund T**, Hou J, Petranovic D and Nielsen J. (2013) Anaerobic alpha-amylase production and secretion with fumarate as the final electron acceptor in yeast. Appl. Environ. Microbiol., 79(9):2962-2967

V. Feizi A, **Österlund T**, Bordel S, Petranovic D and Nielsen J, (2013). Genome-scale modeling of the protein secretory machinery in yeast. PLoS ONE; 8(5):e63284

VI. Liu L*, Feizi A*, **Österlund T***, Hjort C, Nielsen J. (2013) Genome-scale analysis of the high-efficient protein secretion system of *Aspergillus oryzae*. (Submitted)

Additional publications not included in this thesis:

VII. Hou J, Tang H, Liu Z, **Österlund T,** Nielsen J and Petranovic D. (2013) Management of the endoplasmic reticulum stress by activation of the heat shock response in yeast. FEMS Yeast Res. DOI: 10.1111/1567-1364.12125

VIII. Hou J, **Österlund T,** Liu Z, , Petranovic D, Nielsen J. (2012) Heat Shock Response Improves Heterologous Protein Secretion in *Saccharomyces cerevisiae*. Appl. Microbiol. Biotechnol., 97(8):3559-3568

IX. Liu Z*, Liu L*, **Österlund T***, Hou J, Huang M, Fagerberg L, Petranovic D, Uhlen M and Nielsen J. (2013) Improved Production of Heterologous Amylase by Inverse Metabolic Engineering of Yeast. (Submitted)

X. Kristiansson E, **Österlund T**, Gunnarsson L, Arne G, Larsson DGJ and Nerman O. (2013) A novel method for cross-species gene-expression analysis. BMC Bioinformatics, 14:70

XI. van Grinsven K, Kumar R, **Österlund T**, Nielsen J, Teusink B, et al. Global analysis of stress response in *S. cerevisiae.* Manuscript in preparation

* Equal contribution

Contribution summary

Paper I	Compiled the figures and tables and wrote the manuscript.
Paper II	Reconstructed the model, analyzed the data, performed simulations, performed the random sampling and wrote the manuscript.
Paper III	Analyzed the data, constructed the networks, performed the controllability analysis and simulations and wrote the manuscript.
Paper IV	Performed transcriptome analysis and integrated analysis and edited the manuscript.
Paper V	Supervised the work and took part in writing the manuscript.
Paper VI	Performed transcriptome analysis and integrated analysis and wrote the manuscript.

Not included in thesis:

Paper VII	Performed transcriptome and integrated analysis and edited the manuscript.
Paper VIII	Performed transcriptome and integrated analysis and edited the manuscript.
Paper IX	Performed analysis of whole genome resequencing data and identification of mutations. Analyzed the transcriptome data, performed integrated analysis and took part in manuscript writing.
Paper X	Took part in development of the method and analyzed transcriptome data.
Paper XI	Performed transcriptome and integrated analysis and edited the manuscript.

Contents

Preface

This thesis is submitted for the partial fulfillment of the degree doctor of philosophy. It is based on work carried out between 2009 and 2013 at the Department of Chemical and Biological Engineering, Chalmers University of Technology, under the supervision of Professor Jens Nielsen. The research was funded by the Knut and Alice Wallenberg Foundation, Vetenskapsrådet, the European Research Council and the Chalmers Foundation.

Tobias Österlund

January 2014

1. Introduction

Biological systems are complex and involve many different processes such as metabolism, cell growth, cell division etc. The complexity of these processes arises from the interactions of several thousands of components, e.g. genes, proteins and metabolites (Sauer et al, 2007). The baker's yeast *Saccharomyces cerevisiae* is one of the most studied organisms and individual components (genes, enzymes, proteins) have been characterized (Botstein et al, 1997). To fully understand the behavior of complex systems it is beneficial to study the system as whole rather than individual components. The systemic properties includes the connectivity of the system, and how components interact with each other and trying to understand the systemic properties will help to understand the complexity of biological systems. The term "emergent property" applies also to biological systems and it describes how a complex structure or system can emerge from relatively simple interactions of the involved components (Ideker et al, 2001). One example of an emergent property is the symmetrical pattern of a snowflake where the interactions of water molecules give rise to a snow crystal.

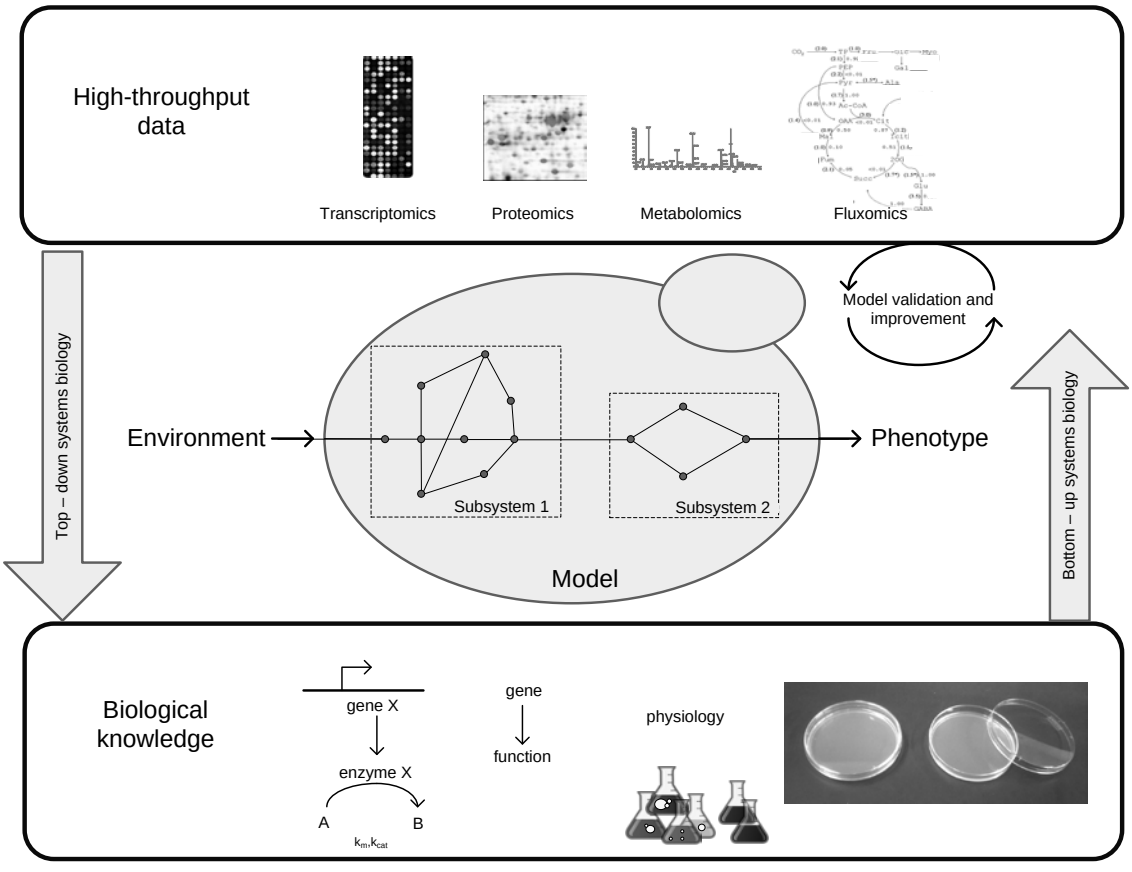


***Figure 1**- Systems biology uses mathematical modeling to be able to describe the biological system. Models that can predict the phenotype based on the genotype and the environment can be constructed from current biological knowledge (bottom-up systems biology) and knowledge can be created from analyzing high throughput data (top-down systems biology).*

Biological networks capture the systemic properties of the system of interest on the genomic level, and serves as a structured description of the system. Systems biology is a field where mathematical models and biological networks are used to describe biological phenomena. Figure 1 shows the concept of systems biology where models are used to capture the properties of the cell and describe the cellular behavior in different conditions or after perturbations. The genome plays an important role in defining the components of the system by defining genes and gene products. Other components can be small molecules, for example metabolites or metal ions. By using information about functionality of genes, gene products and pathways and retrieving information from publications and text books etc models can be constructed from biological knowledge (bottom part of Figure 1). This is called the bottom-up approach to systems biology (Palsson, 2006). Recent developments in experimental methods have generated large amounts of high-throughput omics data including transcriptomics, proteomics and metabolomics data. Transforming this data into biological knowledge or conclusions using mathematical models or techniques is called top-down systems biology.

Integrative analysis takes advantage of reconstructed biological networks and models by using the network structure and topology as a scaffold and using algorithms to integrate high-throughput data into the network. Many biological studies aim to understand the biology behind different diseases or effects of different cellular perturbations or stresses. Transcriptome analysis is commonly used to identify genes that are involved in the response and to find mechanisms that are likely to occur in the cell. Due to the large amount of data produced, analysis methods that make it easier to draw conclusions from the data are needed. Different methods exist, such as clustering methods that make it possible to characterize biological meaningful groups of genes with similar changes in expression, i.e. co-regulation (Eisen et al, 1998). However, these methods do not include any known information about the biological network.

To be able to extend the interpretation of the results from which genes that are involved in the response to which metabolites and pathways that are involved the information that a genome-scale reconstruction of the metabolic network contains can be included by integrating the transcriptome data into a genome scale metabolic model (GEM). The integrated analysis can be used as a tool to draw more general conclusions from the data in order to get as complete pictures as possible of the mechanisms going on in the cell.

Methods for integrating omics data into genome scale metabolic models exist, e.g. the Reporter Metabolite algorithm (Patil & Nielsen, 2005). However, in order to get an increased understanding of the biology behind different perturbations or conditions in the cell we need to have a comprehensive genome scale models that describes as many pathways and cellular mechanisms as possible.

The Reporter Metabolite algorithm was later extended to be able to integrate high-throughput data into other biological networks, e.g. transcriptional regulatory networks, KEGG pathways and GO-term networks (Oliveira et al, 2008). This makes integrated analysis a valuable tool both for reducing the dimensionality of the omics data by identifying interesting GO-terms or KEGG pathways that have key roles and

also for studying transcriptional regulation by using the topology and structure of the transcriptional regulatory network.

This thesis focuses on biological network reconstruction and analysis. If we can construct comprehensive biological networks that describe the current biological knowledge using a systemic viewpoint, i.e. capturing the interactions and complexity between components in the network, we can use these networks as a scaffold for integrative analysis in order to analyze data from a systemic point of view and get an increased understanding of the complex biological system.

The biological networks that are constructed and studied in this thesis are the following:

- The metabolic network
- The transcriptional regulatory network
- The protein secretion network

The metabolic network was constructed for *Saccharomyces cerevisiae* using previously constructed yeast metabolic networks as a starting point. The aims of this reconstruction were to construct a comprehensive description of yeast metabolism as possible and construct a high-quality genome-scale metabolic model that suits well for simulation of metabolism. The newly constructed yeast genome-scale metabolic model is called iTO977 and is presented in **Paper II**.

The concept of metabolic network reconstruction was applied to the protein secretion machinery of *S. cerevisiae* which resulted in a protein secretion machinery model that describes protein folding and processing in the Endoplasmic Reticulum and Golgi and protein sorting and transport within the cellular compartments. The model is presented in **Paper V**. This model is the first model describing the protein secretion machinery and the comprehensive structure of the model makes it useful as a scaffold for integrative analysis. A model of the protein secretion machinery in the high protein producing filamentous fungi *Aspergillus oryzae* was also constructed using the *S. cerevisiae* model as a starting point. This model was used to integrate transcriptome data from three α-amylase producing strains to study the response of protein secretion on the secretory machinery (**Paper VI**).

Both the metabolism and the protein secretory pathway are regulated at many different levels in the cell. One mechanism of regulation is transcriptional regulation where transcription factor regulate the activity of their target genes by identifying and binding to specific sequence motifs. The topology and structure of different transcriptional regulatory networks were analyzed using the concept of network controllability and the results are presented in **Paper III**. The condition-specific transcriptional regulation and response was studied for *S. cerevisiae* in **Paper II** and **Paper III**. Here we used transcriptome data from many different conditions and integrated into the different biological networks in order to identify transcriptionally controlled reactions and transcription factors that respond to environmental cues.

There are many applications of biological network reconstructions and different applications of genome-scale metabolic models for yeast are presented in **paper I.** In this thesis the focus is mainly on network reconstruction and the following two applications of biological network reconstructions.

- Using the networks and models as tools for improved strain construction for industrial applications, e.g. metabolic engineering and protein production
- Using the networks as a tool for biological interpretation and discovery.

In **paper IV** is anaerobic and aerobic protein production in yeast studied using data integration into the metabolic network. Here we use the integrative analysis for biological interpretation and discovery, but for an industrial application (to get improved protein production).

1.1. Thesis structure

The thesis is divided into five chapters. In Chapter 2 - Genome-scale metabolic modeling of yeast the background and results for **Paper I** and part of **Paper II** are presented. In Chapter 3 - Condition-specific regulation of yeast metabolism the background and results for part of **Paper II** and **Paper III** are presented. And finally, background and results for **Paper IV**, **Paper V** and **Paper VI** are presented in Chapter 4 - Systems biology for protein production. In Chapter 5 summary and perspectives of all the work is presented.

2. Genome-scale metabolic modeling of yeast

A genome-scale metabolic model (GEM) is a collection of metabolic reactions, compounds and genes that can be used to simulate the behavior of the cell. This chapter will focus on the history of genome-scale reconstructions of the *Saccharomyces cerevisiae* metabolism and describe the reconstruction of a new, updated genome-scale metabolic model for *S. cerevisiae* called iTO977.

2.1. Framework for genome-scale metabolic modeling and reconstruction

The sequencing of whole genomes in the 1990s opened up the possibility for genome-scale reconstruction of the metabolism of microorganisms in the post-genomic era (Schilling et al, 1999) and the first organisms with reconstructed genome-scale metabolic models were bacteria such as *H. influenzae* (Schilling & Palsson, 2000) and *E.coli* (Edwards & Palsson, 2000). The process of constructing a high-quality genome-scale metabolic model (GEM) is described in Figure 2. A step-by-step guide for high qualitative GEM reconstruction has been published in Nature Protocols (Thiele & Palsson, 2010).

2.1.1. Reconstruction of a draft metabolic network

The first step of the reconstruction is to generate a draft metabolic network which describes the relationship between genes, reactions and metabolites. The metabolic genes in the genome are annotated to metabolic functions using information from biochemistry, literature and databases, e.g. KEGG (Kanehisa et al, 2006) and BRENDA (Schomburg et al, 2002) or by homology to related organisms that already have genome-scale reconstructions. The term “genome-scale” indicates that the reconstruction covers all parts of metabolism, not only the central carbon metabolism. Each reaction in the reconstruction should have a reference in the literature (or database) and be connected to an enzyme commission number (E.C. number) and to one or more ORFs coding for a metabolic enzyme or transporter. Enzyme complexes or isoenzymes can be described as AND/OR relationships for each reaction in the metabolic network.

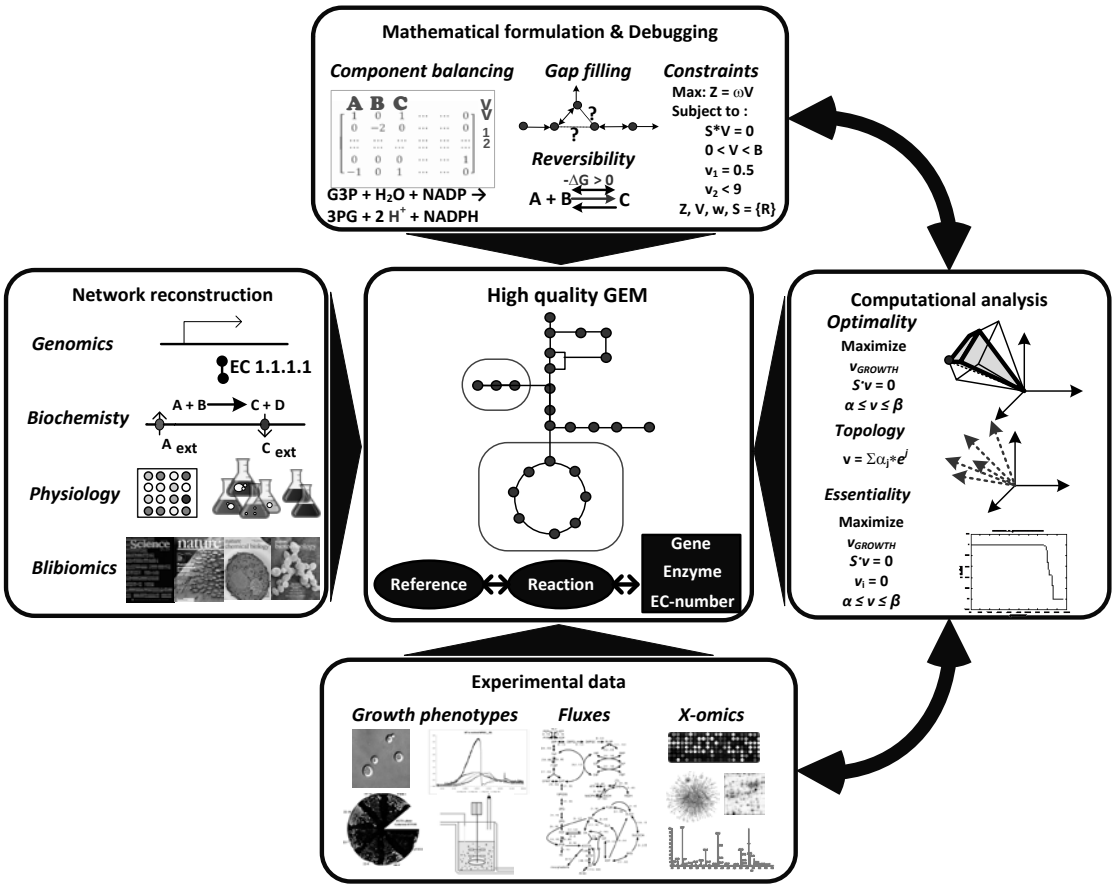


Figure 2 *Workflow for reconstruction of a high-quality genome-scale metabolic model, including network reconstruction, mathematical formulation, comparing simulations with data and model improvement.*

2.1.2. Mathematical formulation and simulation

The next step of the GEM reconstruction process is to convert the draft metabolic network to a mathematical model that can be used for simulations. This can be obtained by mass balancing around each of the intracellular metabolites in the model.

