CHALMERS

In silico Analysis of Transcription Factor Mutants in *Saccharomyces cerevisiae*

Master of Science in Bioinformatics and Systems Biology

Lokeshwaran Manoharan

Supervisor: Asst. Prof. Sergio Bordel Velasco

Examiner: Prof. Jens Nielsen

Department of Systems Biology

Chalmers University of Technology

Gothenburg, Sweden SE- 412 96

December 2011.

Abstract

Transcriptional regulation of metabolism is considered to be one of the key aspects in controlling the behavior of the cells in response to different environmental conditions. It is important to figure out which transcription factors are involved in regulating the metabolic process in which condition to better understand the relation between regulation and metabolism. This has been addressed through this study in *Saccharomyces cerevisiae* primarily by computational simulations using regulatory genome-scale model containing regulatory interactions between 961 metabolites [348 genes and 57 transcription factors]. This regulatory model was simulated for gene expression predictions in five different environmental conditions for all available TF mutants and wildtype which accounts for [(57+1)*5] 290 assays/conditions. These gene expression predictions from the regulatory model in those assays are then integrated with the corresponding metabolic model [comprising 798 metabolites, 619 genes] optimizing for maximal growth.

While the integration of other conditions were mostly unsuccessful, it was the regulatory predictions from the minimal media condition with galactose as the carbon source which seemed to properly simulate the crab-tree effect in case of the wildtype observed through the flux of ethanol production. It was found that only for the TF-KO mutant ΔMET4 there was a tremendous shift in metabolism from respiration towards fermentation in this particular condition. There were also several inconsistencies with the similar experimental research work achieved through ^{13}C flux analysis of TF mutants in *Saccharomyces cerevisiae.* It was understood from this study that the regulatory model is highly incomplete. A more complete regulatory model would be absolutely necessary for proper integration of regulatory and metabolic models that can emulate real-time experiments.

Acknowledgement

This project was carried in Nielsen's lab for Systems and Synthetic Biology at Chalmers University of Technology. I would like to express my appreciation to everyone who supported me throughout my Master thesis work. Mainly, I would like to thank Prof. Jens Nielsen who has given me this wonderful opportunity to work in his lab with amazing infrastructure. I owe my deepest gratitude to my supervisor Asst. Prof. Sergio Bordel Velasco for his guidance, ideas and encouragement he has given me throughout this thesis work. I would also like to thank the PhD students Tobias Osterlund, Rasmus Agren and Amir Feizi for their assistance, guidance and for sharing their experiences. I would also like to thank my friend Saber Ahmad Akhondi who helped me with designing and structuring my thesis report. Finally, I would like to dedicate this work to my parents who have been very supportive to me all through my education from my childhood.

Abbreviations

CL – Carbon Limited condition

COBRA – Constraint Based Reconstruction and Analysis

EPA – Extreme Pathway Analysis

FBA – Flux balance Analysis

GLAC – Galactose as carbon source in minimal media condition

KO – Knock-Out genes

LP – Linear Programming

MM – Minimal Media with Glucose as carbon source

NL – Nitrogen Limited condition

PCA – Principal Component Analysis

rFBA – Regulatory Flux Balance Analysis

S Matrix – Stoichiometric Matrix

SBML – Systems Biology Markup Language

SBT – Sorbitol with glucose in minimal media

SR-FBA – Steady-state Regulatory Flux balance Analysis

SVD – Singular Value Decomposition

TF – transcription Factor

TRN – Transcriptional Regulatory Network

TRS – Transcriptional Regulatory System

Contents

1 Introduction

It is a well studied fact that metabolism and regulation are the main factors involved in the growth and prosperity of the living cells. Metabolism mainly involves in converting the nutrients available from the environment into desirable products for the survival of the cells. Regulation involves in changing the metabolic processes according to the compounds and conditions present in the environment. It is to be noted that regulation through transcription factors hold most of the regulatory processes which are required in different conditions. It is necessary to analyze both regulation and metabolism together to better understand the behavior of the living cells in different conditions. This project aims to address these questions related to the transcriptional regulation and their relation to the metabolic processes in Yeast *Saccharomyces cerevisiae.*

1.1 Background

It is of primary importance to understand the mechanism behind the behavior of a cell, so that it would be easy to modulate the cells according to the required perspectives of the researchers (Csete and Doyle 2002). These regulatory mechanisms are generally studied from the gene expression and protein binding profiles (Karlebach and Shamir 2008; Snyder and Gallagher 2009). It has also been shown that this information is not sufficient to understand the regulation process completely (Bonneau 2008). So, there is a need for a more comprehensive study of the regulation processes and their mechanisms.

It has also been discussed below [section 1.4.3] that transcriptional regulation is considered to be the main mechanism of regulation and which has also been studied extensively in the recent years through many different techniques in *Saccharomyces cerevisiae* (Lee et al. 2002; Chen et al. 2004; Harbison et al. 2004). Although severe research is focused on understanding the transcription regulation mechanism through protein binding and gene expression analysis, the inference of these mechanisms and their integration to metabolism is still unclear.

The recent explosion in the field of sequencing technologies and high-throughput analysis of cellular processes has opened the fields like computational biology and bioinformatics to a larger extent (Yang et al. 2009). The computational analysis of biochemical processes have proved to be efficient in predicting and providing interesting insights especially in the field of metabolic engineering (Edwards et al. 2001). Computational systems biology is considered as the promising field in understanding of life and possible discoveries for human prosperity (Kitano 2002).

1.2 Aim

The primary aim of this project is to avail the computational techniques and genome scale models to address the questions related to transcription regulation of metabolism in Yeast *Saccharomyces cerevisiae*.

1.3 Outline

The project was carried out in two large steps which are analyzing the transcription factor mutants in regulatory model in five different conditions and then to integrate the results with the metabolic model. The first step which is to analyze the transcription factor mutants, the regulatory

mechanistic model designed by the chip-chip method (Harbison et al. 2004) containing most of the transcriptional rules involved in the transcriptional regulation in *Saccharomyces cerevisiae* was used. This model was analyzed by a method where certain way of matrix formalism of the reactions was accomplished which will be explained in detail below [section 2.1].