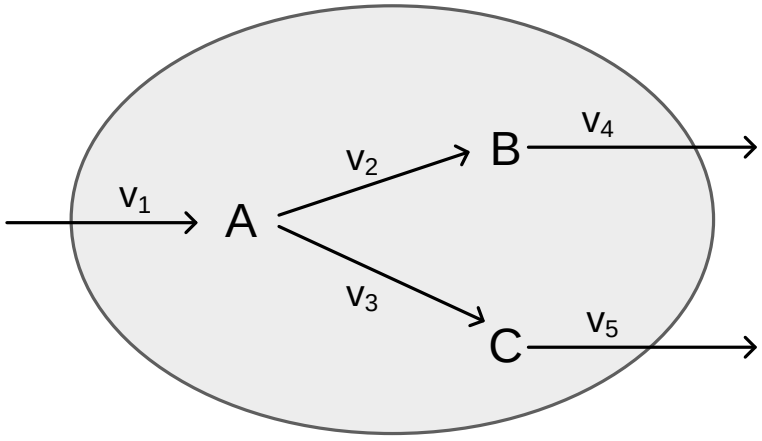


Figure 3 - Example illustrating a very simple metabolic network with 3 metabolites and 5 reactions; uptake reaction: v1, intracellular reactions: v2, v3, and excretion reactions v4 and v5.

Figure 3 shows an example of a metabolic network. The mass balance equation around metabolite A in this network can be written as:

$$\frac{d[A]}{dt} = v_1 - v_2 - v_3 - \mu[A] \qquad (1)$$

where v_1, v_2 and v_3 are the fluxes (reaction rates) of reaction 1, 2 and 3 respectively in the unit mol/(gDW*h), μ is the specific growth rate in h^{-1}, [*A*] is the concentration of metabolite A in mol/gDW and d[*A*]/dt is the (infinitesimal) change in concentration of metabolite A in mol/(gDW*h). The reaction rates of the reactions producing metabolite A (v_1) have a positive sign and the reaction rates of the reactions consuming metabolite A (v_2 and v_3) have a negative sign in the mass balance equation, reflecting the stoichiometry of the system. Similar mass balance equations can be set up for each of the internal metabolites in the system.

The term $-\mu[A]$ is called the dilution term and represents the fact that the cell volume increases as the cell grows, and hence the concentration of metabolite A will decrease proportionally with the growth rate. However, we assume that this effect can be neglected in biological systems, since the fluxes affecting the intracellular metabolite concentrations are normally much larger than the metabolite concentration itself (Stephanopoulos et al, 1998).

Another assumption often made when simulating genome-scale metabolism using for example Flux Balance analysis (FBA), is that we can ignore the dynamics of the system, i.e. we assume that the concentrations of the intracellular metabolites are in steady state and does not change over time. Normally the metabolism is a very fast process compared to other processes in the cell. Changes in metabolite concentrations occur faster than changes in environment or changes in growth. We can therefore assume that the concentrations of intracellular metabolites are in steady state, i.e. the metabolism will react relatively fast to changes in the environment (Varma & Palsson, 1994).

Following this two assumptions we get d[*A*]/dt=0 and the term $\mu[A]$ can be neglected, whereby the mass balance equation (1) in the example becomes,

$$0 = v_1 - v_2 - v_3 \qquad (2)$$

The mass balance equation for the whole system, using matrix notation, becomes,

$$0 = S_{in} \cdot v \qquad (3)$$

where S_{in} is the stoichiometric matrix for the intracellular metabolites with the rows representing metabolites and the columns representing reactions, and v is the flux vector. For the simple example in Figure 3 this becomes,

$$\boldsymbol{S_{in}} = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccccc} v_1 & v_2 & v_3 & v_4 & v_5 \end{array} \\ \begin{pmatrix} 1 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \end{pmatrix} \end{array} \quad (4)$$

$$v = (\quad v_1 \quad v_2 \quad v_3 \quad v_4 \quad v_5)^T \quad (5)$$

The metabolic model can be used to estimate the intracellular fluxes by solving the linear equation (3) in order to get the values for the fluxes in the vector *v*. However, normally for metabolic systems the number of reactions is higher than the number of metabolites, i.e. the S_{in} matrix has more columns than rows, as in the example. The number of unknown variables is higher than the number of equations, which leads to an underdetermined system of equations, which does not have one unique solution, but many possible solutions that satisfy equation (3). One way to come around this is to use an objective function that should be maximized or minimized while all the other intracellular fluxes remain in steady state. For microorganisms such as bacteria or yeast the objective function is often set to maximize the growth (or biomass production) of the cell, following the assumption that microorganisms have evolved to grow as fast as possible (Ibarra et al, 2002). However, a recent study suggest that the flux distribution of microorganisms subject to several, competing cellular objectives (Schuetz et al, 2012). Flux balance analysis (FBA) (Varma & Palsson, 1994) can be formulated as following,

$$\begin{cases} \text{Maximize } c \cdot v, \\ \text{subject to,} \\ S_{in} \cdot v = 0, \\ v_i^{min} \leq v_i \leq v_i^{max} \end{cases} \quad (6)$$

where *c* is the vector of objective functions and v_i^{min} and v_i^{max} are additional constraints (upper and lower bounds) for each reaction rate. There are many other different methods for analysis of GEMs which have been reviewed by Lewis et al. (2012). Methods that does not require an objective function includes random sampling over the space of feasible steady state solutions (Bordel et al, 2010) and topology-based methods, e.g. Elementary Flux Modes (Schuster & Hilgetag, 1994)

2.1.3. Quality control and model validation

Once the draft metabolic network has been converted to a mathematical model it is possible to use the simulation framework to find errors in the model. One important step is so-called gap-filling where dead-end reactions and orphan reactions (reactions without gene associations) can be identified, and the if there exist a missing reaction or gene this can be inserted to the model (Orth & Palsson, 2010). The RAVEN toolbox (Agren et al, 2013a) includes methods for quality control and gap-filling. The toolbox also offers a framework for automatic reconstruction of GEMs. Given the protein sequences for the organism of interest in FASTA format the toolbox can generate a

draft metabolic model by searching for similarities with genes in GEMs of closely related organisms, or enzymes coding for reactions in KEGG.

The simulation capabilities of the model can be validated by comparing simulations with experimental data. The optimal flux distribution obtained from model simulations can be compared to experimentally measured flux distributions or phenotypic data, e.g. growth rate, glucose uptake rate, product formation rate etc. (Pramanik & Keasling, 1997; Price et al, 2004). The model reconstruction process is an iterative process where the model is improved iteratively until the agreement with experiments is good.

2.2. History of yeast genome-scale modeling

During the last 10 years the yeast genome-scale metabolic network has been updated and improved leading to several large-scale reconstructions of the metabolism. The history and impact of genome-scale metabolic modeling of yeast is described in **Paper I**. An updated and comprehensive genome-scale model of yeast metabolism called iTO977 is presented in **Paper II**.

The first genome-scale metabolic model for *Saccharomyces cerevisiae* was also the first metabolic network reconstruction for an Eukaryotic organism (Forster et al, 2003). The model was named iFF708 where FF stands for the authors, Förster and Famili, and 708 is the number of genes included in the model. After the first reconstruction was made several updated versions followed which were all based on iFF708. Table 1 shows the different yeast GEMs published to date including number of compartments, reactions, metabolites and genes.

Table 1 – Genome-scale metabolic models for yeast

Name	Scope	Comps	Rxns	Mets	Genes	Reference
iFF708	First genome-scale model	3	1145	825	708	(Forster et al, 2003)
iND750	8 compartments	8	1149	646	750	(Duarte et al, 2004)
iLL672	Model reduction	3	1038	636	672	(Kuepfer et al, 2005)
iMH805/775	Transcript. Regulation	8	1149	646	775	(Herrgård et al, 2006)
iIN800	Lipid metabolism	3	1446	1013	800	(Nookaew et al, 2008)
iMM904	Applied metabolome	8	1577	1228	904	(Mo et al, 2009)
Yeast 1.0	Consensus network	15	1761	1168	888	(Herrgard et al, 2008)
Yeast 4	Updated consensus GEM	16	1865	1398	932	(Dobson et al, 2010)
Yeast 5	Updated consensus GEM	16	2110	1655	918	(Heavner et al, 2012)
iTO977	Comprehensive	4	1566	1353	977	(Österlund et al, 2013)
Yeast 7	Updated fatty acid metabolism	16	1882	1454	901	(Aung et al, 2013)

Footnote: Yeast 2, Yeast 3 and Yeast 6 are intermediate versions of the yeast consensus model and were released online at http://www.comp-sys-bio.org/yeastnet/

However, the many different versions of the yeast model lead to a problem. The models differed in scope, in the way they were constructed, and they also had different naming of metabolites which made comparison and merging of different models problematic. To solve this several groups was meeting for a yeast metabolism jamboree in Manchester 2008 to construct a consensus yeast metabolic network. The yeast consensus network is called Yeast 1.0 and was also introducing a standard for naming and annotation of metabolites using e.g. ChEBI identifiers (Degtyarenko et al, 2008) and InChI codes (Coles et al, 2005). The Yeast 1.0 network was not a GEM

ready for simulations due to missing information about reversibility of some reactions and missing biomass equations. Therefore it was further updated and yeast 4 was a GEM that also worked for simulations. Yeast 5 and 7 are updated versions of the yeast 4 model, and iTO977 is based on Yeast 1.0 and the iIN800 model.

After the first yeast genome-scale model was reconstructed it was started to be applied for modeling yeast cells for different purposes. In **Paper I** we have divided the use of yeast GEMs into four different application categories which will be described more in detail in this section. Figure 4 shows the cumulative number of publications that uses yeast genome-scale modeling between years 2003-2010. Several publications have showed the successful use of genome-scale metabolic modeling to suggest strategies for strain improvement in metabolic engineering (application category 1). Cases where the product yield has been improved as a result of metabolic modeling include for example ethanol production (Bro et al, 2006), succinic acid production (Agren et al, 2013b), vanillin production (Brochado et al, 2010) and sesquiterpene production (Asadollahi et al, 2009).

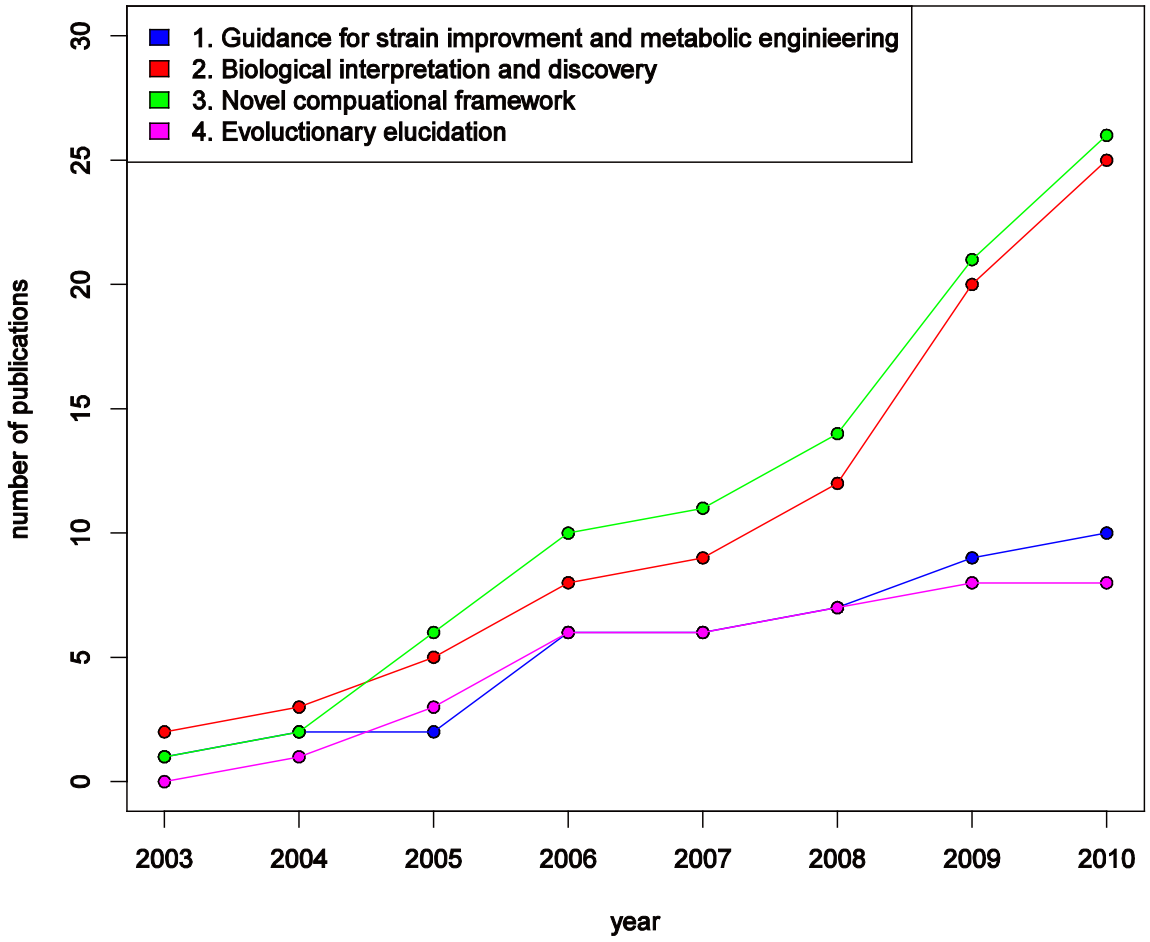


Figure 4 *The cumulative number of publications applying yeast genome-scale modeling in each of the four different application categories between 2003-2010. Data from Österlund et al. (2012).*

The second category uses GEMs as a tool for biological interpretation and discovery. The reporter metabolite algorithm (Patil & Nielsen, 2005) allows integration of

transciptome data from microarray or RNA sequencing experiments into the metabolic network to identify metabolites around which the most significant transcriptional changes occur. Several studies have used data integration in order to investigate how different perturbations to the cell (for example gene knockout or over-expression) influence the metabolism (Cimini et al, 2009; Papini et al, 2010).

The third category includes using yeast models for testing new computational methods and the fourth category uses yeast models to study evolution, e.g. function of gene duplication (Kuepfer et al, 2005) or gene evolution from bacteria to yeast (Mahadevan & Lovley, 2008).

2.3. iTO977 – an updated model of yeast metabolism

The iTO977 genome-scale metabolic model was constructed using the consensus network (Yeast 1.0) and the iIN800 model as starting points to make a draft metabolic network. The aim of the model was to construct a more comprehensive description of the yeast metabolism that also suits well for simulations. The iTO977 model is presented in **Paper II**. Figure 5A shows the relationships between the different yeast models and the construction of iTO977.

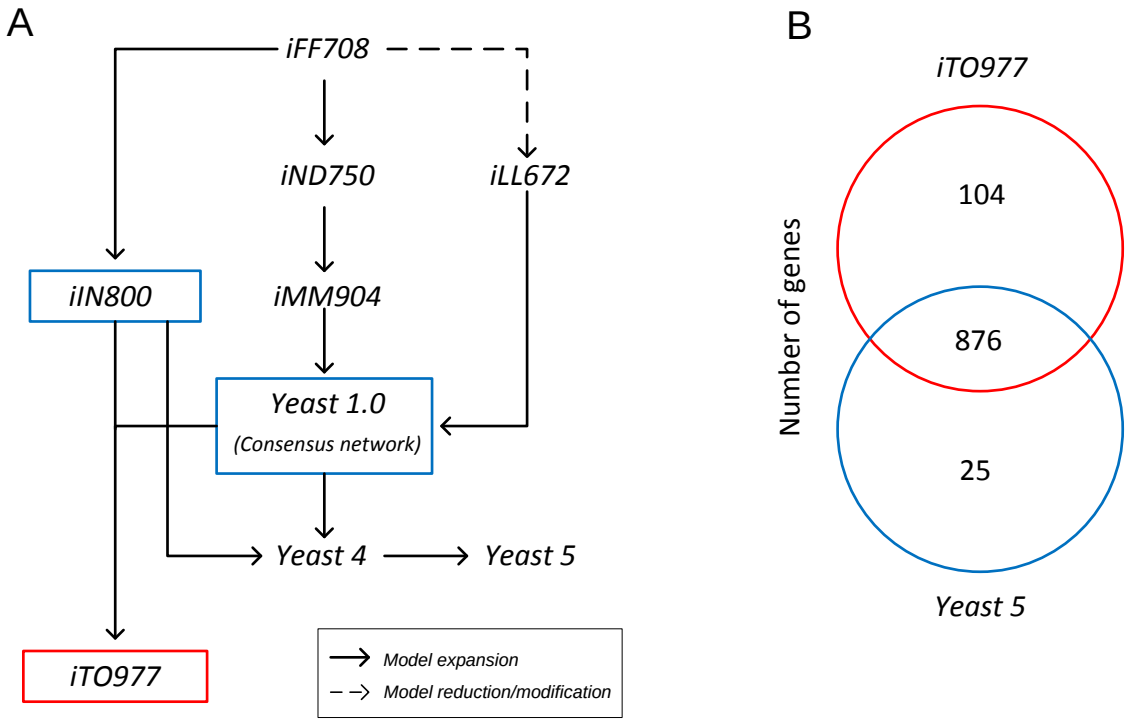


Figure 5 *– (A) Pedigree showing the relationship between some of the yeast models. ITO977 was based on the consensus network and the iIN800 model. (B) The venn diagram shows the overlap of genes included in iTO977 and the Yeast 5 model.*

Information about reversibility was added to some of the reactions in the Yeast 1.0 network during the reconstruction process. New reactions and pathways were added to the draft network after merging the two starting-point reconstructions in order to get a comprehensive description of the yeast metabolism. Gap filling and quality control

steps were implemented using the RAVEN toolbox (Agren et al, 2013a). Further, the simulation capabilities of the model was ensured both by checking that the model was functional, i.e. able to produce all metabolites and grow on different substrates when running *in silico* simulations, and by comparing simulations to experimental data.

A summary of the number of genes, metabolites and reactions of the iTO977 model compared to other yeast models is present in Table 1 on page 9. Figure 5B shows a comparison in terms of number of genes between the iTO977 model and the Yeast 5 model (Heavner et al, 2012). The iTO977 model includes 104 additional ORFs and many of the genes included in iTO977 but not in Yeast 5 belongs to newly added reactions and pathways, e.g. the biosynthesis of lipid-linked oligosaccharides (Burda & Aebi, 1999) and glycosylphosphaditylinositol (GPI) biosynthesis (Grimme et al, 2001). These two example pathways included in the model will make it easier to merge the metabolism model with the model of protein secretion (Feizi et al, 2013) which is presented in section 4.2.