With the regulatory model, the expression of a gene can be predicted by means of utilizing the flux balance analysis [FBA] achieved by linear programming [LP]. The analysis of each transcription factor mutant in each condition is achieved by predicting the expression state of each gene in the regulatory model for every combination of mutant and condition. This was achieved with the help of MATLAB and COBRA toolbox [section 2.2] to make and analyze the mutants in each condition.

The expression profiles of each mutant in each condition are analyzed comparatively with that of the wildtypes in each condition. This would help in understanding the function of each transcription factor in each condition in the regulatory mechanism of *Saccharomyces cerevisiae.* The entire expression profiles from all the conditions are combined together and a statistical analysis of the data [section 3.3] has been done. This provides the overall understanding of the behavior of the organism according to the changes in the environment. Later, the results from the regulatory model are then integrated with the respective metabolic model for phenotypic changes. This can be achieved by taking genes from the regulatory model and fixing the bound values of all the reactions related to those genes and simulating for maximal growth rate.

1.4 Theory

Understanding the mechanisms that control the activities and their ability to adapt to the environmental conditions of a microbial cell are the main areas where most of the biological research fields are focused on for the past few years. This is because the above mentioned mechanisms play an important role in cases such as medical discoveries and industrial production of biochemical compounds by microbial cells (Haynes and Silver 2009; Kreeger and Lauffenburger 2010). These mechanisms can be explained by means of a set of chemical reactions known as metabolism and their regulation.

1.4.1 Yeast

Saccharomyces cerevisiae is a single celled lower eukaryote which has been considerably studied extensively for a very long period of time as a model organism in the fields of medicine and biochemical industries (Barnett 2003). The study of biochemical pathways in *Saccharomyces cerevisiae* related to their phenotype, genetic variations and various other factors have been well progressed through the years (Nevoigt 2008). The technologies such as genetic engineering and transformation have been made highly efficient over the years(Cereghino and Cregg 1999) and the availability of other information such as biomarkers and reporter genes has been well studied and accumulated for this particular micro-organism (Gietz and Woods 2001; Gueldener et al. 2002). The current application of Yeast in different fields of biotechnology is explained in the Figure 1 below.

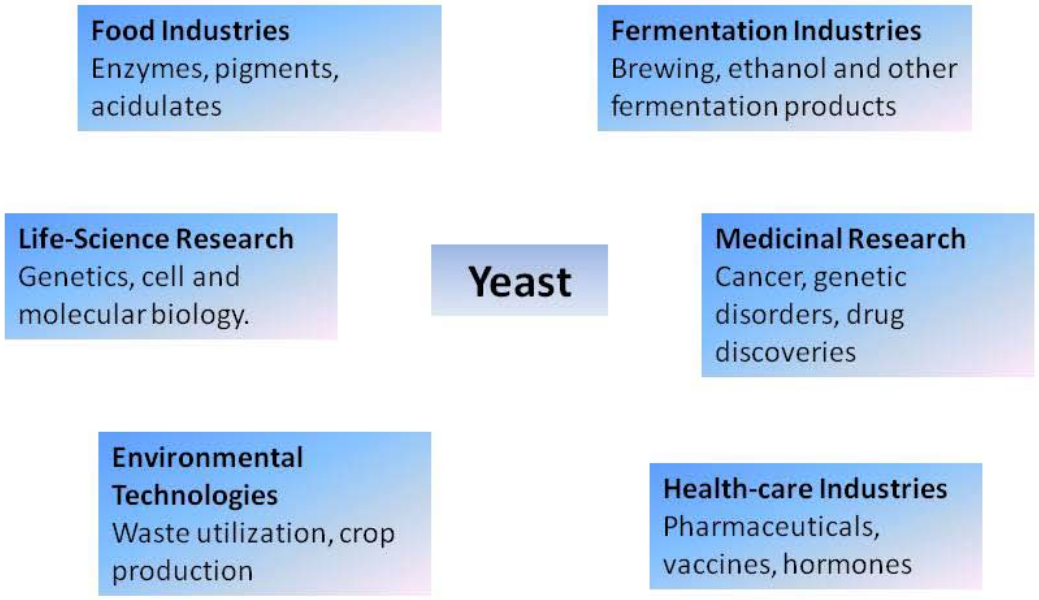


Figure 1- Varied applications of yeast. (C.Guthrie and Fink 1991; J.R. Broach 1991)

The availability and development of sequencing technologies have made a significant impact over the genomic studies of *Saccharomyces cerevisiae.* It adds to the well known fact that *Saccharomyces cerevisiae* was the first eukaryotic organism whose complete genome was sequenced (Goffeau et al. 1996). There are several databases which have been developed regarding genome of *Saccharomyces cerevisiae* notably [http://www.yeastgenome.org] and several knockout strains has been developed and used for the functional analysis of this organism (Winzeler et al. 1999). The metabolic engineering strategies have also been applied to this organism to get this strain improved in various cellular functions (Ostergaard et al. 2000) with the introduction of this field (Bailey et al. 1990).

1.4.2 Metabolism

Metabolism is a process by which any cell survives and maintains itself by means of consuming the nutrients available in the environment of the cell. Metabolism can be represented by means of a whole set of chemical reactions that occur inside the cell. These chemical reactions are interconnected both in serial and in parallel by other chemical reactions and catalyzed by proteins known as enzymes. These chemical reactions contribute the metabolic pathways by which one compound is converted to another which is further converted to another and so on. The central carbon metabolism of *Saccharomyces cerevisiae* is well illustrated in [Figure 2].