Another difference between the iTO977 model and the Yeast 5 model is the number of compartments represented in the model. The iTO977 model has 4 compartments, namely cytoplasm, mitochondria, peroxisome and extracellular space, while the Yeast 5 model includes 15 different compartments. One advantage with including fewer compartments in the model is that the connectivity is improved without the need of including transport reactions, and the complexity of the model is reduced. This will improve the simulation capabilities of the model. Also, the four compartments included in the iTO977 model are the compartments with the highest confidence of localization of the enzymes in the *Saccharomyces* genome database (SGD) (Cherry et al, 2012).

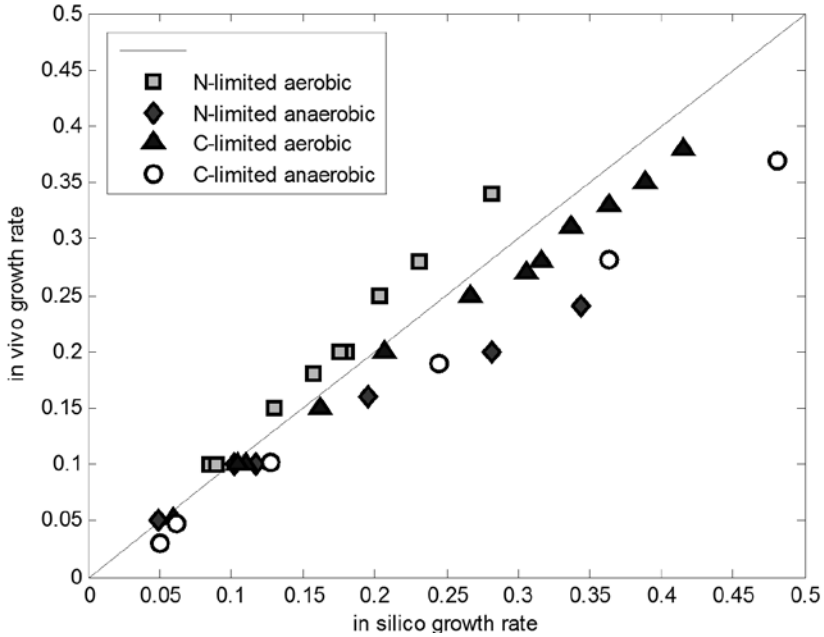


Figure 6 *– Comparison of simulated growth rate (x-axis) and experimental growth rate (y-axis). Simulations and experiments were carried out under 4 different conditions, nitrogen and carbon limitation and aerobic and anaerobic growth conditions.*

The iTO977 model was validated in two different ways when it comes to simulations. The first validation was the ability of the model to predict the growth rate in different conditions. Figure 6 shows the experimental and simulated growth rate under four different conditions.

Experimental data was taken from chemostat cultures where the metabolism is in steady-state and the cells are controlled to grow with the same rate – the dilution rate. Data was collected from chemostats with different dilution rates and under four different conditions: carbon limited aerobic (Bakker et al, 2000; Gombert et al, 2001; Jewett et al, 2013; Overkamp et al, 2000; Vemuri et al, 2007), nitrogen limited aerobic (Aon & Cortassa, 2001; Tai et al, 2007; Usaite et al, 2006; Vemuri et al, 2007) carbon limited anaerobic (Nissen et al, 1997; Tai et al, 2007) and nitrogen limited anaerobic (Lidén et al, 1995; Tai et al, 2005). The growth rate was maximized as an objective function for the simulations and the experimentally measured values for glucose uptake rate, ammonium uptake rate and oxygen uptake rate was used to constrain the model according to each experiment.

The second type of validation that was performed was the ability of the model to predict viability, i.e. growth of single and double gene knock-outs. Figure 7 shows the result of the gene deletion analysis.

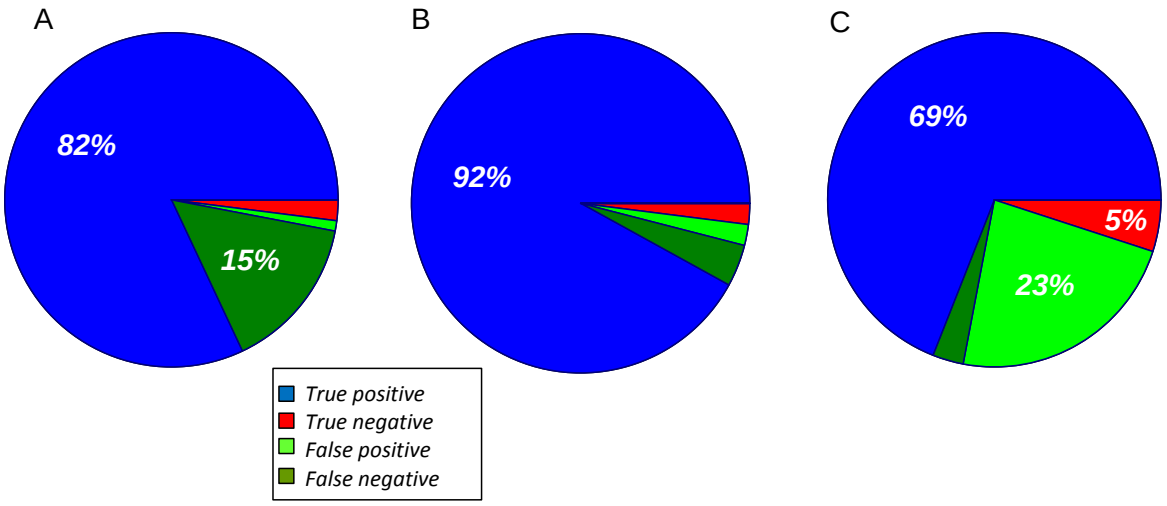


Figure 7 – *Performance of prediction of viability of single and double gene knockouts using the iTO977 model. The simulated growth phenotypes were compared to experimental growth phenotypes. (A) Single gene knock-outs, minimal media. (B) Single gene knock-outs, rich YPD media. (C) Double gene knock-outs, rich YPD media.*

In this analysis we compared the growth phenotype or fitness for simulated knock-outs with experimental values for single gene knock-outs (Cherry et al, 2012; Förster et al, 2003; Mo et al, 2009) and double gene knock-outs (Costanzo et al, 2010). The gene deletion was simulated by constraining the flux of all reactions associated with the

deleted gene to zero. The mutation was considered as lethal if the growth rate was reduced with 10% or more as compared with the wild type growth rate. The simulations were performed both in a simulated minimal media allowing only uptake of glucose, ammonium, phosphate, sulfate and oxygen and in a simulated rich YPD media where also amino acid uptake and nucleotide uptake was allowed.

3. Condition-specific regulation of yeast metabolism

False predictions using genome-scale metabolic models can be due to missing information about regulation in the models, i.e. the models assume that all enzymes are present at all times, in sufficient amounts to catalyze the reactions. In this chapter different types of regulation is discussed, and different methods for integrating regulation into genome-scale metabolic models are presented. Further, two examples are presented where transcriptional regulation in yeast is investigated under different growth conditions (**Paper II** and **Paper III**).

3.1. Regulation of metabolism

Understanding how the cell regulates different processes and especially metabolism is important for understanding most biological systems. However, understanding regulation is complicated, and regulation gets more complicated with increased complexity of the organism. Here different types of regulation and different ways to model regulation are presented.

3.1.1. Different types of regulation

Regulation of enzyme activity in yeast and other Eukaryotes can occur at many different levels in the cell. Here are three examples:

a) **Transcriptional regulation:** i.e. transcription factors that regulate the transcription of a gene. DNA-binding transcription factors can recognize and bind to specific sequence motifs, so called transcription factor binding sites, which can be located upstream of the coding sequence (Alberts et al, 2007). When the transcription factor is bound to the transcription factor binding site it can either activate or repress the transcription of the gene by interacting with for example the RNA polymerase complex (Hahn & Young, 2011).

b) **Post-translational modifications:** Signaling pathways can include e.g. phosphorylation (post-translational modifications) of a protein in order to activate it. Ubiquitination is another type of post-translational modification where a small protein, ubiquitin, is attached to the protein to regulate its activity. Acetylation is a reversible process where an acetyl group is attached to the protein. Examples of proteins that can be acetylated are histones and tubulines (Jensen, 2006). Performing post-translational modifications is generally a faster way for the cell to regulate the activity of enzymes and proteins than by transcriptional regulation.

c) **Enzyme level regulation:** Experimental assays can be set up to measure enzyme kinetics for a specific enzyme. Often one can simplify the kinetics of enzyme-catalyzed reactions with irreversible Michaelis Menten kinetics where the enzyme activity is dependent of the substrate concentration and the enzyme concentration according to Equation 7:

$$v = v_{max} \frac{[S]}{K_m + [S]}. \quad (7)$$

Different isoenzymes can have different K_m and v_{max} values. The rate v_{max} is the maximum reaction rate, i.e. when all enzymes are saturated with substrate. The constant K_m corresponds to the substrate concentration that gives the reaction rate equal to $v_{max}/2$. The kinetic parameters v_{max} and K_m can be measured experimentally for the enzyme of interest by having a known amount of enzyme and substrate, and then measure the product at different time points (Stephanopoulos et al, 1998).

In this chapter the focus is mainly on transcriptional regulation.

3.1.2. Modeling regulation of metabolism

Different approaches have been taken to integrate transcriptional regulation into yeast metabolism. Table 2 shows different methods for integrating regulation into genome-scale metabolic models in order to improve the prediction power of the simulations. Most of the methods implement the transcriptional regulatory network as Boolean rules.

Table 2 – Methods for integrating regulation into genome-scale metabolic models

Name	Reference	Scope
Regulatory FBA (rFBA)	Covert et al. (2001)	Constrain reactions in FBA simulations (Boolean rules)
Genetically-constrained metabolic flux analysis	Cox et al. (2005)	Constrain reactions in FBA simulations (Boolean rules)
R-matrix	Gianchandani et al. (2006)	Constrain reactions in FBA simulations (Boolean rules)
Steady-state rFBA (SR-FBA)	Shlomi et al.(2007)	Constrain reactions in FBA simulations using MILP
Integrated dynamic FBA (idFBA)	Lee et al. (2008)	signaling, regulation and metabolism
Integrated FBA (iFBA)	Covert et al. (2008)	signaling, regulation and metabolism
E-flux	Colijn et al. (2009)	constrain reactions directly from OMICS data (no transcriptional regulatory network is needed)
GeneForce	Barua et al. (2010)	correct too stringent Boolean regulatory rules
Probabilistic Regulation of Metabolism (PROM)	Chandrasekaran & Price (2010)	Constrain reactions in FBA simulations using probabilistic rules (not Boolean)
TIGER	Jensen et al. (2011)	Constrain reactions in FBA simulations using MILP, correct too stringent Boolean regulatory rules

In addition to these methods which tries to predict the state of the transcriptional regulatory network and integrate it to existing genome-scale metabolic models there are also several methods that uses high-throughput information from transcriptomics, metabolomics and proteomics studies and use this data in the network reconstruction process in order to construct context-specific models. These methods have for example been used successfully to predict tissue-specific models of human metabolism using a generic human metabolic network as a starting point (Agren et al, 2012; Becker & Palsson, 2008; Jensen & Papin, 2011; Shlomi et al, 2008; Yizhak et al, 2010).

The Boolean modeling approach, applied in rFBA and other methods, model each TF as either on or off (1 or 0) depending on environmental factors, for example extracellular metabolites that can be either present or absent, or on the state of other TFs (Covert et al, 2001). The formulation of the rFBA problem where the regulation is implemented as Boolean rules is stated in Equation 8.

$$\begin{cases} \text{Maximize } c^T v \\ \text{Subject to} \\ \quad - \; S_{in} v = 0 \\ \quad - \; v_i^{min} y_i \leq v_i \leq v_i^{max} y_i \\ y_i \in \{0,1 \\ y_i = f(X_{ext}, TF) \end{cases} \qquad (8)$$

Figure 8 shows how the solution space is reduced when introducing regulatory constraints. The regulatory network can be reconstructed based on ChIP-chip experiments (Iyer et al, 2001; Lieb et al, 2001; Ren et al, 2000) that reveal TF-DNA interactions. For yeast there are several sources of TF-DNA interaction data available (Harbison et al, 2004; Lee et al, 2002; Teixeira et al, 2006).

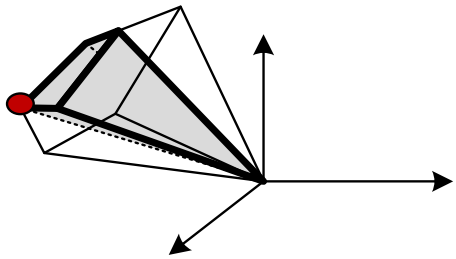


Figure 8 – *Concept drawing of regulatory FBA (rFBA), which uses regulatory rules as additional constraints to the FBA simulation in order to reduce the space of possible steady-state solutions and improve the prediction power of the model. The white hypercone in this figure represents the solution space (space of feasible solutions) and the shaded area represents the solution space after introducing regulatory constraints.*

The rFBA framework has been implemented both for *E. coli* (Covert et al, 2004) and *S. cerevisiae* (Herrgård et al, 2006) where the Boolean rules were constructed from a large number of publications and from TF-DNA interactions from ChIP-chip experiments.

One problem with the representation of the TF regulation as Boolean is that it is simplifying the transcription of a gene to be either 0 or 1, which is not always the case. Another observation is that introduction of Boolean constraints to the reactions in the FBA simulations can be too conservative and lead to unfeasible solutions, i.e. the model will predict growth if a reaction is modeled as off, but the cell should still be able to grow according to experiments. (Jensen et al, 2011). To come around these problems Chandrasekaran & Price (2010) developed the Probabilistic Regulation Of Metabolism (PROM) framework where the probability of TF binding for a TF-target gene pair is estimated from a large number of microarray experiments. Instead of having a value of the TF regulation of either 0 or 1 the probability of TF binding is continuous between 0 and 1. The method requires setting a threshold value in the microarray experiments saying if a gene is expressed or not.

3.2. Identification of transcriptionally controlled reactions using random sampling

The response of metabolism to different perturbations, e.g. gene deletions or change of environmental conditions may occur at different levels in the cell. As an example we have a reaction v_1 which converts metabolite A to metabolite B:

$$A \xrightarrow{v_1} B \qquad (9)$$

The reaction in Equation 9 is catalyzed by enzyme X. Assume that we can measure the change in transcription of enzyme X between condition 1 and condition 2 and also measure the change in flux of the reaction (v_1) between the two conditions. We then have the following possible scenarios:

a) The expression of enzyme X changes significantly between condition 1 and condition 2, and the flux v_1 changes significantly in the same direction (i.e. up-regulated or down-regulated) between the two conditions. We say that the reaction is *transcriptionally controlled.*
b) The expression of enzyme X change significantly, but the flux of the reaction does not change
c) There is no significant change in expression, but the flux of the reaction change significantly between the two conditions.

In scenario b and c the metabolic flux might be post-translationally controlled and metabolically controlled, respectively.

In metabolic engineering suitable over-expression targets (enzymes) should increase the flux of a reaction and direct the flux towards a desired product. Therefore it is good to find reactions of type a) above where the flux of the reaction can be increased by over-expressing the enzyme (transcriptionally controlled reactions).

Bordel et al. (2010) introduced a method for identification of these transcriptionally controlled reactions by comparing measured change in transcription with estimated intracellular fluxes using random sampling of the solution space. In **paper II** we used this algorithm (Bordel et al, 2010) together with high-throughput data from different conditions in order to find transcriptionally changed reactions in the iTO977 genome-scale metabolic model.

The workflow followed the following pattern:

- Transcriptome data (microarrays) from 24 independent chemostat cultivations were collected from a previous study (Jewett et al, 2013). The cells were grown in a controlled environment under aerobic, anaerobic, carbon limited and nitrogen limited conditions.
- The data was analyzed and genes that were regulated as a function of aerobic vs. anaerobic conditions and N-limited vs. C-limited conditions were identified.
- Extracellular flux measurements (Jewett et al, 2013) was used to constrain the iTO977 model to represent the four different conditions, aerobic, anaerobic, N-limited and C-limited. The CO_2 production rate was not constrained.
- Random sampling of the solution space gives a mean and standard deviation of the flux for each reaction in each condition. The change in flux of a reaction can be estimated between conditions and compared to the change in expression of the enzymes catalyzing that reaction.
- Transcriptionally controlled reactions are identified, i.e. reactions where flux and transcription change in the same direction.

Transcriptionally controlled reactions were determined for the aerobic-anaerobic and C-limited vs. N-limited comparisons and the results are shown in Figure 9. To test if the transcriptionally controlled reactions had any common regulators we performed a hypergeometric test where overrepresented TF regulations were identified among the significantly changed genes (adjusted p-value <0.05). The results for the hypergeometric test are presented in Table 3.

Table 3 – Over-represented transcription factors in the different comparisons

Comparison	TF	Hypergeometric p-value
Anaerobic - Aerobic (C-limited)	Opi1p	0.0009
	Pip2p	0.0168
	Gis2p	0.0278
Anaerobic - Aerobic (N-limited)	Opi1p	0.0215
C-limited - N-limited (Anaerobic)	Opi1p	0.0070
C-limited - N-limited (Aerobic)	Yap7p	0.0018
	Opi1p	0.0021
	Dal80p	0.0153
	Dig1p	0.0290

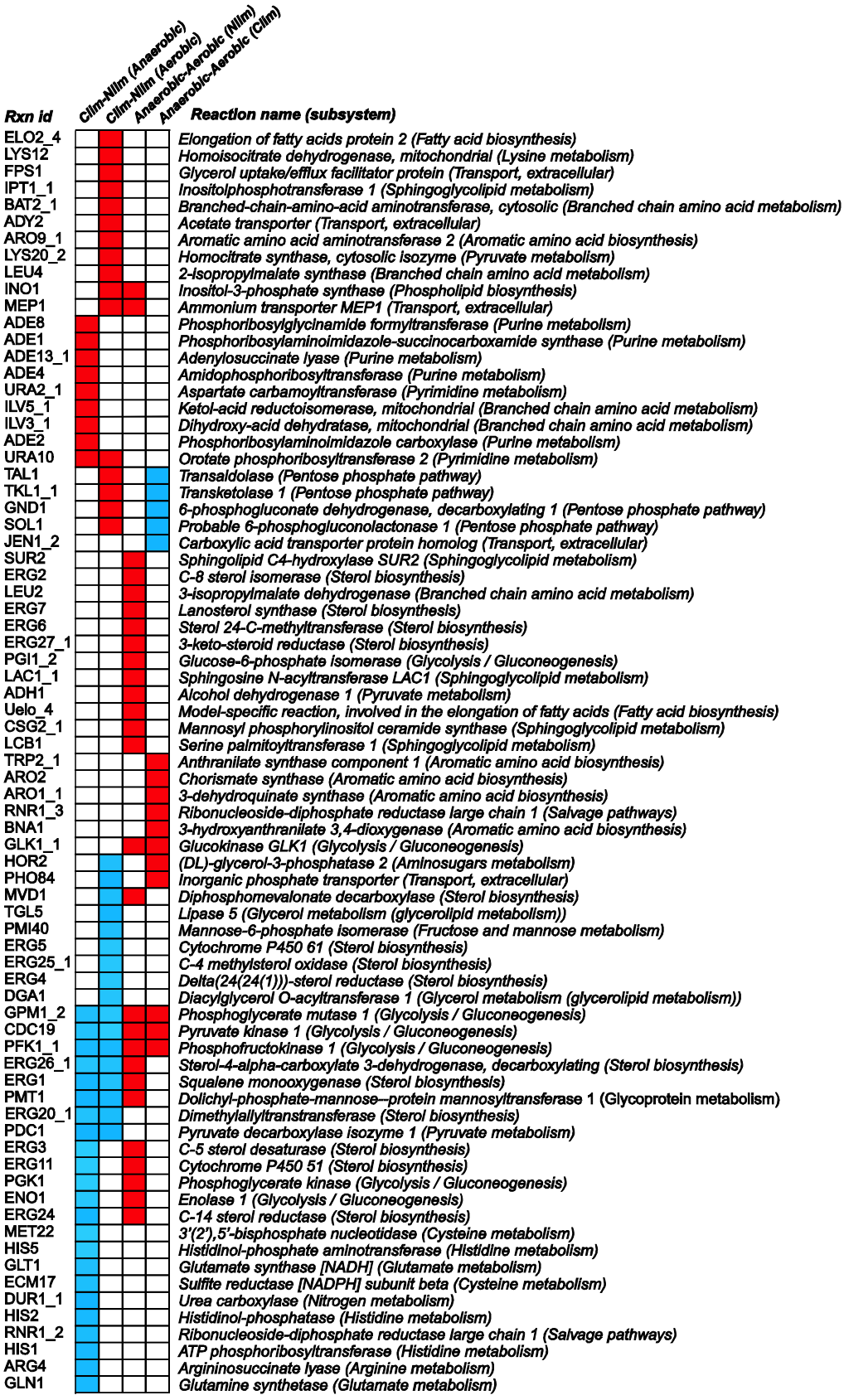
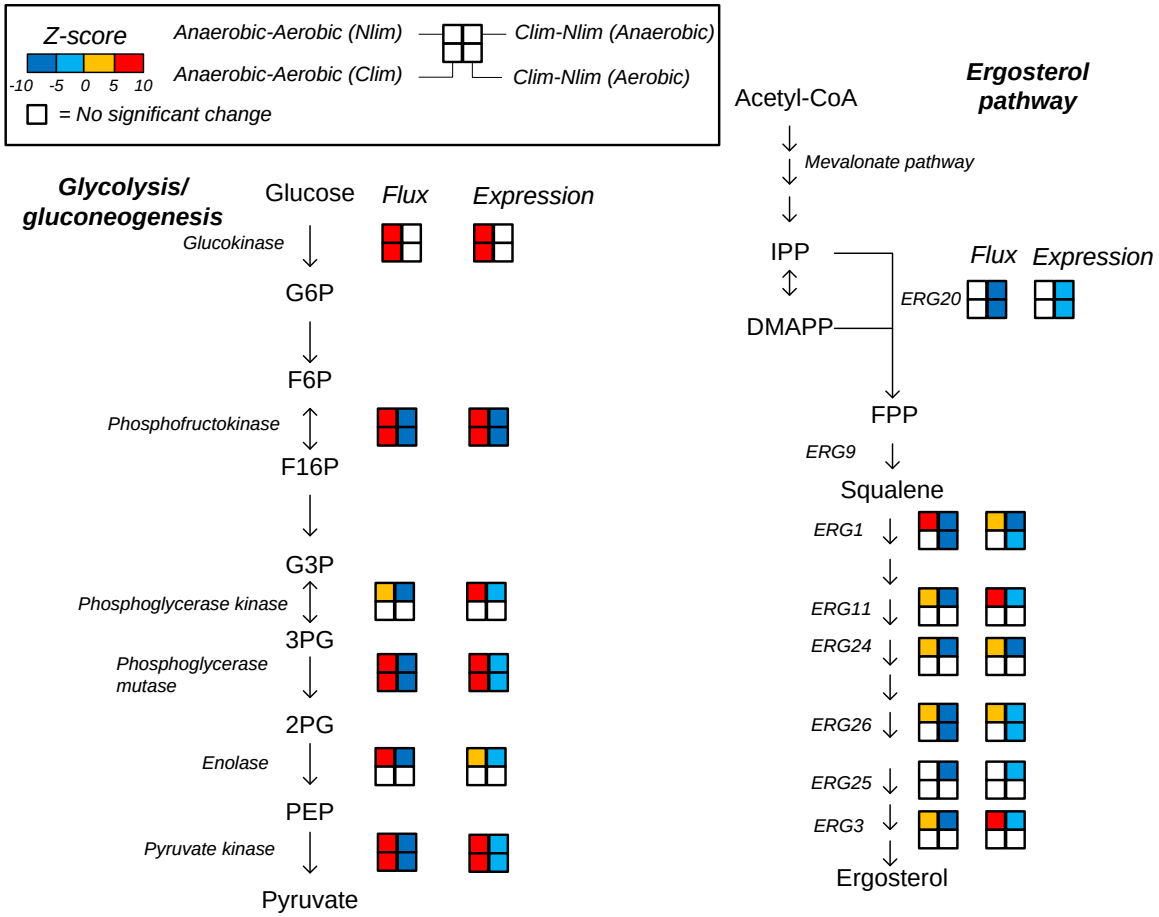


Figure 9 – *Transcriptionally controlled reactions in iTO977 identified by the random sampling algorithm in four different comparisons. Red color means that the reaction is up-regulated in both flux and transcription, blue color means that the reaction is down-regulated in both flux and transcription.*

The transcription factor Opi1p is a negative regulator of phospholipid metabolism and it was overrepresented in all four comparisons, which indicates that phospholipid metabolism is significantly changed both in flux and in transcription of the involved enzymes in all four comparisons. In Figure 9 there are three reactions that are reported as transcriptionally controlled in all four comparisons, namely Phosphofructokinase (*PFK1*), Phosphoglycerate mutase (*GPM1*) and Pyruvate kinase (*CDC19*). These reactions are all glycolytic reactions and the glycolysis is shown in detail in Figure 10.



.

Figure 10 *- Transcriptionally controlled reactions in the glycolysis/gluconeogenesis pathway (left) and in the ergosterol pathway (right).*

Since the data comes from chemostat experiments the growth rate of the cells is controlled to be the dilution rate 0.05 h^{-1}. In batch experiments the aerobic cells would probably grow faster than the anaerobic cells, since aerobic growth (respiration) is more energetically efficient. In a chemostat when the growth rate is constrained, the anaerobic cells instead have to take up more glucose in order to be able to grow with the same speed as the aerobic cells and the glycolytic activity is higher. Figure 10 shows that at least four steps in the glycolysis seems to be transcriptionally controlled and these reactions are upregulated both in flux and transcription of the enzymes when comparing anaerobic to aerobic conditions which is explained by a higher glycolytic activity in the anaerobic cells compared to the aerobic.

Similarly, the cells take up more glucose in N-limited conditions (glucose excess conditions) than in C-limited (glucose limited) conditions since there is more glucose available, and the glycolytic activity is higher. Several of the reactions in the Ergosterol pathway (the conversion of Acetyl-CoA to Ergosterol) seem also to be transcriptionally regulated, which is shown in the right part of Figure 10.

3.3. Controllability of yeast Transcriptional Regulatory Network

Transcriptional regulation is condition dependent where transcription factors (TFs) can respond to environmental cues and regulate the transcription of their target genes. The transcriptional regulatory network can be interconnected, meaning that TFs can form complexes with other TFs or regulate the transcription of other TFs. In **Paper III** we investigate the topology and organization of the yeast transcriptional regulatory network (TRN) by investigating the controllability of the network.

3.3.1 Network Controllability

The concept of network controllability originates from control theory and was introduced for real complex networks by Liu et al (2011). A dynamic system can be formulated as

$$\frac{d\boldsymbol{x}(t)}{dt} = A\boldsymbol{x}(t) + B\boldsymbol{u}(t) \qquad (10)$$

where the vector ***x(t)*** represents the *n* internal variables of the system and ***u(t)*** are the *m* inputs to the system. *A* is a *n* x *n* matrix and *B* is a *n* x *m* matrix with coefficients. The controllability of the system is defined by the controllability matrix $C = (B, AB, A^2B,\ldots,A^{n-1}B)$. The whole system can be controlled given the input vector ***u(t)***if and only if the controllability matrix has full rank, that is rank(*C*) = *n*.

How large part of the network that can be controlled given the input vector ***u(t)*** can be determined from the "maximum matching" graph, which is the longest non-overlapping path in the graph. In the example in Figure 11 we only need to control one input node in the left graph to control the whole system since node A is controlling node B and node B is controlling node C. In the right graph we need to control two nodes to be able to control the whole system, i.e. the nodes A, B and C should be able to reach any state based on the input to the system.

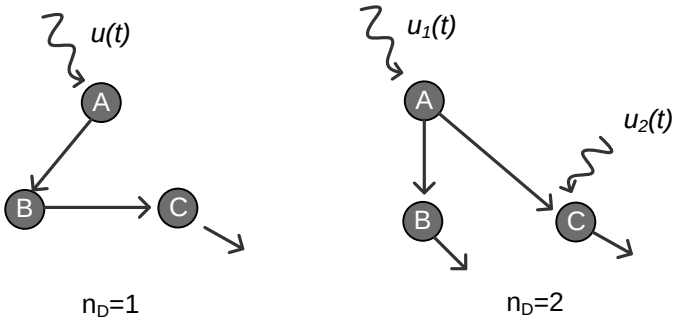


Figure 11 – *Network controllability for two small systems. In the left graph we can control the whole system by controlling only one input node (driver node). In the right graph we need two driver nodes to control the whole system.*

The concept of controllability is taken from electrical engineering and automatic control theory and adapted to biological systems. Even if a biological system cannot be controlled in the same exact way as an electrical system the controllability analysis can be a useful tool to reveal topological structures and motifs in the control of the network.

3.3.2 Controllability analysis reveals circular control motifs

To study the topology and organization of the yeast transcriptional regulatory network (TRN) we applied controllability analysis to investigate how large part of the network that can be controlled when one transcription factor (TF) is controlled as input. If we also can identify TFs that respond to the environment we can investigate how large part of the network that can be controlled if we control the TFs responding to an environmental cue and in that way create a condition-specific transcriptional regulatory network.

The TRNs used in this study were constructed using different ChIP-chip and ChIP-seq (Barski et al, 2007; Robertson et al, 2007) datasets as a starting point. For yeast TF-DNA interactions were defined either by the Harbison et al. dataset (Harbison et al, 2004) with two different probabilities for TF binding ($p<0.001$ and $p<0.005$) and from the Yeastract database (Teixeira et al, 2006) where both direct evidence of TF binding from ChIP-chip experiments and indirect evidence (i.e. change in transcription of the target gene in a microarray experiment where the TF is deleted) were included. The TRNs were constructed to contain only TF-TF interactions, i.e. non-TF genes were filtered out.

The controllability analysis for the yeast TRN revealed information about the topology and structure of the regulatory network. For the Yeastract network it is possible to control 78 % of the network just by controlling one TF as input. The large controllability in this network is due to a big internal loop where the TFs control each other in a circular manner. This internal loop is referred to as a circular control motif (CCM). For the two Harbison networks analyzed we also find CCMs but they are smaller. For the $p<0.005$ network we can control maximum 36 % of the network by controlling one input node and for the $p<0.001$ the maximum controllability is 19 %. To see if this internal loop structure is specific for yeast we also performed the analysis for some other TRNs for other organisms. The results of this analysis are presented in Table 4. For the *E.coli* network (Gama-Castro et al, 2011) we don't find any CCMs and the maximum controllability when controlling one node is 6 %. For the human and mouse networks (Lachmann et al, 2010; Zambelli et al, 2012) we find internal loops and a large part of the network can be controlled just by one input to the system.

Table 4 – Networks and maximum controllability when controlling one input node

Network	Number of TFs	TF-TF interactions	Average degree	Maximum controllability
S.cerevisiae Yeastract	182	1824	18.28	78 %
S.cerevisiae Harbison (p<0.001)	102	231	7.26	19 %
S.cerevisiae Harbison (p<0.005)	102	391	8.67	36 %
E.coli Regulon DB	139	196	3.36	6 %
M.musculus Chea v.1	51	547	18.43	98 %
M.musculus Chea v.2	133	3125	34.99	92 %
H.sapiens Chea v.1	33	103	6.06	64 %
H.sapiens Chea v.2	97	1322	23.08	87 %
H.sapiens Cscan	126	3930	26.46	46 %

Figure 12 shows the network structure of two *S. cerevisiae* regulatory networks (A and B) and one *E. coli* regulatory network (C). The green nodes show the TFs that belong to the internal loop (CCM).

Studying the three different yeast TRNs it seems like the maximum controllability of a network might be a function of the amount of information included in the network, or the average number of neighbors in the graph (average degree). The Harbison 0.001 network is the most conservative in terms of what to consider as a TF binding event, and it has also the lowest controllability of the three yeast networks (19 %). The Yeastract network is the least conservative in terms of what to consider as TF binding, it has the highest average degree and it also has the highest controllability (78 %). In order to test if the controllability is a function of the average degree of the network we simulated random networks (Erdős & Renyi, 1961) and scale-free networks (Barabási & Albert, 1999) with different average degrees. The result of this analysis is plotted in Figure 13. For the simulated scale-free networks the maximum controllability of the network increases slightly with increased average degree, but it never exceeds 29 % for the networks simulated here. In contrast, the simulated random networks have a higher controllability and when the average degree is 8 or higher we can almost control 100 % of the network by only controlling one input node. This suggests that the high controllability for real networks might also be a consequence of more randomness in the organization of the networks. Based on the controllability analysis it seems like the E.coli network behaves more as a perfect scale-free network, the yeast networks behaves somehow in between a scale-free and random network and the behavior of the yeast and human networks are more random.

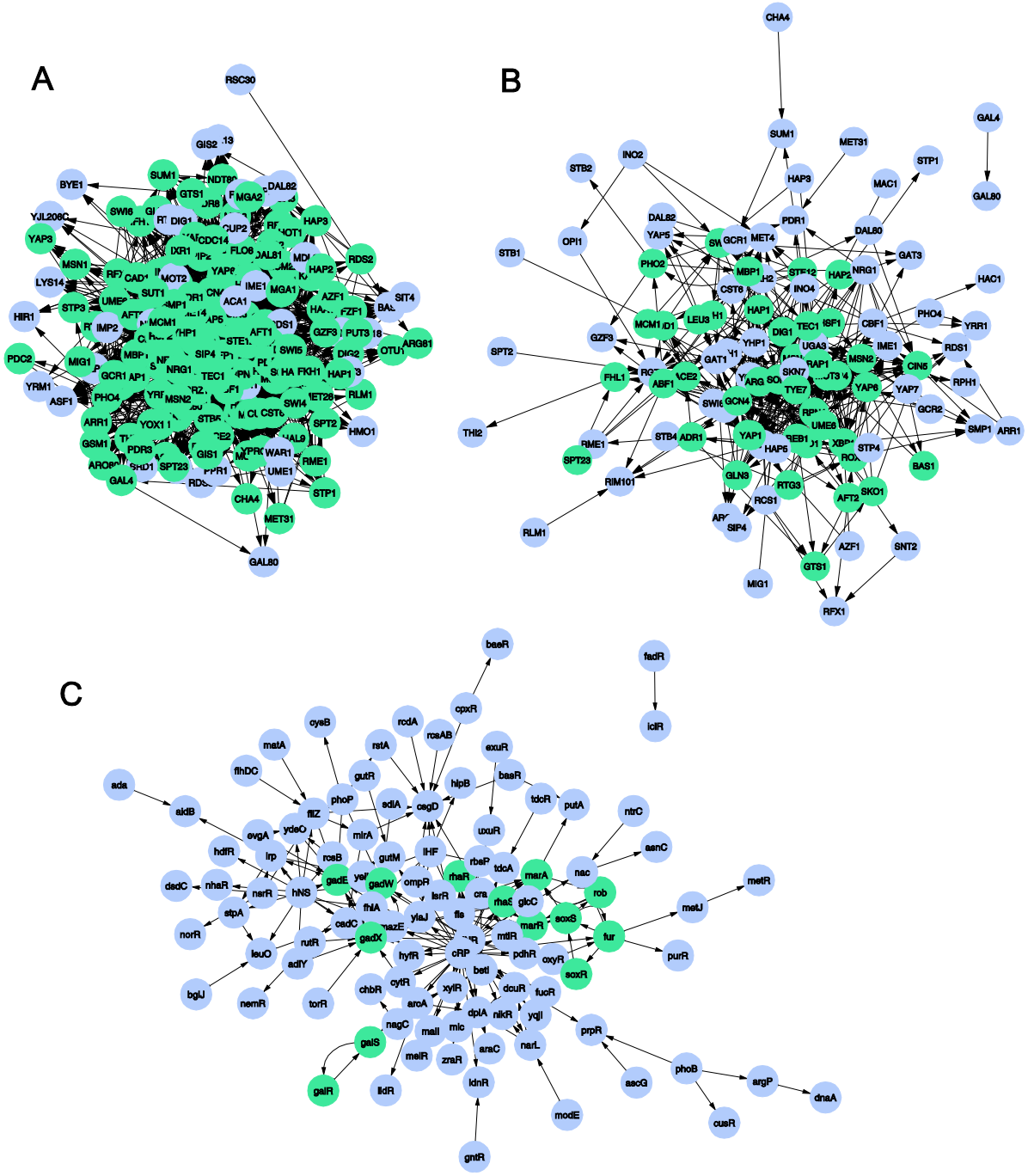


Figure 12 *– Controllability of transcriptional regulatory networks (TRNs) when controlling only one input node (one driver node). The green nodes represents the circular control motif (CCM), an internal loop in the network where all nodes in the CCM can be controlled just by controlling one input node. (A) S.cerevisiae yeastract network (B) S.cerevisiae Harbison network with p<0.001 (C) E.coli Regulon DB network.*