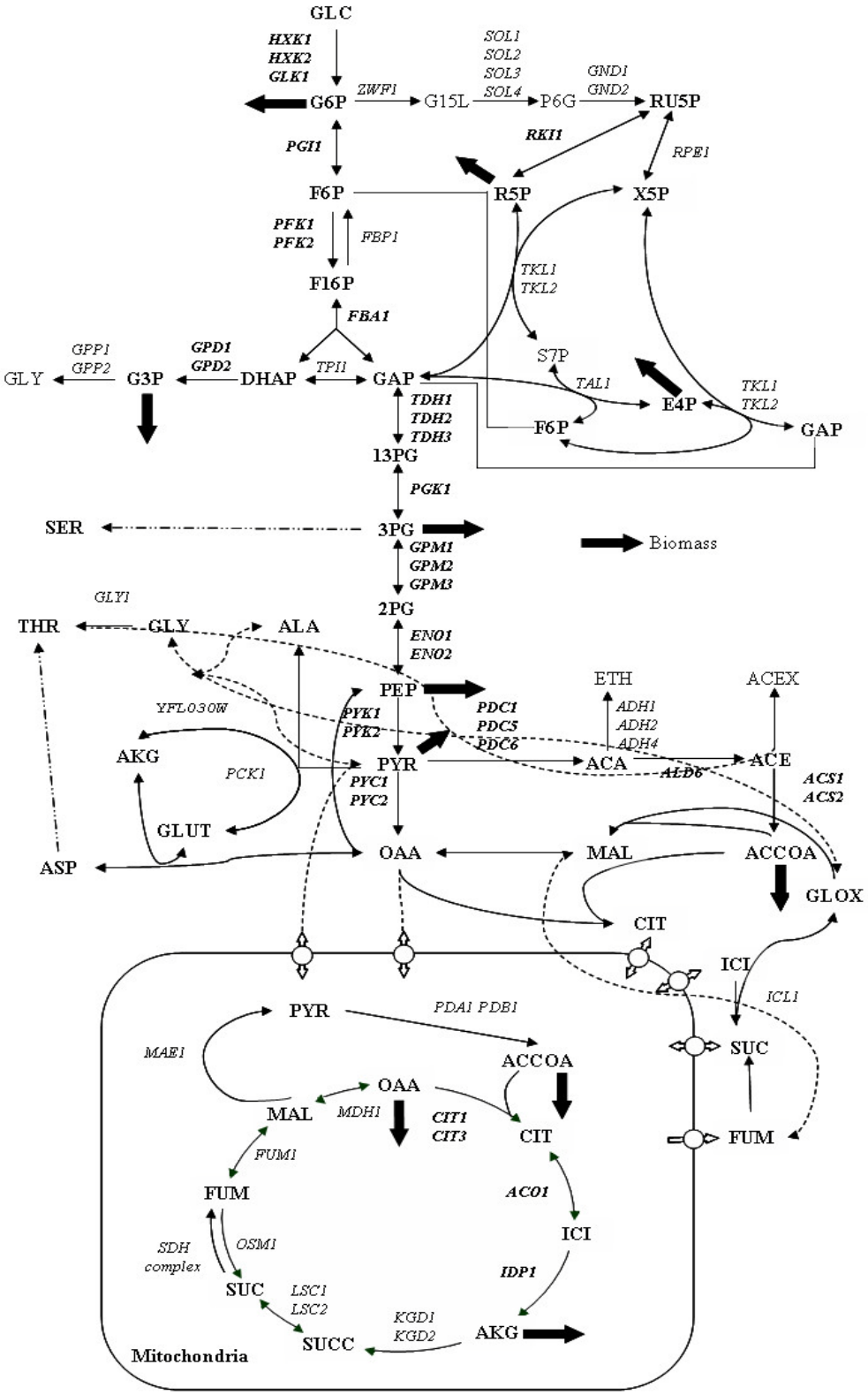


Figure 2: The central carbon metabolism in *Saccharomyces cerevisiae* obtained from Biomedcentral database. [http://Biomedcentral.com] (Patil et al. 2005)

1.4.3 Regulation

Regulation comes to play its part mainly to deal with interactions between the cell and its environmental conditions. The metabolic pathway is controlled by a set of chemical reactions based on the existence of the type of compounds in its surroundings and their properties. These two regulatory mechanisms [allosteric and transcriptional] are the most targeted pathways of study. The third type of regulation is the post-transcriptional modification of enzymes achieved by means of processes such as phosphorylation, glycosylation and so on (Desvergne et al. 2006).

These set of chemical reactions which control the activity and result of metabolism are known as Regulatory pathways. These pathways are responsible for controlling the concentration of metabolites and enzymatic activities. The interplay between these two factors results in the change of metabolic pathway from one direction to the other and so on depending on the conditions. These two types of regulation which control the concentration of metabolites and enzymatic activities are explained in terms of enzymatic and transcriptional regulation.

Enzymatic regulation can be explained in terms of increase or decrease in the activity of the enzyme which results on its respective metabolite concentration (Fell and Thomas 1995). Transcriptional regulation of metabolism is achieved by means of its effect to increase or decrease the expression of a metabolic gene, even sometimes to completely shut the gene. Transcriptional regulation is most widely considered as the primary factor for the resulting concentrations of protein levels (Krappmann et al. 2000).

1.4.4 Transcription Factors

Transcription factors are the regulating proteins where they bind to specific genetic DNA regions and control the process of transcription from DNA to mRNA. They either activate the process of transcription or repress it and they achieve this by means of allowing or restricting the RNA polymerase (Karin 1990; Latchman 1997). It is also to be noted that they usually have more than one DNA binding sites where sometimes two or more transcription factors together activate or repress the expression of one particular gene. There are also other substances like kinases and methylases which take part in the regulation of gene expression, but they lack any kind of DNA binding domain which is why they are not classified as transcription factors.

Transcription factors are also considered to be the main factors that involve in differentiation of cells into specific tissues (Olson 1990). Malfunction in transcription factors can cause number of diseases; hence the understanding of its mechanism in regulation is of very important. The absence of action of a particular transcription factor ‘PitI' seemed to have caused dwarfism and pituitary malfunctions in human (Radovick et al. 1992). There are also evidences relating malfunction of transcription factors causing genes to become oncogenes causing cancer in human.

1.4.5 Genome Scale Models

The study of cellular behavior according to the conditions [constraints] that define the allowable changes to the phenotype has become important and achievable through computational simulations in the past few years. The ability to reconstruct the metabolic networks with the help of whole-genome sequencing technologies have made a significant impact on *in silico* analysis of

micro-organisms (Covert et al. 2001; Edwards et al. 2002). These *in silico* models can be utilized and analyzed for various simulations and predictions related to biochemical pathways.