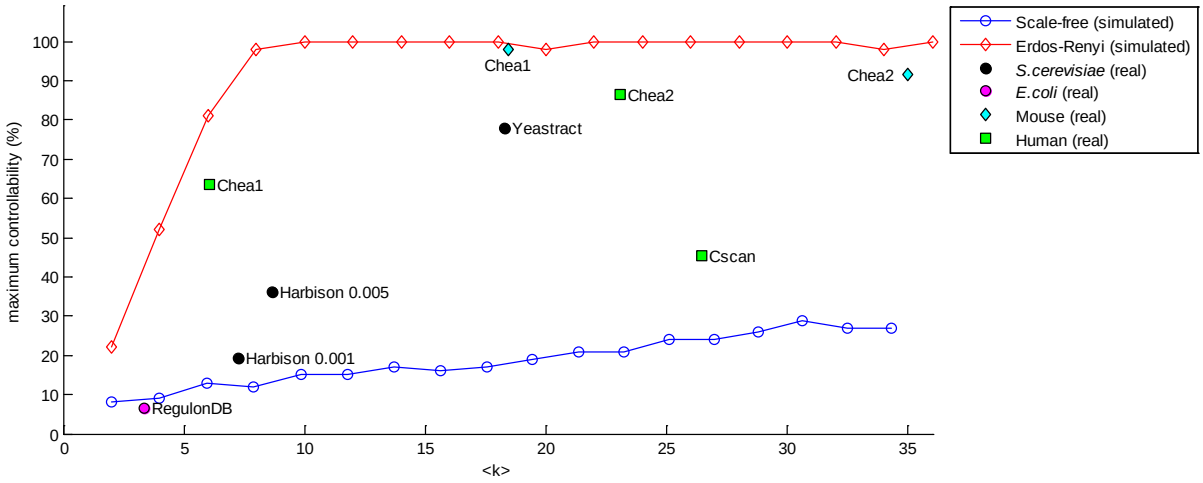


Figure 13 – *Maximum controllability for real and simulated networks as a function of the average degree of the network <k>.*

To investigate if the networks are scale-free or not it is possible to plot the degree distribution on a log-log scale and fit a power law distribution to the points. If the network is scale-free the degree distribution should follow a power law distribution, i.e. $P(k) \sim k^{-\alpha}$ where k is the degree of a node and α is a parameter that normally have a value between 2 and 3 for a scale free network (Barabási & Albert, 1999). Figure 14 shows that the *S. cerevisiae* Harbison network and the *E. coli* network show most scale-free behavior (i.e. the fit of the power law distribution is reasonable). The Yeastract network is not scale free since the number of nodes with low degree, e.g. 1, 2 and 3 is less than expected for a scale free network. The graphs in Figure 14 show that the networks for mouse and human are not scale-free since the fit of the power law distribution is poor.

For the *S. cerevisiae* networks we have three networks with different average degree (and different confidence in the included TF-DNA interactions). For these networks the controllability increases with increasing degree, suggesting that if we increase the average degree for the yeast networks the network behaves more random.

Study transcriptional regulation in human and mouse might be problematic due to increased complexity and also incomplete transcriptional regulatory networks. For example there are around 1400 DNA binding transcription factors in the human genome (Vaquerizas et al, 2009), but the TF regulatory networks constructed from ChIP-seq involves maximum 97 transcription factors.

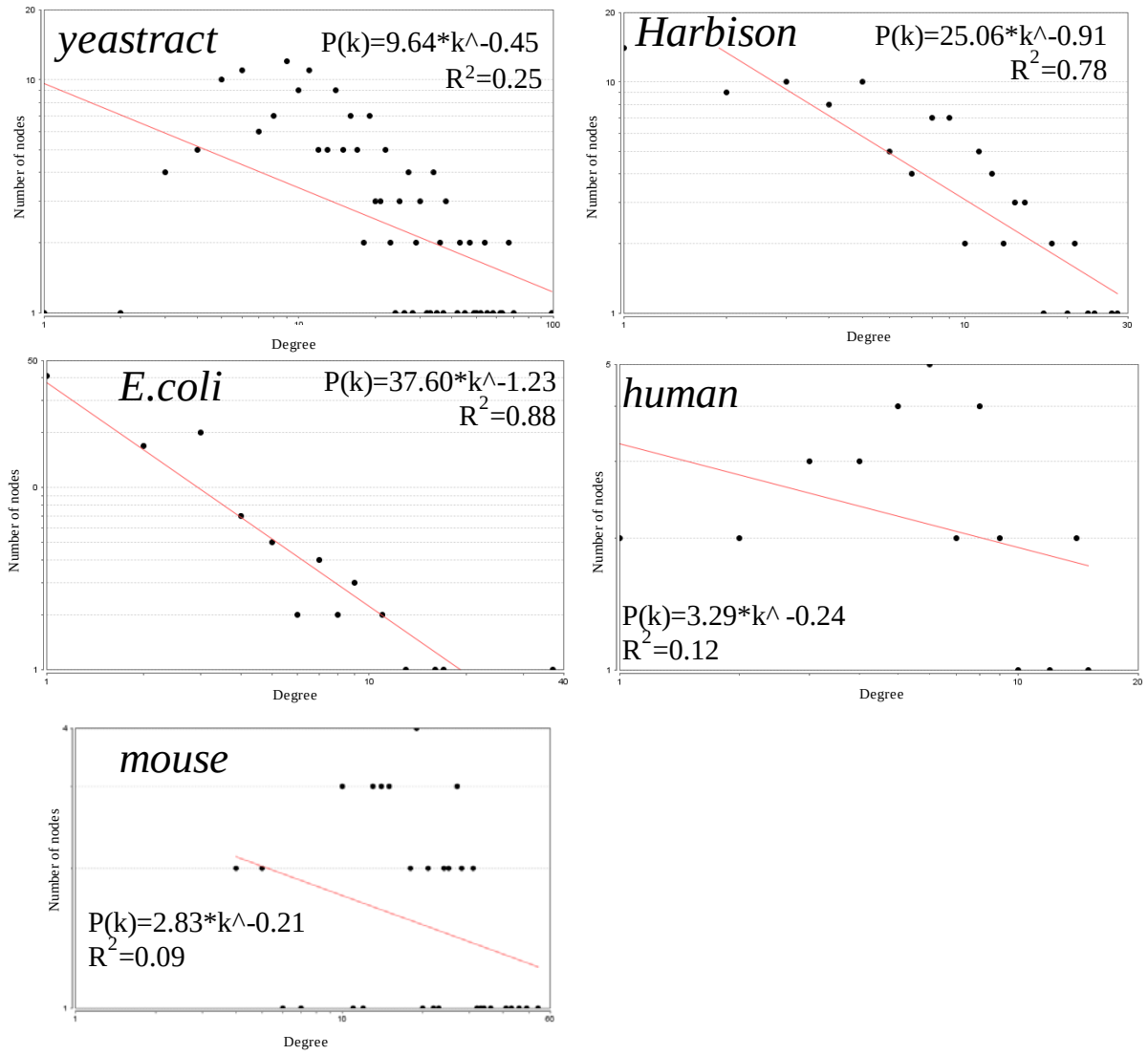


Figure 14 – *Degree distribution plots on log-log scale for different networks. The red line and the formula represent the fitted power law distribution and the R^2-value represents how good the fit is.*

3.3.3 Yeast transcription factors responding to environment

Yeast transcriptional regulation is condition specific. In order to investigate the condition-specific transcriptional response in yeast we investigated gene expression profiles from 233 chemostat experiments where the environment has been controlled. The gene expression for each gene was modeled with an ANOVA model including the environmental factors described in Table 5.

Table 5 – Environmental factors

Factor	Levels
Oxygen availability	Aerobic, Anaerobic
Nutrient limitation	Carbon, Nitrogen, Zinc, Iron, Phosphorus, Sulfur
Dilution rate	0.02, 0.03, 0.05, 0.1, 0.2, 0.25
Carbon source	Glucose, Ethanol, Acetate, Maltose, Galactose
Extra compound	None, Acetate, Benzoate, Propionate, Sorbate, Formate, CO2, Ethanol

In order to find transcription factors that are affected by a specific environmental cue we did the following:

1) Identify TFs whose target genes in the TRN change in expression between different conditions
2) Identify TFs who does not change in expression between different conditions
3) Find TFs in the iMH805/775 Boolean rules collected from primary literature that are reported to be affected by extracellular cues (Herrgård et al, 2006).

In order to test if the target genes of a transcription factor are changed in expression as a function of a specific environmental cue a hyper-geometric test was performed for each transcription factor and each environmental factor. The idea is to test if the number significantly changed genes among the genes regulated by that transcription factor (the target genes) is over-represented among all significantly changed genes. The comparison of the hypergeometric test p-value and the adjusted ANOVA p-value for the transcription factor itself is presented for the different environmental factors in Figure 15.

The results from the hyper-geometric test for the factors oxygen availability, nutrient limitation and dilution rate are presented in Figure 16.

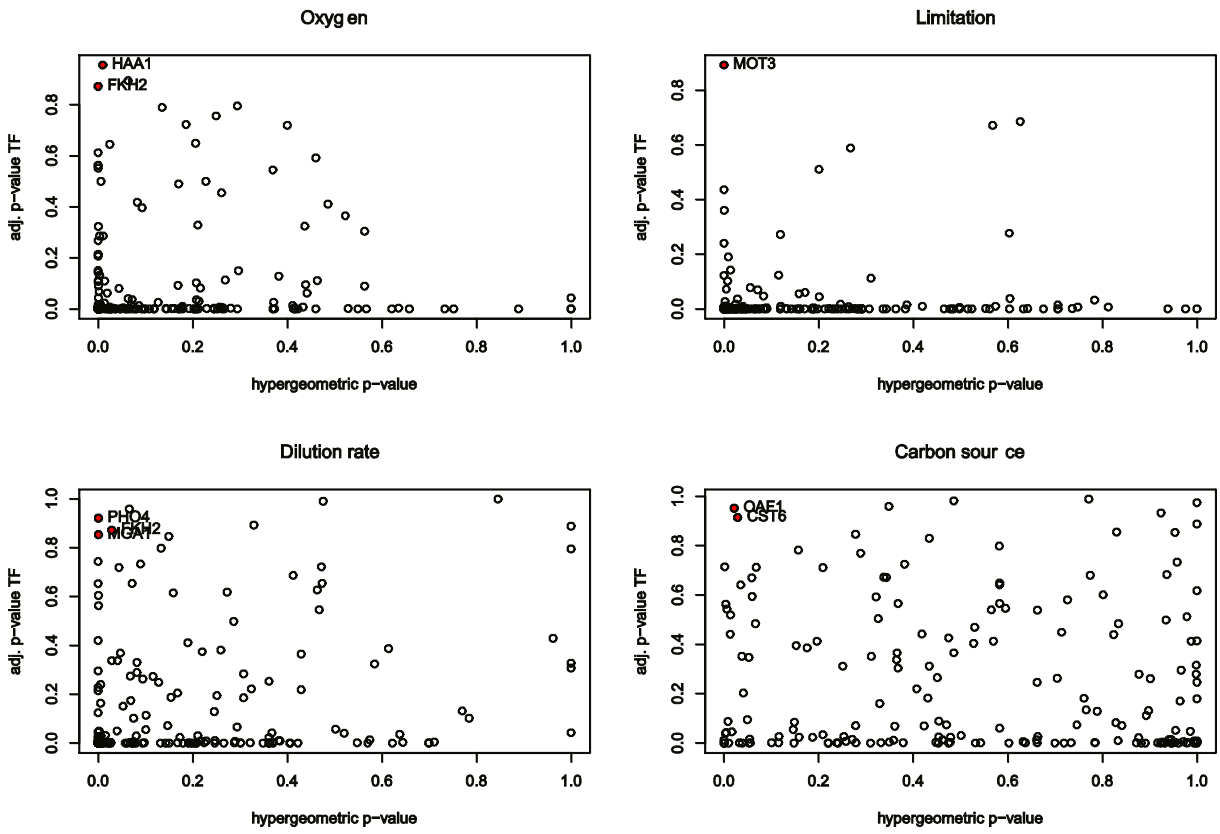


Figure 15 – *Transcription factors affected by environmental cues. The x-axis shows the p-value from the hyper-geometric test for the target genes and the y-axis shows the adjusted p-value for the transcription factor gene itself. The TFs marked with filled circles have a hyper-geometric p-value for the target genes < 0.05 and an adj. p-value for the TF gene > 0.8.*

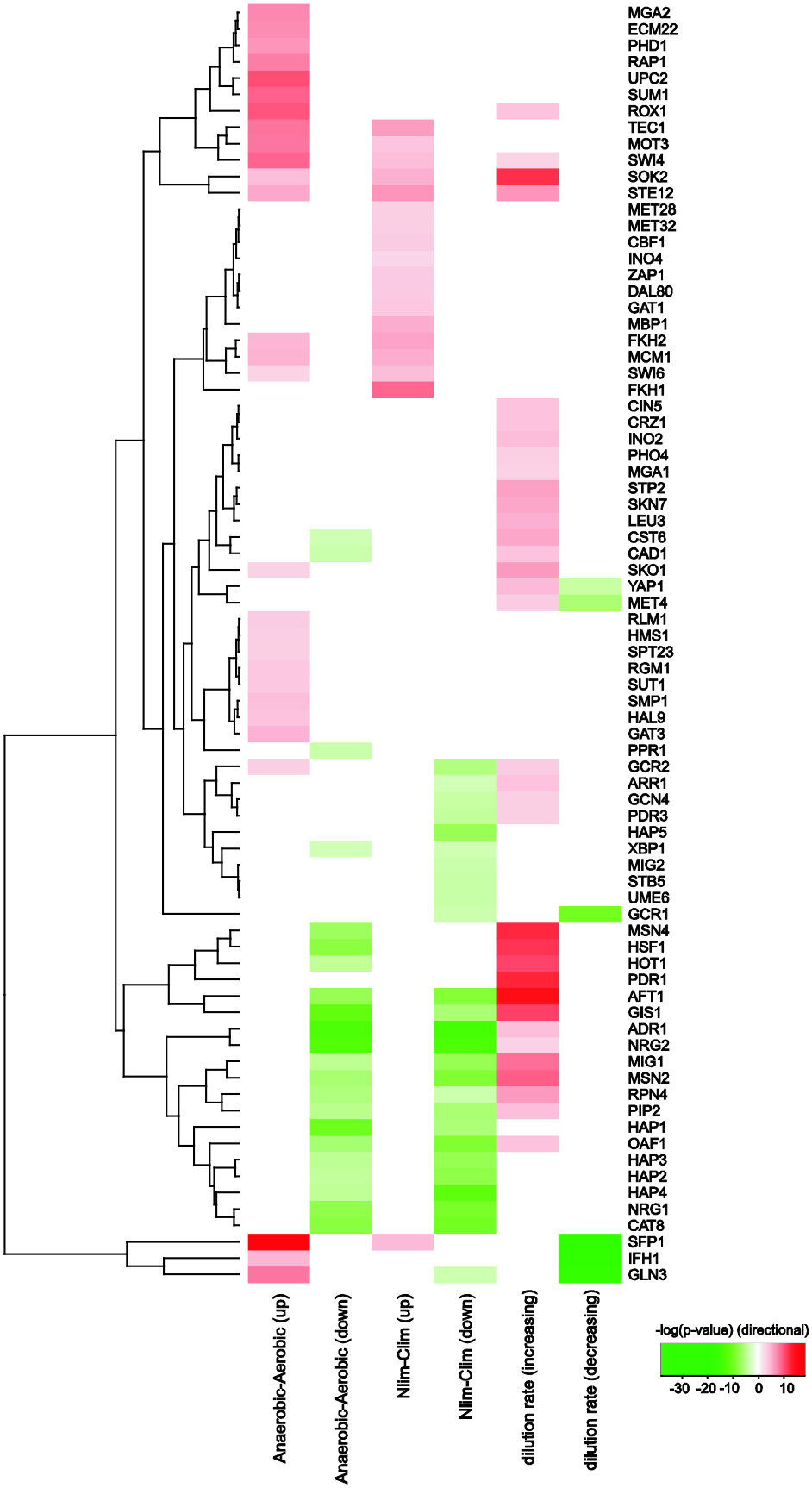


Figure 16 – *Hypergeometric test – red color corresponds to up-regulation of the target genes and green color corresponds to down-regulation of the target genes.*

4. Systems biology for protein production

Proteins are part of the building blocks of the cell and they carry important cellular functions, e.g. enzymatic activity, signaling, immune response etc. Many pharmaceutical proteins and industrial enzymes, including insulin and amylase, are produced by engineered production host cells (Demain & Vaishnav, 2009; Walsh, 2010). When choosing the host organism for recombinant protein production it is important to consider properties of the secretory pathway in the production organism. Bacteria can be used for production of small simple proteins, but for production of proteins that requires complex post-translational modifications and proteins with many disulfide bonds a Eukaryotic production system is a better choice (Graf et al, 2009). Figure 17 shows an overview of the secretion pathway in yeast. If a newly translated protein has a signal peptide (SP) it will be recognized by the signal recognition particle (SRP) and translocated into the Endoplasmic reticulum. The protein will be folded and pass many steps of post-translational modification and sorting on the way to its final localization. Understanding the protein secretion pathway and its regulation can help engineering of recombinant protein production in yeast for improved protein production. (Hou et al, 2012)

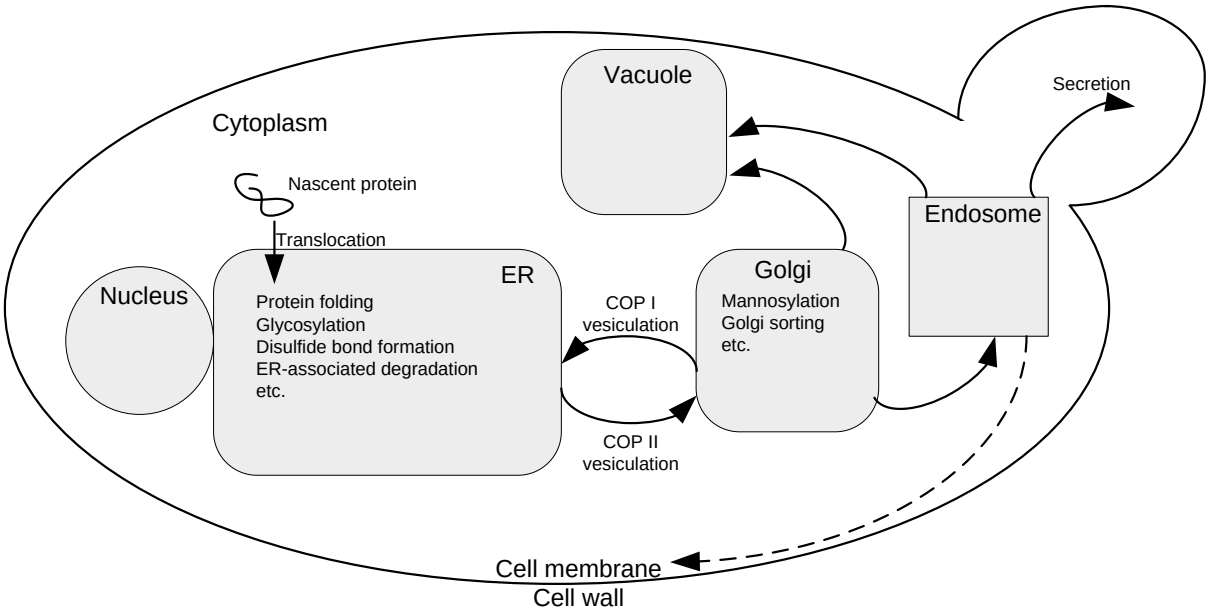


Figure 17 – *A schematic picture of the protein secretory machinery in yeast. Unfolded proteins enter the Endoplasmic reticulum (ER) where it is folded, glycosylated etc. Misfolded proteins can be degraded using for example the ER-associated degradation (ERAD) pathway. Otherwise it is vesiculated and transported to Golgi for further modifications and sorting. Depending on the localization of the protein it might be transported to other compartments, or be secreted.*

In this chapter application of systems biology to study recombinant protein production in yeast is illustrated by several examples. First, it is shown that yeast can produce higher levels of recombinant α-amylase under anaerobic conditions than under aerobic cultivations, and the response of metabolism to protein secretion under anaerobic and aerobic conditions is investigated using integrated analysis (**Paper IV**). Further the framework for genome-scale modeling is used to reconstruct a detailed model of the protein secretion machinery in yeast, including classification of the different components of the secretory pathway into different subsystems (**Paper V**). Last but not least a model of the secretory machinery in *Aspergillus oryzae* is built based on the secretion model for yeast, and data from fermentations and transcriptome experiments is used for integrative analysis (**Paper VI**).