The frame work where a model can be analyzed with the help of constraints is named as COBRA – Constraint Based Reconstruction and Analysis. Basically there are four kinds of constraints explained (Price et al. 2004) which are physico-chemical constraints, topobiological constraints, environmental constraints and regulatory constraints. Having these constraints, the metabolic model can be simulated for a solution space. Then these models can be studied for different factors by changing the constraints which eventually would reduce the solution space and the pathways can be analyzed in those regions.

Mathematically, these constraints can be given as balances and bounds where balances are constraints of conserved values and bounds are the limiting values of the parameters. The best example to explain balance is the mass balance which is assumed in every metabolic model where the combination of Stoichiometry S and the velocity v is always maintained as $S.\ v = 0.$ Here 'v' is given as bounds where the velocity of each pathway can vary between given lower and upper bound values. By this way concentration and fluxes of all reactions can be calculated where the model tries to maximize one particular reaction with the given constraints.

Having an objective function for example the excretion of biomass in a metabolic model, the framework will try to maximize that objective function by means of a mathematical concept known as Linear Programming [LP]. From here, the optimal fluxes of all other pathways to maximize the growth rate can be calculated from the solution space. Here gene additions or deletions can also be analyzed by means of a constraint where to add a gene all the reactions in which the gene is involved are added whereas to delete they are removed from the model. Regulatory constraints can also be added to the metabolic model, where the regulatory pathways are given in Boolean form representing the expression of a particular gene is ON/OFF form along with their constraints.

2 Related Literature

Important method that has been adopted in analyzing the regulatory model and a toolbox for MATLAB which was used to study the models of *Saccharomyces cerevisiae* will be discussed in this section along with a similar experimental research work. Since the regulatory models are a bit different from the metabolic model, a different type of approach is certainly required and the discussed method is to address those aspects [Section 2.1]. A toolbox dedicated to the study of biochemical processes inside a cell most certainly comes in handy [Section 2.3]. A similar research work was carried out experimentally and published [section 2.3] which was used as a comparative tool for the whole project.

2.1 Matrix Formalism

As discussed above, the availability of multiple genome sequences significantly increased the research on the mechanism of the regulation of genes in a genome. As the number of genes and the related components involved in the regulation of metabolism is quite large, a certain framework of the transcriptional regulatory networks [TRN] was absolutely required. Varied types of networks were designed and analyzed with methods starting from Boolean (Kauffman et al. 2003), Bayesian (Friedman et al. 2000) to stochastic equations (McAdams and Arkin 1997). But these methods can only be applied to small TRNs as the availability of related data is very minimal. Although large-scale TRNs are constructed, it has not been studied thoroughly yet.

The predictive power of the above mentioned models has also found to be really low due to the paucity in relevant data. This publication (Gianchandani et al. 2006) has introduced a novel method to analyze the regulatory pathways which addresses the above mentioned limitations using a Boolean model of the TRNs. The complete transcriptional regulatory system [TRS] is one which includes the TRN along with their inputs and outputs of the particular network. There by connecting the environmental cues [inputs] to the regulatory reactions [TRN] which leads to the expression states [outputs] of the genes in the TRN.

Networks comprising the interactions between the components can be expressed in the form of graphical maps which can further be expressed mathematically in forms of incidence matrices. Matrix analysis methods have been traditionally used for the network analysis and the prediction of their states (Price et al. 2004). The formation of TRS and R_{rules} from the TRN is best described in the [Figure 3]. Figure 3A represents the whole TRS where it includes all the portions of a TRN. In Figure 3, B and C show the graphical maps of the regulatory reactions and D shows the formation of the R_{rules} matrix by combining their incidence vectors into a matrix. It is clear that the formation of R_{rules} is similar to the formation of stoichiometric matrix *S*. The mentioned example [in Figure 3] will be utilized to explain the proposed method all through this section [Section 2.1].