4.1. Anaerobic α-amylase production in yeast

In **paper IV** we investigate recombinant α-amylase production in *Saccharomyces cerevisiae* under aerobic and anaerobic conditions (Liu et al, 2013). Most of the studies investigating recombinant protein production have been performed under aerobic conditions. The protein folding occurs in the endoplasmic reticulum and might require the formation of a disulfide (S-S) bridge.

$$\text{R-SH} + \text{HS-R} \xrightarrow{oxidation} \text{R-S-S-R} + 2\,\text{H}^+ + 2\,e^- \qquad (11)$$

This process (Equation 11) is an enzymatic redox reaction and requires electron transfer to an electron acceptor (Sevier & Kaiser, 2002). In *S. cerevisiae* it is known that electrons are transferred to oxygen under aerobic conditions forming reactive oxygen species (ROS) (Nguyen et al, 2011; Tu et al, 2000). For amylase production under aerobic conditions it has been reported that the oxygen uptake rate and ATP consumption rate is 2-fold higher than in the wild type, suggesting that this is due to increased oxidation in connection with electron transfer from the ER (Tyo et al, 2012). However, under anaerobic conditions the final electron acceptor is unknown. This study focuses on studying the global response to protein production under anaerobic conditions compared to aerobic in order to understand the mechanisms for anaerobic protein folding and secretion.

The two yeast strains used in this study are the AAC strain (α-amylase producing strain) and the NC strain (negative control strain). The AAC strain was constructed by transforming the starting strain CEN.PK530-1C with a plasmid containing the α-amylase gene under control of the *TPI1* promoter and *TPI1* terminator. The plasmid also contains a copy of the *POT1* gene from *Schizosaccharomyces pombe*. The starting strain has a *TPI1* deletion which makes it unable to grow on glucose and grows slowly on other carbon sources. The *S. pombe POT1* gene encodes for the same function as *TPI1* which requires that the cell expresses the vector in order to grow (Liu et al, 2012). The NC strain was constructed by transforming the starting strain with an empty plasmid.

The two strains AAC and NC were grown under aerobic and anaerobic conditions and the levels of α-amylase were measured. Figure 18 shows the protein yield of α-amylase production under aerobic and anaerobic conditions. The α-amylase production

in the AAC strain was around 3-fold higher under anaerobic conditions than aerobic. We therefore looked into anaerobic protein production compared to aerobic, trying to find mechanisms that can explain why anaerobic cultivations produced more α-amylase than aerobic cultivations. In order to do this we performed microarray experiments to measure gene expression in the two strains under the two different conditions.

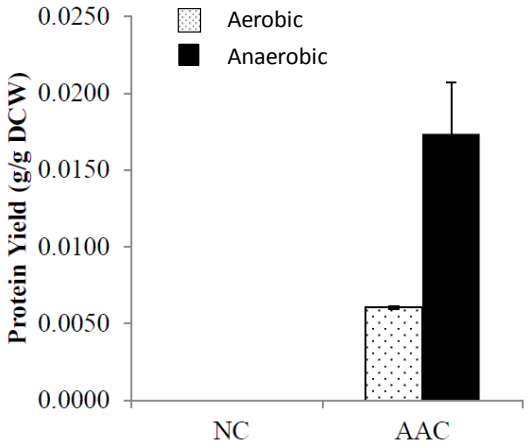


Figure 18- *The yield of α-amylase production under aerobic and anaerobic cultivations.*

When comparing anaerobic to aerobic conditions 2427 genes were significantly changed in transcription (adjusted p-value < 0.05) for the AAC strain and 2638 genes were significantly changed for the NC strain. To specifically look into the metabolism and to reduce the dimensionality of the data we used the reporter metabolite algorithm (Patil & Nielsen, 2005).

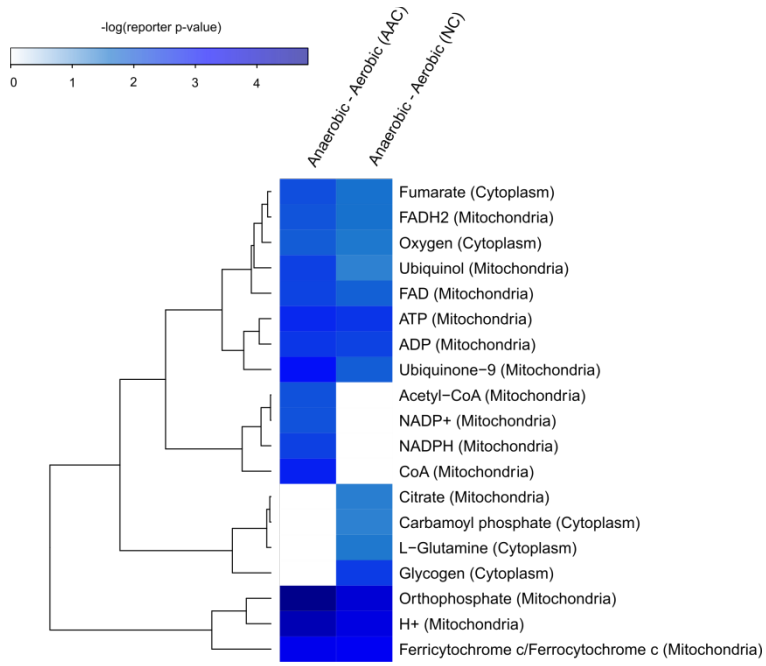


Figure 19 *– Significant reporter metabolites when comparing anaerobic to aerobic conditions for the AAC strain and NC strain respectively.*

The algorithm uses the metabolic network structure from the metabolic model described in chapter 1.1 to identify key metabolites in the network, i.e. metabolites where the neighboring genes change transcriptionally between the two conditions compared. Figure 19 shows the results of the reporter metabolite analysis.

The metabolites fumarate and oxygen in the cytoplasm and ubiquinol and FAD/FADH2 in the mitochondria all appear as significant anaerobic-aerobic reporter metabolites for both strains. However all these are more significant (have a lower reporter p-value) in the amylase producing AAC strain than in the NC strain which means that the genes involved in the metabolism around these metabolites are more significantly changed (up- or down-regulated) in the AAC strain than in the NC strain when comparing anaerobic to aerobic conditions. Looking specifically at the anaerobic- aerobic response in the AAC strain we find a number of significantly regulated genes connected to these metabolites (Table 6).

Table 6 – Significantly changed neighboring genes (adj. p-value < 0.05) for selected reporter metabolites when comparing anaerobic – aerobic cultivations in the AAC strain

Reporter metabolite	Gene	Description	Direction	Log fold change	Adjusted p-value
Fumarate	*FRD1*	Fumarate reductase	↑	1.86842	1.32E-12
	OSM1	Fumarate reductase	↑	0.29533	8.54E-03
	FUM1	Fumarase	↓	-1.25366	2.33E-09
	SFC1	Mitochondrial succinate-fumarate transporter	↓	-0.33579	2.49E-04
FAD/FADH2	*FAD1*	FAD synthesis	↑	0.20803	5.75E-03
	ERV2	Disulfide bond formation	↑	0.58947	3.00E-07
	SDH3	Succinate dehydrogenase	↓	-1.10214	1.41E-07
	FLX1	FAD transporter	↓	-0.30566	2.13E-04
Ubiquinol/ Ubiquinone	*URA1*	Pyrimidine synthesis	↓	-0.30293	3.82E-03
	COQ1	Ubiquinone synthesis	↓	-0.61828	3.76E-07
	COQ2	Ubiquinone synthesis	↓	-0.28939	5.58E-03
	COQ4	Ubiquinone synthesis	↑	0.18317	8.39E-03
	COQ6	Ubiquinone synthesis	↑	0.46284	1.13E-05
	COQ9	Ubiquinone synthesis	↑	0.57733	3.09E-05

Based on this analysis, together with the macroscopic flux analysis, we proposed a model for the electron transfer from disulfide bond formation in the ER under anaerobic conditions which is presented in Figure 20. The protein Pdi1p is responsible for disulfide bond formation in the ER and can under aerobic conditions transfer two electrons per disulfide bond from thiol substrates to oxygen via Ero1p or Erv2p (Gross et al, 2006). Under anaerobic conditions there is no oxygen available to take care of the two electrons produced in Equation 7. Instead it can be transferred to free FAD in the ER forming $FADH_2$ and the $FADH_2$ can then be transported over the ER membrane.

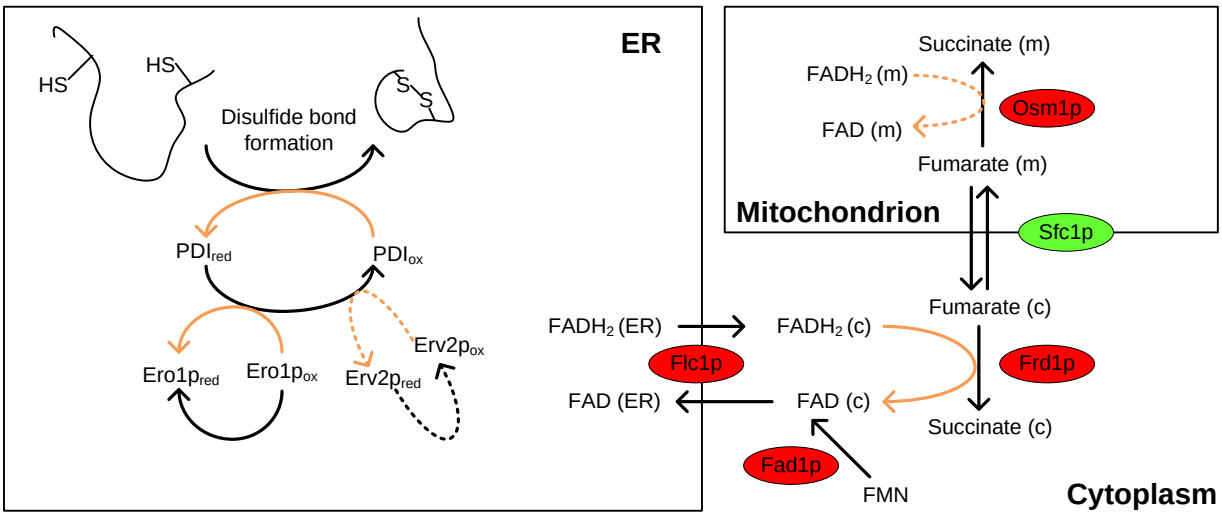


Figure 20 – *Model for electron transfer from disulfide bond formation in the ER with fumarate as the final electron acceptor in the cytoplasm. Orange lines indicate possible electron transfer routes/redox reactions. Dashed lines indicate alternative electron transfer routes. Red ellipses indicate enzymes that are transcriptionally up-regulated in anaerobic conditions compared to aerobic. Green ellipse means down-regulated in anaerobic conditions.*

The cytosolic $FADH_2$ can then be oxidized in the cytosol by the reaction catalyzed by Frd1p converting fumarate to succinate. Fumarate could also be transported into mitochondrion and there be converted to succinate and at the same time oxidize $FADH_2$ to FAD. The hypothesis that fumarate is working as an electron acceptor under anaerobic conditions was tested experimentally by adding fumarate to the growth medium when cultivating the cells in aerobic and anaerobic conditions (Figure 21).

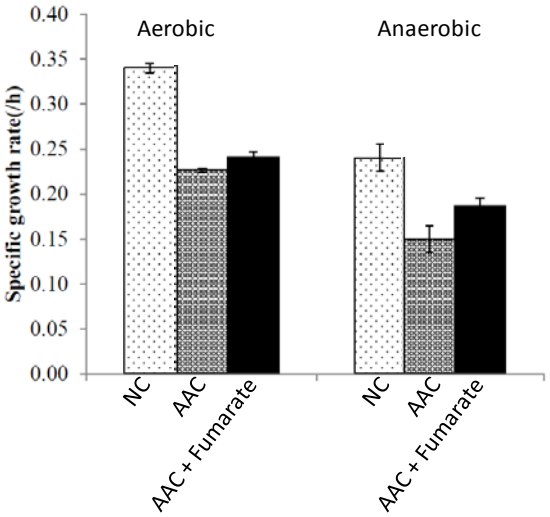


Figure 21 *– Specific growth rate under aerobic and anaerobic conditions for the two strains NC and AAC and for the AAC strain under addition of fumarate.*

The AAC strain (amylase producing strain) grows slower than the NC (control) strain since increased protein production might increase the stress in ER and oxidative stress. The results in Figure 21 show that the growth of the AAC strain was partly restored due to addition of fumarate to the media, especially under anaerobic conditions. This results strengthen the hypothesis that fumarate might act as an important player in connection with anaerobic protein folding and might work as the final electron acceptor for anaerobic disulfide bond formation.

In summary, in **paper IV** we used microarrays as a tool to study gene expression in aerobic and anaerobic conditions. The data was integrated into the metabolic model using the reporter metabolite algorithm. Based on the results from transcriptome and fermentation experiments we suggest that fumarate act as the final electron acceptor from disulfide bond formation under anaerobic conditions.

4.2. A genome-scale model of protein secretion in yeast

The secretory pathway is an important pathway in Eukaryotic cells and is responsible for post-translational modifications (PTMs) and protein sorting etc. (Schekman, 2010). Figure 17 on page 31 shows an overview of the secretory pathway in yeast. In **paper V** we used a systems biology approach to model the protein secretion machinery in *Saccharomyces cerevisiae*. The idea of the reconstruction is to get a systemic view of the protein secretion network as well as capturing the protein specific functions of the protein secretion. Depending on the sequence and properties of the clients to the secretory machinery they may take different routes in the secretory pathway. The protein secretory model has the following features:

1. Annotation of 162 protein components and 1 RNA component in terms of function in the protein secretion machinery. The 163 secretory machinery components are divided into 16 subsystems and 8 different compartments. The processes are represented by pseudo-chemical reactions, so called template reactions.
2. Classification of 1197 client proteins based on their secretory features, including localization and PTMs. The client proteins can belong to one of 185 theoretical secretory classes where each class contains a specific combination of PTMs, sorting and transport steps processed by the 163 components of the secretory machinery.
3. An algorithm for simulating protein secretion in yeast based on the protein-specific information matrix (PSIM) which is constructed from the template reactions and the secretory features.
4. Possibility for future connection of the protein secretion model with other systems in the cell, e.g. the metabolic network due to cofactors and metabolites included in the model.

The model is illustrated in Figure 22 where all the different routes a client protein can take are illustrated by arrows. The 16 different subsystems of the model are presented in Table 7. Different client proteins that enter the secretory machinery through the

translocation pathway (S1) will pass through different combinations of these subsystems.

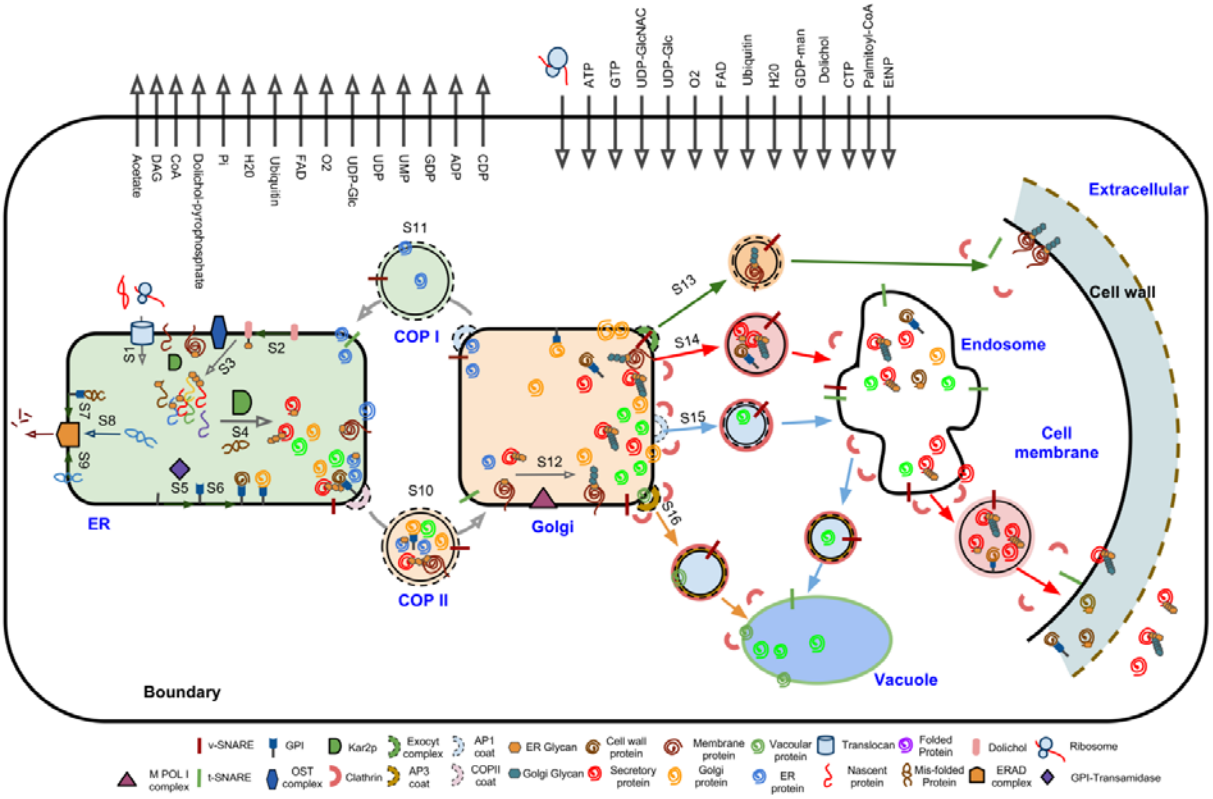


***Figure 22** – Representation of the genome-scale model of protein secretion in yeast. The different subsystems denoted S1-S16 are explained in Table 7.*

Table 7 - Subsystems in the protein secretion model

Subsystem	Name	Number of machinery proteins
S1	Translocation	18
S2	Dolichol pathway	15
S3	ER glycosylation	15
S4	Protein folding	10
S5	GPI biosynthesis	20
S6	GPI transfer	0
S7	ERADC	3
S8	ERADL	14
S9	ERADM	0
S10	COPII	22
S11	COPI	10
S12	Golgi processing	11
S13	Low density secretory vesicle (LDSV)	9
S14	High density secretory vesicle (HDSV)	1
S15	CPY pathway	10
S16	ALP pathway	5

Which subsystems that will process the client protein is decided by 7 different secretory features which are all contained in one way or another in the protein sequence (Table 8). The secretory features for each protein were extracted from three different databases: Uniprot (Bairoch et al, 2005), SGD (Weng et al, 2003) and KEGG (Kanehisa et al, 2006).

Table 8 – Type of protein specific information used to generate the PSIM

Secretory feature	Values
Signal peptide	Yes, No
Number of N-linked glycosylation sites	0,1,2,3…
Number of O-linked glycosylation sites	0,1,2,3…
Number of disulfide bonds	0,1,2,3…
Localization	ER, Cell membrane, Vacuole, Golgi, Extracellular
GPI anchoring*	Yes, No
Number of transmembrane domains*	0,1,2,3…

* For membrane proteins only

After extracting the secretory features for each protein we could construct the protein specific information matrix (PSIM) which has dimensions 5882x7 where each row represents a yeast protein and each column represents one secretory feature reported in Table 8. The PSIM can be used to group the yeast proteins into secretory classes where each class has a specific combination of the secretory features.

4.3. Modeling α-amylase production in *Aspergillus oryzae*

Aspergillus oryzae, or Koji mold, is a species of filamentous fungus that is used for food and Asian beverage production from soy beans and rice (Knuf & Nielsen, 2012). The high secretion capabilities of this organism makes it suitable as a host for protein production and it has been used to express heterologous proteins (Lubertozzi & Keasling, 2009; Ward, 2012). In **paper VI** we investigate *α*-amylase production in *A. oryzae* by combining large-scale gene expression analysis with modeling for three different amylase producing *A. oryzae* strains grown in batch cultivations. The three strains CF1.1, CF32 and A16 were constructed in different ways by transforming different plasmids containing the TAKA-amylase gene and different promoters. The wild type A1560 strain can also produce amylase, but by engineering the strain we could make it produce higher amounts.

Figure 23 shows the physiological parameters for the α-amylase production under batch cultivations. The CF32 strain has the highest amylase production but slow growth rate, and hence a relatively low productivity. The A16 strain has a high amylase yield and relatively high growth rate, and therefore the highest productivity. If we can investigate the global response to amylase production and understand why the CF32 strain grows slower than the other strains we might be able to engineer the CF32 strain in order to achieve an even higher productivity.

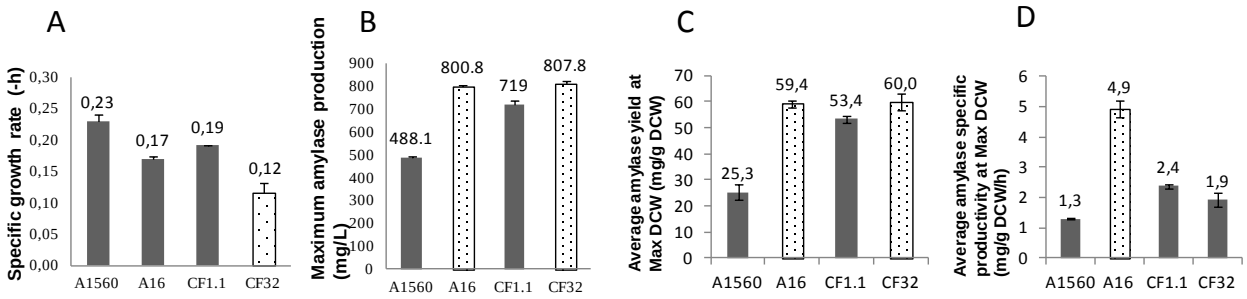


Figure 23 – *Growth and amylase production for the three amylase producing strains and the wild type A1560 strain in batch cultivations. A) Specific growth rate (h^{-1}) B) Maximum amylase production (mg/L) C) Yield of α-amylase at the point of maximum biomass (mg/g DCW) D) α-amylase productivity at the point in C (mg/gDCW/h).*

We studied the transcriptional response to *α*-amylase production using microarrays. 1709 genes changed in transcription (adj. p-value <0.05) when comparing the CF32 to the wild type A1560 strain. For the A16-A1560 comparison 1222 genes were identified as significantly changed and 655 genes for the CF1.1-A1560 comparison.

We performed reporter GO-term analysis (Oliveira et al, 2008) in order to reduce the dimensions of the data and obtained an overview of which pathways or processes that are changed between the amylase producing strains and the wild type strain. Figure 24 shows the results of the reporter GO-term analysis. Gene ontology terms (www.geneontology.org) are bioinformatic classifications of genes into processes and functions (Ashburner et al, 2000). 7699 of the genes in Aspergillus Genome Database (AspGD) has been classified into one or more GO-terms (Arnaud et al, 2012). The GO-terms with up-regulated genes (the red cluster in Figure 24) are mainly GO-terms related to protein secretion, indicating that up-regulation of components involved in the protein secretion machinery (PTMs, protein folding, trafficking etc.) plays an important role for improved production of α-amylase. This made us interested in investigating the protein secretion machinery of *A. oryzae* more in detail.

The protein secretion machinery model for *A. oryzae* was constructed based on the yeast secretion model described in the previous section and in **Paper V.** First we indentified homologs between the *A. oryzae* proteome and the 163 components of the yeast secretory machinery. *A. oryzae* has almost twice as many protein coding genes as *S. cerevisiae*, and based on this we would also expect the number of proteins involved in the protein secretion machinery to be larger. However, the *A. oryzae* genome is less studied than the *S. cerevisiae* genome and many of the genes in *A. oryzae* have unknown function. Also, if we do pairwise bidirectional blast comparisons between the two genomes we will only find maximum one homolog for each protein in the yeast secretory machinery. If we instead assume that *A. oryzae* should have a higher number of inparalogs, i.e. genes with similar functions due to gene duplications, we can use the OrthoMCL database (Chen et al, 2006) which consists of gene families and also includes the inparalogs. The list of A.oryzae components was further extended using PSI-blast best hits and genes with secretory function reported in (Wang et al, 2010).

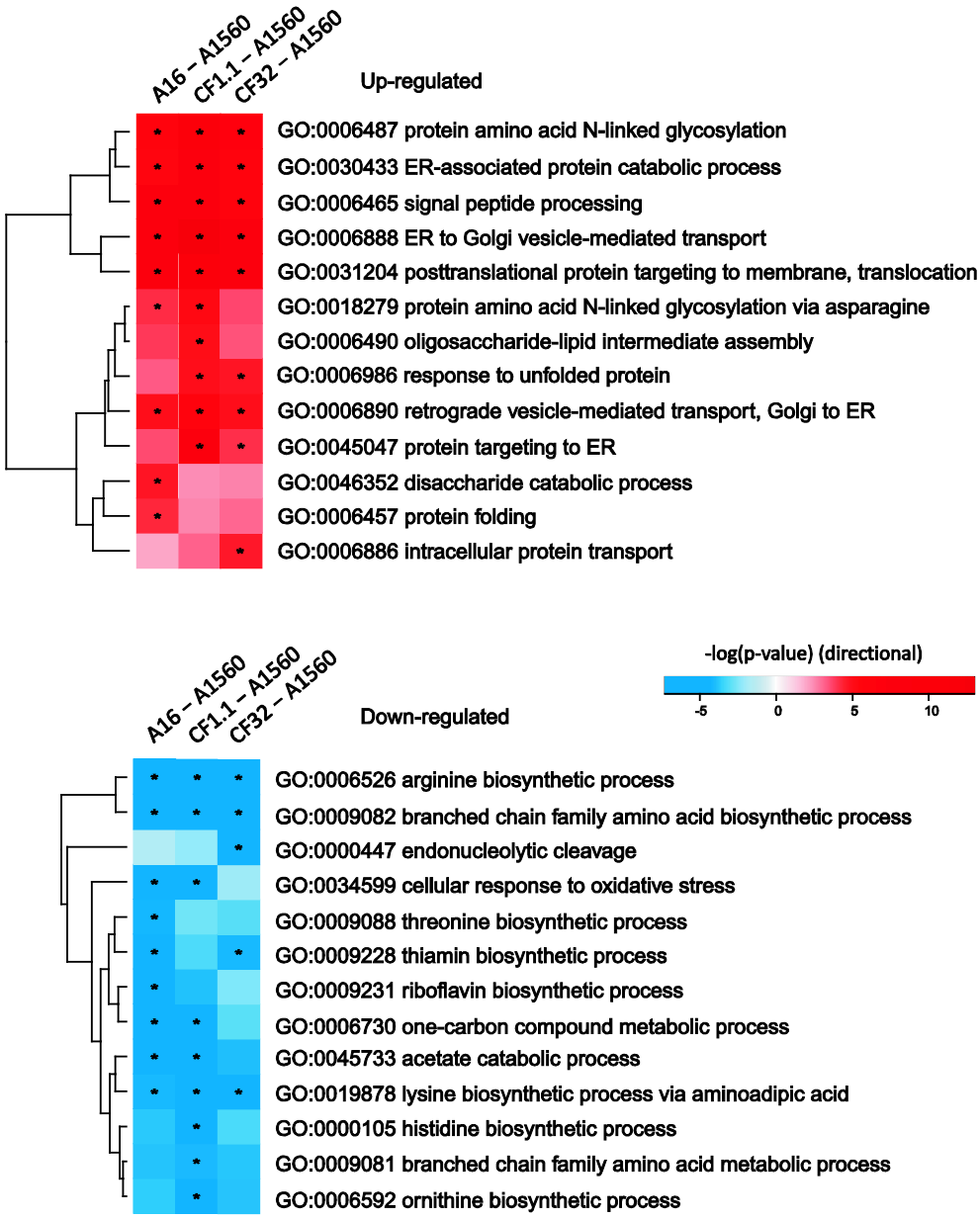


Figure 24 – *Reporter GO-terms for the three amylase producing strains compared to the wild type A1560 strain. Red color indicates that the genes belonging to the GO-term are up-regulated and blue color indicates that the GO-term genes are down-regulated. An asterisk means that the reporter p-value is less than 0.0001.*

To define the client proteins of the A. oryzae secretory machinery we used the signalP 4.0 algorithm (Petersen et al, 2011) to predict the proteins with a signal peptide. The algorithm predicts 1107 proteins with a signal peptide and most of the cleavage sites for the signal peptides are predicted to be located 17-25 amino acids from the N-terminal. Based on the information retrieved we could simulate the secretory machinery for *A. oryzae.*

5. Summary and perspectives

Biological network reconstructions have been presented for *S.cerevisiae* and *A. oryzae* in this thesis, including the metabolic network, transcriptional regulatory network and protein secretion network. In order to understand the biological systems from a holistic view we need comprehensive descriptions of the biological systems. The biological network reconstructions can be used to model the cell and to integrate high-throughput data using the network structure in order to analyze the data in a context-specific manner.

The genome-scale metabolic model presented here (iTO977) is larger in scope and more comprehensive than previous models and allows both simulation of the whole system and serves as a knowledge base for annotation of reactions and pathways, which can be used for integrative analysis.

One disadvantage with the FBA modeling framework is that information about regulation is missing in the model formulation. False predictions using genome-scale metabolic models can therefore be due to missing information about regulation. Attempts have been done to integrate the regulatory network into the FBA modeling framework for *S. cerevisiae* and *E.coli* and the aim is to be able to make more accurate predictions of the metabolic fluxes.

Here we used controllability analysis to unravel topological properties and hierarchical structures of the transcriptional regulatory network, and the analysis shows that the *S. cerevisiae* transcriptional regulatory network contains circular control motifs where the transcription factors can control each other in a circular fashion, in a big internal loop. For the *E. coli* transcriptional regulatory network we don't see this pattern. Because of the topology and structure of the yeast transcriptional regulatory network it might be very hard to predict how a metabolic flux will change after certain perturbations, due to redundancy and overlap in the transcriptional regulatory network. This analysis points to the fact that regulation is complex and need to be studied more in detail in order to be able to make accurate predictions using modeling approaches.

Recently the first whole-cell *in silico* model for an organism was published, namely for the small bacteria *Mycoplasma genitalium* (Karr et al, 2012). The whole-cell model describes the cellular processes divided into 28 submodels which each describe one biological process, e.g. metabolism, transcription and translation etc. *M. genitalium* contains only 525 genes. The model was reconstructed by examining several hundreds of publications and investigating many high-throughput datasets. We are still far from being able to reconstruct a whole cell model for *S. cerevisiae* due to many reasons, e.g. much higher complexity and many more genes and components.

The long term goal and extension of the work presented in this thesis would be to connect the biological network reconstructions and be able to model the whole system together. This is not a very easy task and requires an advanced modeling framework that also can take different time scales of the different processes into account. Furthermore, these networks could be combined with other models, for example with a description of cell cycle regulation and behavior in order to try to come a step closer to a model that can describe the whole cell.

Acknowledgements

Now when this period of my life is going to an end I can look back on these years that formed my PhD and conclude that I have learnt a lot and I would like to thank all the people that have been involved in one way or another, in no particular order.

First I would like to thank my supervisor Jens Nielsen for support and encouragement during my PhD. It meant a lot to me. Thanks to my co-supervisor Intawat for support and cooking delicious Thai food. Sergio, you helped me a lot, especially during the crucial time of my PhD when I was stuck and had low motivation. At this time you stepped in and gave me new ideas and suggestions. Thank you very much!

I would also like to thank my collaborators Amir, Lifang, Flora, Jin and Rahul. I really enjoyed working with all of you and we all worked efficiently together and had many fruitful discussions.

The modeling gang, Fredrik, Rasmus, Leif, Adil, Kwanjeera, Natapol, Francesco, Manuel, Ed, Subazini, Ibrahim, Luis, Kaisa, Shaghayegh, Pouyan and Saeed (and also Christoph) have all been involved in nice discussions about modeling, bioinformatics and even statistics and also created a nice working atmosphere and good time in the MC2 building. Thanks to Gatto for introducing the tomato method.

The Thai community, P'Mint. P'Ball, P'Aey, P'In, P'Ple, P'Bon, Tangmo, Wanwipa and Pramote. We have been friends and had a lot of fun together during these years. Thanks for delicious Thai khanom, for teaching me Thai and thanks to Sakda for cooking delicious but spicy food for me.

Thanks to the Persian community for teaching me Persian and cooking delicious Persian food for me. Shaq, Rosita, Saeed, Amir x 2, Pouyan, Raya.

Martina and Dina have helped me a lot, both with the Croatian language and with all kind of practical and scientific things. Thanks a lot!

Thanks to Erica and Helena for being there and helping me with practical things.

Thanks to my former supervisor Erik, and to Joakim Larsson's research-junta for making my master studies such a nice time and continuing during my PhD.

Thanks to Christoph, Florian, Anastasia, Verena, Juan, Ilse, Heidi, Francesco, Leif, Fredrik and Rahul etc. for a lot of fun activities inside and outside the lab. Thanks to Knuf for joining me for hiking, currywurst and whisky.

Thanks to all the people in the sysbio group and all the people I forgot. It has been great working with you.

Thanks to my good old friends from Borås: Daniel, Sara, Mattias, Maria, Erik, Sara, Henrik, Andreas and Sebastian for having a lot of fun and going for skiing and other fun activities together.

And last but not least, big thanks to my parents, farmor Britten and brothers for supporting me in the hard situations. Elnaz jan, thank you a lot for being beside me and supporting me and helping me. I can't explain how much it means to me.

References

Agren R, Bordel S, Mardinoglu A, Pornputtapong N, Nookaew I, Nielsen J (2012) Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS computational biology* **8:** e1002518

Agren R, Liu L, Shoaie S, Vongsangnak W, Nookaew I, Nielsen J (2013a) The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum. *PLoS computational biology* **9:** e1002980

Agren R, Otero JM, Nielsen J (2013b) Genome-scale modeling enables metabolic engineering of Saccharomyces cerevisiae for succinic acid production. *Journal of industrial microbiology & biotechnology***:** 1-13

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2007) *Molecular biology of the cell*: Garland Science.