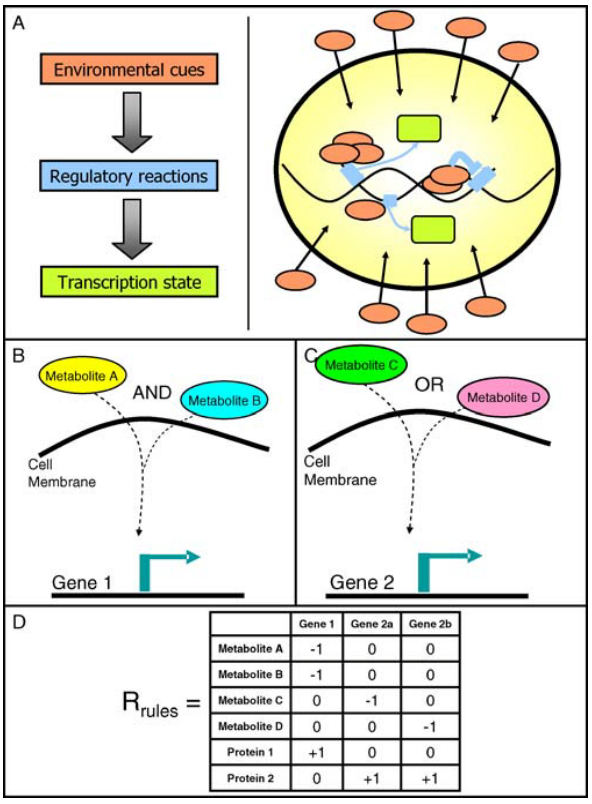
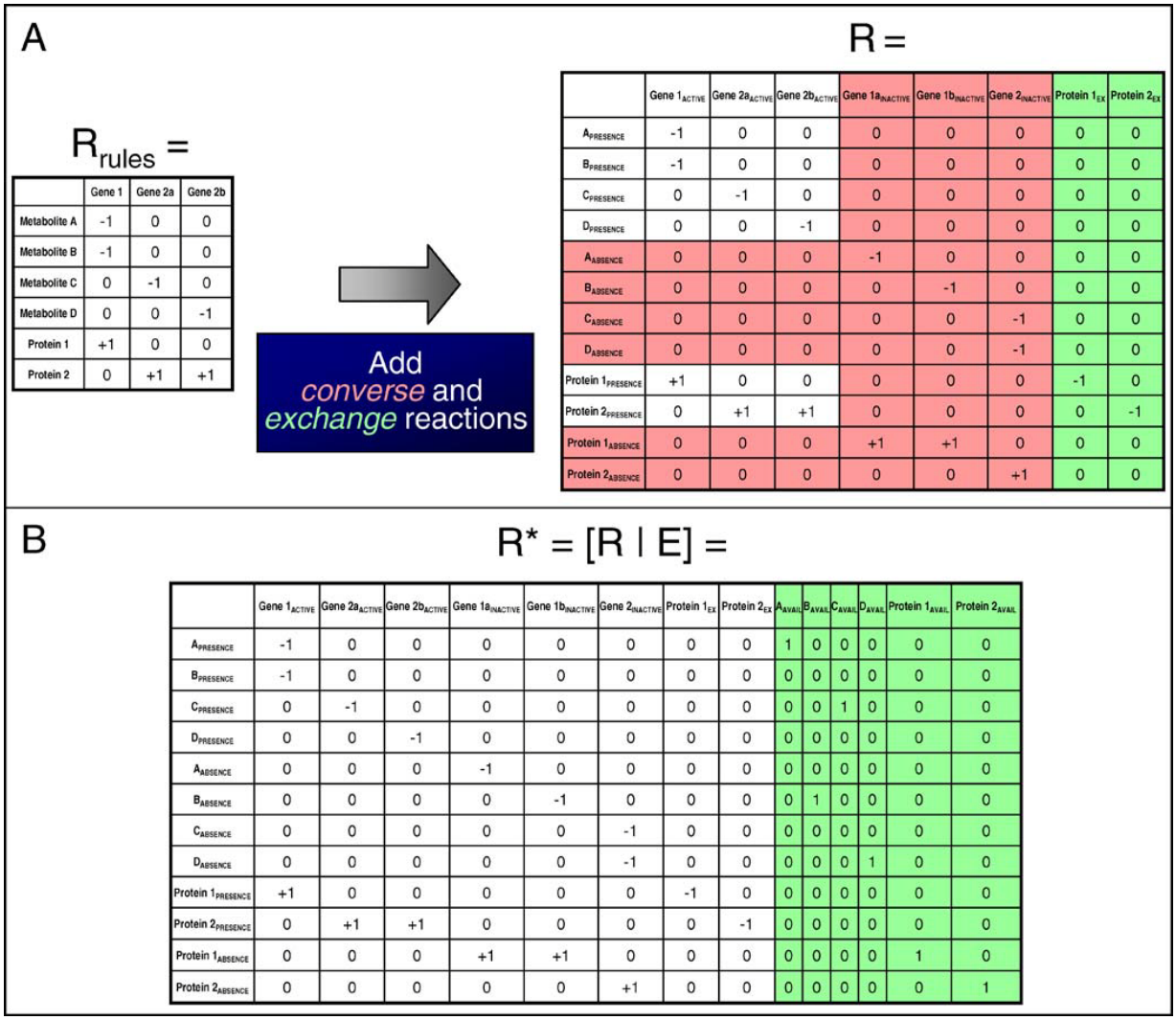


Figure 3: Formation of R_{rules} from TRN A: TRS; B and C: Intermediate Reactions; D: Matrix Formation. (Gianchandani et al. 2006)

The conversion to TRS from TRN is accomplished by including two main features of the TRN to the R_{rules} matrix. The first one being the converse of the regulatory rules being added, for example the absence of metabolites [A, B, C, D] and their related states of the genes are added. This is important because it will be easy to understand the state of a gene when a metabolite is absent or inactive because there are many cases where a gene will be active if a particular compound is absent in the system. The second feature is to include the exchange reactions by which the quasi-stoichiometric formalism can be maintained [steady state].

Basically, R matrix would have all the rules that inter-relate all the regulatory reactions inside a TRS. The matrix R* is constructed by means of adding the matrix E, which consist of all the environmental cues in the TRS. In below B [Figure 4], only the availability of metabolites and proteins involved in this TRS is mentioned. But, in reality other kinds of environmental cues like pH can also be given along with their rules. In the described example the environmental cues are the presence of metabolites A, C and the absence of proteins 1 and 2. The formation of R and R* matrices are explained for the described example in figure [Figure 4].



	Gene 1	Gene 2a	Gene 2b
Metabolite A	-1	0	0
Metabolite B	-1	0	0
Metabolite C	0	-1	0
Metabolite D	0	0	-1
Protein 1	+1	0	0
Protein 2	0	+1	+1

	$Gene\ 1_{ACTIVE}$	$Gene\ 2a_{ACTIVE}$	$Gene\ 2b_{ACTIVE}$	$Gene\ 1a_{INACTIVE}$	$Gene\ 1b_{INACTIVE}$	$Gene\ 2_{INACTIVE}$	$Protein\ 1_{EX}$	$Protein\ 2_{EX}$
$A_{PRESENCE}$	-1	0	0	0	0	0	0	0
$B_{PRESENCE}$	-1	0	0	0	0	0	0	0
$C_{PRESENCE}$	0	-1	0	0	0	0	0	0
$D_{PRESENCE}$	0	0	-1	0	0	0	0	0
$A_{ABSENCE}$	0	0	0	-1	0	0	0	0
$B_{ABSENCE}$	0	0	0	0	-1	0	0	0
$C_{ABSENCE}$	0	0	0	0	0	-1	0	0
$D_{ABSENCE}$	0	0	0	0	0	-1	0	0
$Protein\ 1_{PRESENCE}$	+1	0	0	0	0	0	-1	0
$Protein\ 2_{PRESENCE}$	0	+1	+1	0	0	0	0	-1
$Protein\ 1_{ABSENCE}$	0	0	0	+1	+1	0	0	0
$Protein\ 2_{ABSENCE}$	0	0	0	0	0	+1	0	0

	$Gene\ 1_{ACTIVE}$	$Gene\ 2a_{ACTIVE}$	$Gene\ 2b_{ACTIVE}$	$Gene\ 1a_{INACTIVE}$	$Gene\ 1b_{INACTIVE}$	$Gene\ 2_{INACTIVE}$	$Protein\ 1_{EX}$	$Protein\ 2_{EX}$	A_{AVAIL}	B_{AVAIL}	C_{AVAIL}	D_{AVAIL}	$Protein\ 1_{AVAIL}$	$Protein\ 2_{AVAIL}$
$A_{PRESENCE}$	-1	0	0	0	0	0	0	0	1	0	0	0	0	0
$B_{PRESENCE}$	-1	0	0	0	0	0	0	0	0	0	0	0	0	0
$C_{PRESENCE}$	0	-1	0	0	0	0	0	0	0	0	1	0	0	0
$D_{PRESENCE}$	0	0	-1	0	0	0	0	0	0	0	0	0	0	0
$A_{ABSENCE}$	0	0	0	-1	0	0	0	0	0	0	0	0	0	0
$B_{ABSENCE}$	0	0	0	0	-1	0	0	0	0	1	0	0	0	0
$C_{ABSENCE}$	0	0	0	0	0	-1	0	0	0	0	0	0	0	0
$D_{ABSENCE}$	0	0	0	0	0	-1	0	0	0	0	0	1	0	0
$Protein\ 1_{PRESENCE}$	+1	0	0	0	0	0	-1	0	0	0	0	0	0	0
$Protein\ 2_{PRESENCE}$	0	+1	+1	0	0	0	0	-1	0	0	0	0	0	0
$Protein\ 1_{ABSENCE}$	0	0	0	+1	+1	0	0	0	0	0	0	0	1	0
$Protein\ 2_{ABSENCE}$	0	0	0	0	0	+1	0	0	0	0	0	0	0	1

Figure 4: A: formation of matrix R by including 'converse' and 'exchange' reactions. B: formation of R* by including 'environmental cues'.(Gianchandani et al. 2006)

From this formation of R* it can be seen that the influence of environmental cues can be analyzed by means of studying the exchange reactions in this matrix. Generally R* can be used to calculate the expression states of the genes in the TRS, the same way how a stoichiometric matrix S is used [Section 3.2]. There have been many methods to approach for the analysis of Stoichiometric matrices, one such method is the extreme pathway analysis [EPA] (Schilling et al. 2000). EPA on a stoichiometric matrix S gives a series of vectors that define the edges of a convex solution space. Here, all the possible combinations of solutions will be available inside the solution space where the system maintains the steady state.

Similarly, for R* it gives a series of vectors that define the expression states of the genes inside the TRS. Mathematically, a vector **f** that satisfies **[R*].f=0** will be the expression states of the genes for the given environmental cues that is given in the R*. It is to be noted that the intermediate reactions/regulatory reactions can only have non-negative bound values. This method was also applied to a large scale model and found to work properly. With the help R* formalism the other

physical states of the TRS can also be determined from the other spaces of the matrix. To conclude, this method helps in better understanding of TRS and calculation of the expression states of the genes in any given condition.

2.2 COBRA Toolbox for MATLAB

As it has been discussed in the above mentioned sections previously [Sections 1.4.4& 2.1], the use of genome scale models to predict the cellular processes have grown to a new level in the recent years. This has been achieved from the reconstruction of metabolic and regulatory networks in terms of several biochemical reactions. These biochemical reactions can be utilized whole as a model to study different properties of the cellular functions through constraint-based analysis. This section targets the availability of a software package for Matlab dedicated to the computational simulations with these models. This Matlab toolbox called COBRA [Constraint Based Reconstruction Analysis](Becker et al. 2007) is utilized in this project for the constraint based analysis of the regulatory and metabolic models of *Saccharomyces cerevisiae.*

The protocol (Becker et al. 2007) explains the utilization of these models in the format of SBML [Systems Biology Markup Language] and to use it in varied kinds of analysis. Some of the methods used in this project and the advantages of this toolbox will be explained in this section. The predictions made with the help of this toolbox can only be treated as hypothesis which is the case with almost all kinds of computational simulations as an experimental verification is absolutely required.

This toolbox helps in finding the reaction rates of each of the biochemical reactions happening inside a cell through Flux Balance Analysis [FBA]. This is certainly the most common method used in the field of systems biology, which will be explained in detail in the following section [Section 3.2]. This toolbox helps in achieving this by means of simulating the maximization of one particular objective function or reaction which is usually the production of biomass. This can be achieved in two steps which are the formation of a stoichiometric matrix and the optimization using a linear programming method. The simulations based on gene deletions can also be predicted by means of switching off the reactions related to that particular gene.

It is also to be noted that this toolbox can also be used in the study of signal transduction models apart from metabolic and regulatory models. To conclude, the software package is very useful predicting the flux rates of the biochemical pathways quantitatively. It is also can be used to simulate in both steady-state and dynamic growth behavior. The important aspect of this package is the analysis based on the constraints, where the predictions are made with the help of this toolbox according to the environmental cues and gene deletions in this project.

2.3 Condition-Dependant Study of Regulation

It is a well established fact that transcriptional regulation holds the main key in explaining the process of regulation inside the cells. The objective of this project was to study the effects of environmental cues like the presence and absence of substances like carbon and nitrogen source over the network of transcription regulation of metabolism though computational simulations. This can be achieved through simulations of gene expressions for transcription factor deletion mutants in different environmental conditions.

A similar research work was recently carried out in studying the transcription factor deletions through ^{13}C analysis of flux distributions in central carbon metabolism of *Saccharomyces cerevisiae* experimentally (Fendt et al. 2010). The research work was carried out in studying 119 TF mutants in five different conditions through ^{13}C analysis (Sauer 2006) which is an FBA method carried out experimentally. The analysis were studied in glucose [minimal media], galactose instead of glucose, low pH, high osmolarity and urea as a nitrogen source as the five different conditions in which the analysis has been done.