Aon JC, Cortassa S (2001) Involvement of nitrogen metabolism in the triggering of ethanol fermentation in aerobic chemostat cultures of Saccharomyces cerevisiae. *Metabolic Engineering* **3:** 250-264

Arnaud MB, Cerqueira GC, Inglis DO, Skrzypek MS, Binkley J, Chibucos MC, Crabtree J, Howarth C, Orvis J, Shah P (2012) The Aspergillus Genome Database (AspGD): recent developments in comprehensive multispecies curation, comparative genomics and community resources. *Nucleic acids research* **40:** D653-D659

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

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT (2000) Gene Ontology: tool for the unification of biology. *Nature genetics* **25:** 25-29

Aung HW, Henry SA, Walker LP (2013) Revising the Representation of Fatty Acid, Glycerolipid, and Glycerophospholipid Metabolism in the Consensus Model of Yeast Metabolism. *Industrial Biotechnology* **9:** 215-228

Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M (2005) The universal protein resource (UniProt). *Nucleic acids research* **33:** D154-D159

Bakker BM, Bro C, Kötter P, Luttik MAH, Van Dijken JP, Pronk JT (2000) The mitochondrial alcohol dehydrogenase Adh3p is involved in a redox shuttle in Saccharomyces cerevisiae. *Journal of bacteriology* **182:** 4730-4737

Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *science* **286:** 509-512

Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007) High-resolution profiling of histone methylations in the human genome. *Cell* **129:** 823-837

Barua D, Kim J, Reed JL (2010) An automated phenotype-driven approach (GeneForce) for refining metabolic and regulatory models. *PLoS computational biology* **6:** e1000970

Becker SA, Palsson BO (2008) Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol* **4:** e1000082

Bordel S, Agren R, Nielsen J (2010) Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. *PLoS computational biology* **6:** e1000859

Botstein D, Chervitz SA, Cherry JM (1997) Yeast as a model organism. *Science (New York, NY)* **277:** 1259

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

Brochado AR, Matos C, Moller BL, Hansen J, Mortensen UH, Patil KR (2010) Improved vanillin production in baker's yeast through in silico design. *Microbial Cell Factories* **9**

Burda P, Aebi M (1999) The dolichol pathway of N-linked glycosylation. *Biochimica et biophysica acta* **1426:** 239-257

Chandrasekaran S, Price ND (2010) Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **107:** 17845-17850

Chen F, Mackey AJ, Stoeckert CJ, Roos DS (2006) OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic acids research* **34:** D363-D368

Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR (2012) Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic acids research* **40:** D700-D705

Cimini D, Patil KR, Schiraldi C, Nielsen J (2009) Global transcriptional response of Saccharomyces cerevisiae to the deletion of SDH3. *Bmc Systems Biology* **3**

Coles SJ, Day NE, Murray-Rust P, Rzepa HS, Zhang Y (2005) Enhancement of the chemical semantic web through the use of InChI identifiers. *Organic & biomolecular chemistry* **3:** 1832-1834

Colijn C, Brandes A, Zucker J, Lun DS, Weiner B, Farhat MR, Cheng T-Y, Moody DB, Murray M, Galagan JE (2009) Interpreting expression data with metabolic flux models: predicting Mycobacterium tuberculosis mycolic acid production. *PLoS computational biology* **5:** e1000489

Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JLY, Toufighi K, Mostafavi S (2010) The genetic landscape of a cell. *Science Signalling* **327:** 425

Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO (2004) Integrating high-throughput and computational data elucidates bacterial networks. *Nature* **429:** 92-96

Covert MW, Schilling CH, Palsson B (2001) Regulation of gene expression in flux balance models of metabolism. *J Theor Biol* **213:** 73-88

Covert MW, Xiao N, Chen TJ, Karr JR (2008) Integrating metabolic, transcriptional regulatory and signal transduction models in Escherichia coli. *Bioinformatics* **24:** 2044-2050

Cox SJ, Shalel Levanon S, Bennett GN, San KY (2005) Genetically constrained metabolic flux analysis. *Metab Eng* **7:** 445-456

Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M, McNaught A, Alcántara R, Darsow M, Guedj M, Ashburner M (2008) ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic acids research* **36:** D344-D350

Demain AL, Vaishnav P (2009) Production of recombinant proteins by microbes and higher organisms. *Biotechnology advances* **27:** 297-306

Dobson P, Smallbone K, Jameson D, Simeonidis E, Lanthaler K, Pir P, Lu C, Swainston N, Dunn W, Fisher P, Hull D, Brown M, Oshota O, Stanford N, Kell D, King R, Oliver S, Stevens R, Mendes P (2010) Further developments towards a genome-scale metabolic model of yeast. *BMC Systems Biology* **4:** 145

Duarte N, Herrgard M, Palsson B (2004) Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome research* **14:** 1298 - 1309

Edwards J, Palsson B (2000) The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences of the United States of America* **97:** 5528

Eisen M, Spellman P, Brown P, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences* **95:** 14863

Erdős P, Renyi A (1961) On the strength of connectedness of a random graph. *Acta Mathematica Hungarica* **12:** 261-267

Feizi A, Österlund T, Petranovic D, Bordel S, Nielsen J (2013) Genome-Scale Modeling of the Protein Secretory Machinery in Yeast. *PloS one* **8:** e63284

Forster J, Famili I, Fu P, Palsson B, Nielsen J (2003) Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. *Genome research* **13:** 244 - 253

Förster J, Famili I, Palsson BØ, Nielsen J (2003) Large-scale evaluation of in silico gene deletions in Saccharomyces cerevisiae. *OMICS A Journal of Integrative Biology* **7:** 193-202

Gama-Castro S, Salgado H, Peralta-Gil M, Santos-Zavaleta A, Muñiz-Rascado L, Solano-Lira H, Jimenez-Jacinto V, Weiss V, García-Sotelo JS, López-Fuentes A (2011) RegulonDB version 7.0: transcriptional regulation of Escherichia coli K-12 integrated within genetic sensory response units (Gensor Units). *Nucleic acids research* **39:** D98-D105

Gianchandani EP, Papin JA, Price ND, Joyce AR, Palsson BO (2006) Matrix formalism to describe functional states of transcriptional regulatory systems. *PLoS Comput Biol* **2:** e101

Gombert AK, Moreira dos Santos M, Christensen B, Nielsen J (2001) Network identification and flux quantification in the central metabolism of Saccharomyces cerevisiae under different conditions of glucose repression. *Journal of bacteriology* **183:** 1441-1451

Graf A, Dragosits M, Gasser B, Mattanovich D (2009) Yeast systems biotechnology for the production of heterologous proteins. *FEMS yeast research* **9:** 335-348

Grimme SJ, Westfall BA, Wiedman JM, Taron CH, Orlean P (2001) The essential Smp3 protein is required for addition of the side-branching fourth mannose during assembly of yeast glycosylphosphatidylinositols. *The Journal of biological chemistry* **276:** 27731-27739

Gross E, Sevier CS, Heldman N, Vitu E, Bentzur M, Kaiser CA, Thorpe C, Fass D (2006) Generating disulfides enzymatically: reaction products and electron acceptors of the endoplasmic reticulum thiol oxidase Ero1p. *Proceedings of the National Academy of Sciences of the United States of America* **103:** 299-304

Hahn S, Young ET (2011) Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* **189:** 705-736

Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature* **431:** 99-104

Heavner BD, Smallbone K, Barker B, Mendes P, Walker LP (2012) Yeast 5–an expanded reconstruction of the Saccharomyces cerevisiae metabolic network. *BMC systems biology* **6:** 55

Herrgard M, Swainston N, Dobson P, Dunn W, Arga K, Arvas M, Bluthgen N, Borger S, Costenoble R, Heinemann M (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nature Biotechnology* **26:** 1155 - 1160

Herrgård M, Lee B, Portnoy V, Palsson B (2006) Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae. *Genome research* **16:** 627

Hou J, Tyo KE, Liu Z, Petranovic D, Nielsen J (2012) Metabolic engineering of recombinant protein secretion by Saccharomyces cerevisiae. *FEMS yeast research* **12:** 491-510

Ibarra RU, Edwards JS, Palsson BO (2002) Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* **420:** 186-189

Ideker T, Galitski T, Hood L (2001) A new approach to decoding life: systems biology. *Annual review of genomics and human genetics* **2:** 343-372

Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409:** 533-538

Jensen ON (2006) Interpreting the protein language using proteomics. *Nature Reviews Molecular Cell Biology* **7:** 391-403

Jensen P, Lutz K, Papin J (2011) TIGER: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. *BMC systems biology* **5:** 147

Jensen PA, Papin JA (2011) Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics* **27:** 541-547

Jewett MC, Workman CT, Nookaew I, Pizarro FA, Agosin E, Hellgren LI, Nielsen J (2013) Mapping Condition-Dependent Regulation of Lipid Metabolism in Saccharomyces cerevisiae. *G3: Genes| Genomes| Genetics* **3:** 1979-1995

Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita K, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Research* **34:** D354 - D357

Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B, Assad-Garcia N, Glass JI, Covert MW (2012) A whole-cell computational model predicts phenotype from genotype. *Cell* **150:** 389-401

Knuf C, Nielsen J (2012) Aspergilli: Systems biology and industrial applications. *Biotechnology Journal* **7:** 1147-1155

Kuepfer L, Sauer U, Blank LM (2005) Metabolic functions of duplicate genes in Saccharomyces cerevisiae. *Genome Res* **15:** 1421-1430

Lachmann A, Xu H, Krishnan J, Berger SI, Mazloom AR, Ma'ayan A (2010) ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26:** 2438-2444

Lee JM, Gianchandani EP, Eddy JA, Papin JA (2008) Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLoS Comput Biol* **4:** e1000086

Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I (2002) Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science* **298:** 799-804

Lewis NE, Nagarajan H, Palsson BO (2012) Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology* **10:** 291-305

Lidén G, Persson A, Niklasson C, Gustafsson L (1995) Energetics and product formation by Saccharomyces cerevisiae grown in anaerobic chemostats under nitrogen limitation. *Applied microbiology and biotechnology* **43:** 1034-1038

Lieb JD, Liu X, Botstein D, Brown PO (2001) Promoter-specific binding of Rap1 revealed by genome-wide maps of protein–DNA association. *Nature genetics* **28:** 327-334

Liu Y-Y, Slotine J-J, Barabási A-L (2011) Controllability of complex networks. *Nature* **473:** 167-173

Liu Z, Tyo KE, Martínez JL, Petranovic D, Nielsen J (2012) Different expression systems for production of recombinant proteins in Saccharomyces cerevisiae. *Biotechnology and bioengineering* **109:** 1259-1268

Liu Z, Österlund T, Hou J, Petranovic D, Nielsen J (2013) Anaerobic α-amylase production and secretion with fumarate as the final electron acceptor in yeast. *Applied and environmental microbiology*

Lubertozzi D, Keasling JD (2009) Developing< i> Aspergillus</i> as a host for heterologous expression. *Biotechnology advances* **27:** 53-75

Mahadevan R, Lovley D (2008) The degree of redundancy in metabolic genes is linked to mode of metabolism. *Biophysical journal* **94:** 1216-1220

Mo ML, Palsson BØ, Herrgård MJ (2009) Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC systems biology* **3:** 37

Nguyen VD, Saaranen MJ, Karala A-R, Lappi A-K, Wang L, Raykhel IB, Alanen HI, Salo KE, Wang C-c, Ruddock LW (2011) Two endoplasmic reticulum PDI peroxidases increase the efficiency of the use of peroxide during disulfide bond formation. *Journal of molecular biology* **406:** 503-515

Nissen TL, Schulze U, Nielsen J, Villadsen J (1997) Flux distributions in anaerobic, glucose-limited continuous cultures of Saccharomyces cerevisiae. *Microbiology* **143:** 203-218

Nookaew I, Jewett M, Meechai A, Thammarongtham C, Laoteng K, Cheevadhanarak S, Nielsen J, Bhumiratana S (2008) The genome-scale metabolic model iIN800 of Saccharomyces cerevisiae and its validation: a scaffold to query lipid metabolism. *BMC Systems Biology* **2:** 71

Oliveira A, Patil K, Nielsen J (2008) Architecture of transcriptional regulatory circuits is knitted over the topology of bio-molecular interaction networks. *BMC Systems Biology* **2:** 17

Orth JD, Palsson BØ (2010) Systematizing the generation of missing metabolic knowledge. *Biotechnology and bioengineering* **107:** 403-412

Overkamp KM, Bakker BM, Kötter P, Van Tuijl A, De Vries S, Van Dijken JP, Pronk JT (2000) In Vivo Analysis of the Mechanisms for Oxidation of Cytosolic NADH by Saccharomyces cerevisiaeMitochondria. *Journal of bacteriology* **182:** 2823-2830

Palsson B (2006) Properties of Reconstructed Networks. *Cambridge: Systems Biology*

Papini M, Nookaew I, Scalcinati G, Siewers V, Nielsen J (2010) Phosphoglycerate mutase knock out mutant Saccharomyces cerevisiae: Physiological investigation and transcriptome analysis. *Biotechnology Journal*

Patil K, Nielsen J (2005) Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proceedings of the National Academy of Sciences* **102:** 2685-2689

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* **8:** 785-786

Pramanik J, Keasling J (1997) Stoichiometric model of Escherichia coli metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnology and bioengineering* **56:** 398-421

Price ND, Reed JL, Palsson BØ (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews Microbiology* **2:** 886-897

Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E (2000) Genome-wide location and function of DNA binding proteins. *Science* **290:** 2306-2309

Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature methods* **4:** 651-657

Sauer U, Heinemann M, Zamboni N (2007) Getting closer to the whole picture. *Science(Washington)* **316:** 550-551

Schekman R (2010) Charting the secretory pathway in a simple eukaryote. *Molecular biology of the cell* **21:** 3781-3784

Schilling CH, Palsson BØ (2000) Assessment of the Metabolic Capabilities of *Haemophilus influenzae Rd* through a Genome-scale Pathway Analysis. *Journal of theoretical biology* **203:** 249-283

Schilling CH, Schuster S, Palsson BO, Heinrich R (1999) Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnology progress* **15:** 296-303

Schomburg I, Chang A, Schomburg D (2002) BRENDA, enzyme data and metabolic information. *Nucleic acids research* **30:** 47-49

Schuetz R, Zamboni N, Zampieri M, Heinemann M, Sauer U (2012) Multidimensional optimality of microbial metabolism. *Science* **336:** 601-604

Schuster S, Hilgetag C (1994) On elementary flux modes in biochemical reaction systems at steady state. *Journal of Biological Systems* **2:** 165-182

Sevier CS, Kaiser CA (2002) Formation and transfer of disulphide bonds in living cells. *Nature reviews Molecular cell biology* **3:** 836-847

Shlomi T, Cabili MN, Herrgard MJ, Palsson BO, Ruppin E (2008) Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol* **26:** 1003-1010

Shlomi T, Eisenberg Y, Sharan R, Ruppin E (2007) A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol Syst Biol* **3:** 101

Stephanopoulos G, Aristidou AA, Nielsen J (1998) *Metabolic engineering: principles and methodologies*: Academic Press.

Tai SL, Boer VM, Daran-Lapujade P, Walsh MC, de Winde JH, Daran JM, Pronk JT (2005) Two-dimensional Transcriptome Analysis in Chemostat Cultures - Combinatorial Effects of Oxygen Availability and Macronutrient Limitation in Saccharomyces cerevisiae. *Journal of Biological Chemistry* **280:** 437-447

Tai SL, Daran-Lapujade P, Luttik MAH, Walsh MC, Diderich JA, Krijger GC, van Gulik WM, Pronk JT, Daran JM (2007) Control of the glycolytic flux in Saccharomyces cerevisiae grown at low temperature. *Journal of Biological Chemistry* **282:** 10243-10251

Teixeira MC, Monteiro P, Jain P, Tenreiro S, Fernandes AR, Mira NP, Alenquer M, Freitas AT, Oliveira AL, Sá-Correia I (2006) The YEASTRACT database: a tool for the analysis of transcription regulatory associations in Saccharomyces cerevisiae. *Nucleic acids research* **34:** D446-D451

Thiele I, Palsson BØ (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature protocols* **5:** 93-121

Tu BP, Ho-Schleyer SC, Travers KJ, Weissman JS (2000) Biochemical basis of oxidative protein folding in the endoplasmic reticulum. *Science* **290:** 1571-1574

Tyo KE, Liu Z, Petranovic D, Nielsen J (2012) Imbalance of heterologous protein folding and disulfide bond formation rates yields runaway oxidative stress. *BMC biology* **10:** 16

Usaite R, Patil KR, Grotkjær T, Nielsen J, Regenberg B (2006) Global transcriptional and physiological responses of Saccharomyces cerevisiae to ammonium, L-alanine, or L-glutamine limitation. *Applied and environmental microbiology* **72:** 6194-6203

Walsh G (2010) Biopharmaceutical benchmarks 2010. *Nature biotechnology* **28:** 917

Wang B, Guo G, Wang C, Lin Y, Wang X, Zhao M, Guo Y, He M, Zhang Y, Pan L (2010) Survey of the transcriptome of Aspergillus oryzae via massively parallel mRNA sequencing. *Nucleic acids research* **38:** 5075-5087

Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM (2009) A census of human transcription factors: function, expression and evolution. *Nature Reviews Genetics* **10:** 252-263

Ward OP (2012) Production of recombinant proteins by filamentous fungi. *Biotechnology advances* **30:** 1119-1139

Varma A, Palsson BO (1994) Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Bio/technology* **12**

Vemuri G, Eiteman M, McEwen J, Olsson L, Nielsen J (2007) Increasing NADH oxidation reduces overflow metabolism in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences* **104:** 2402-2407

Weng S, Dong Q, Balakrishnan R, Christie K, Costanzo M, Dolinski K, Dwight S, Engel S, Fisk D, Hong E (2003) Saccharomyces Genome Database (SGD) provides biochemical and structural information for budding yeast proteins. *Nucleic Acids Research* **31:** 216 - 218

Yizhak K, Benyamini T, Liebermeister W, Ruppin E, Shlomi T (2010) Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics* **26:** i255-i260

Zambelli F, Prazzoli GM, Pesole G, Pavesi G (2012) Cscan: finding common regulators of a set of genes by using a collection of genome-wide ChIP-seq datasets. *Nucleic acids research* **40:** W510-W515

Österlund T, Nookaew I, Bordel S, Nielsen J (2013) Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling *BMC systems biology*

Österlund T, Nookaew I, Nielsen J (2012) Fifteen years of large scale metabolic modeling of yeast: developments and impacts. *Biotechnol Adv* **30:** 979-988