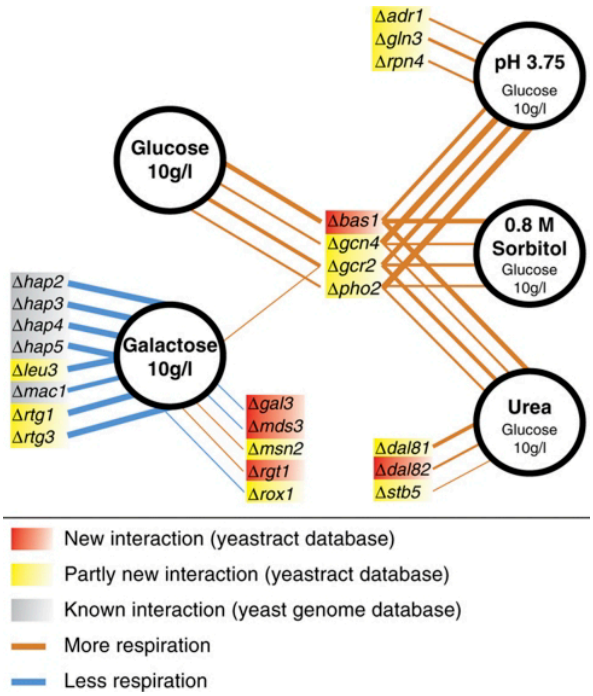


Figure 5: Transcription factors that control the steady-state flux of central carbon metabolism in Yeast. Line thickness represents the variation from wildtypes (Fendt et al. 2010).

The prediction of four new transcription factors that seem to be involved in the regulation of Yeast metabolism in all the conditions as shown in the figure above were reported [Figure 5]. It has also been shown that only very few transcription factors are involved in the process of transcription regulation meaning that most of the mutants behaved similar to that of the wildtypes. It has been predicted that these 23 transcription factors [Figure 5] seem to be involved in regulating the behavior of the cells between respiration and fermentation. This particular research has been used as a comparative tool for the computational simulations that have been carried out throughout this project work.

3 Materials and Methods

This section would mostly concentrate on the different types of methods available in the systematic study of regulation and metabolism and the methods utilized in this project. This section is designed in the same manner as the project was executed. Starting with the models and methods used to analyze the regulatory model of *Saccharomyces cerevisiae,* then the method used in analyzing the results from the regulatory model and finishing it with the process of integration of those results from the regulatory analysis to the metabolic model.

3.1 Boolean regulatory Model

Usually, the TRNs are constructed by the integration of two main methods which are genome-wide location analysis and expression analysis. The genome-wide location analysis has been achieved by myc epitope tagging of the regulators and looking for binding of these regulators to the promoter regions (Lee et al. 2002). These binding data has been statistically analyzed and the data with very significant P values are considered and the interactions obtained as a result are being developed as network motifs. This is because of the reason that usually more than one transcription factor is usually involved in the expression of a gene and also the same transcription factor is involved in activating more than one gene. These network motifs are combined together to form the large regulatory network structures.

This kind of experiment is repeated in different growth conditions to look for environmental constraints i.e., environment-based genome-wide location analysis of regulators could represent more comprehensive TRN. One such study (Harbison et al. 2004) exhibited an environment based regulatory map achieved by combining comparative genomics and genome-wide location analysis. A mechanistic regulatory model explaining the condition based regulator binding to the promoter regions of the genes was achieved.

Such a model is used in this project where it contains information like environmental conditions [inputs], the intermediate reactions along their relation with the environmental conditions [TRN] and the expression state of the associated genes [outputs]. For an extensive explanation, the inputs are mostly the presence and absence of a particular compound like glucose, galactose, and oxygen. The intermediate reactions are the biochemical interactions between the compounds available, transcription factors and other genes. The outputs are the states of genes which results from the availability of substances and the intermediate reactions.

As it is mentioned above, this model of biochemical reactions can be recognized as a TRS having all the related information of a TRN. All the biochemical reactions in the above mentioned model were constructed in accordance with the matrix formalism method explained before [Section 2.1]. It shows that all the interactions, environment and the genes are designed for the study of both presence and absence of each of the metabolites. For example, all the interaction details of a particular metabolite ‘A’ in both presence and absence of it in the model. The same case applies for every environmental compound and intracellular metabolite in the system. The upper and lower bounds of each of the reactions are also maintained according to the rules of the matrix formalism that each reaction is maintained in a Boolean formalism ‘0’ for the absence of a metabolite or a reaction and any other value representing the presence of them.

This particular design of transcriptional regulatory system helps in analyzing the various effects of different TF-KO strains in environmental conditions. This project involved in studying each TF-KO strain of *Saccharomyces cerevisiae* in five different conditions which in case of this project was carbon limited, minimal media[glucose], minimal media with galactose instead of glucose, Nitrogen limited and Minimal media with the presence of sorbitol. The expression analysis of genes in each KO strain in each condition is studied and compared with that of the wildtype in that particular condition. This helped in understanding of the involvement of transcription factors and the respective genes in each of those conditions.

3.2 Flux Balance Analysis

It has also been discussed in previous sections that the computational methods for analysis of metabolic and regulatory networks have been adapted for a long time now (Bailey 1998). The availability of information regarding the metabolic reactions [kinetics] inside a given microbe is scarce and the predictions based on the available information were not proper. Introduction of the method FBA [Flux Balance Analysis] solved most of the problems in computational analysis of biochemical pathways. This method was also exhibited and proved to simulate flux distribution of the metabolic pathways of some microbes to some extent (Edwards et al. 2001).

FBA is formulated on the requirement of only two related information about the biochemical pathway network [here metabolism] of a microorganism (Varma and Palsson 1994). First, all the biochemical reactions and their Stoichiometry related to the metabolic reactions, which is available for most of the organisms. There are also many sources that are available where this kind of information can be extracted. Second, the required process or a pathway that is to be predicted by the system. For example, biomass production can be predicted with the availability of all the biochemical reactions and their Stoichiometry with this method. The main advantage of this method is that it does not require any information about the system like metabolite concentrations or enzyme kinetics in the system.

This method works mainly on the assumption that the metabolic steady state is adapted by the system throughout the process of simulation. This metabolic steady state means that all the metabolites that are formed in the biochemical pathways of metabolism during the intermediate processes are either up-taken or eliminated by the system. This metabolic steady state is also assumed to be true with respect to the experimental techniques as well. This can be explained with the help of the stoichiometric matrix *S*, which contains all the stoichiometric coefficients of all the reactions and the metabolites in the system. Typically, the columns of S matrix contain all the reactions and the rows contain all the metabolites involved in the system. The product of this stoichiometric matrix to the vector containing all the metabolic reaction rates [*v*] will be equal to the metabolite uptake rates [*b*].

The hypothesis and the mathematical explanation the metabolic steady state is best described with the example of a simple system containing two reactions in the figure [Figure 6]. The stoichiometric matrix with rows containing the metabolites and columns containing the reactions is filled with the corresponding stoichiometric coefficients of the system. The vector '*v*' contains the fluxes of the two reactions. The vector '*b*' which results as a product of *S* matrix and *v* would contain the fluxes of

each of the metabolite's uptake rates. This assumption does not give any unique solution of flux rates of the extreme reactions. This method usually gives a solution space containing all possible solutions according to the biochemical reactions. This solution space can be reduced by giving the flux rates calculated from the experimental processes. Mostly, this solution space is solved by linear optimization method where we maximize one particular optimal reaction and find the possible flux rates of all the reactions in order to maximize that particular optimal reaction. FBA has been utilized in the field of computational simulation of biochemical processes in this manner.

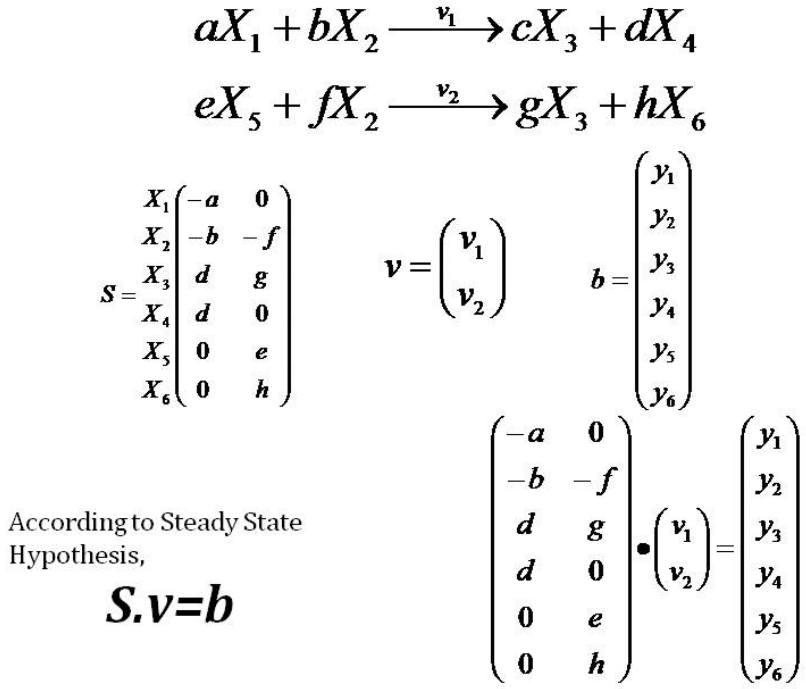


Figure 6: Explaining stoichiometric matrix and steady-state inside a simple system containing two biochemical reactions.

It is to be noted that FBA is similar to the method that was explained in the matrix formalism section [Section 2.1]. The *S* matrix along with the extreme pathway reactions can be compared with that of the R* matrix in the formalism method. The *v* vector can be compared with that of the *f* vector. A similar approach as FBA can be implemented to that of the study of the regulatory model for the expression of genes. It was also discussed that the vector which contains the expression state of genes is also not unique. So, a similar linear optimization utilized in FBA can be used to each gene in the regulatory model for a given condition. The optimized value above 0 would represent the expression of that particular gene in that condition.

The first step to adapt the methods used in FBA to the regulatory model is to correct the bounds of reactions representing environmental conditions in the model. As described previously [Section 3.1], there were five different conditions where the TF mutants are analyzed in this project. These five conditions are made into five different models, each representing one of the conditions. Then

the TF mutants are studied in each of these conditions. The minimal media is represented by giving glucose as a carbon source, where the bounds of 'GluxtX_P' representing the presence of glucose in the environment is given as 1 and the converse [absence] 'GluxtX_A' is given as 0. This 1 and 0 represents the Boolean rules for that particular reaction [intake of glucose from environment] to be active or inactive respectively. For minimal media, the presence of substances like ammonia [nitrogen source], phosphate and oxygen are also represented by denoting 1 as upper bound to their respective up-take reactions.

The carbon limited condition can be represented in the same way as minimal media where all the other input reactions have same bounds except 'GluxtX_P' and 'GluxtX_A'. In this case the converse of values used in the previous case is adapted i.e. 'GluxtX_P' has bound values as 0 and 'GluxtX_A' has bound values 1. The nitrogen limited condition is also represented as same as minimal media with the converse bound values of ammonia uptake has been utilized. In case of galactose condition, the absence of glucose is represented in the same way as in carbon limited condition along with correcting the bound values for the presence of galactose. The fifth condition which is the minimal media in the presence of sorbitol is similar to minimal media along with the corrected bound values for the sorbitol uptake reaction.

As the models, each representing a particular condition is defined the expression state of each gene which is represented as an exchange reaction [output] in this model can be calculated by applying constraint based analysis optimizing for a particular gene. The regulatory model also contains the information of transcription factors and their regulatory reactions in the model. As explained in the previous section [Section 2.2], the optimal function can be represented by given 1 to a particular extreme reaction for which the optimization process would be carried out. With the help of the COBRA 'objective function' command used to represent the objective reaction starting from the first gene [extreme reaction]. After denoting the objective reaction, 'solveLP' command used for optimization of the solution space is used to find the possible expression state of that particular gene in that particular condition.

A function was created where the objective function changes starting from the first gene expression reaction [extreme] to the next until the last gene in the model and the optimization is achieved simultaneously. As a result, this function would give a vector of possible expression states of all the genes in that particular condition. This particular vector achieved in the first step with the first model [minimal media for instance] would represent the possible expression states of all the genes of wildtype *Saccharomyces cerevisiae* in minimal media. The next step would be to engineer TF-KO strains that can be analyzed in the same condition after the wild types.

The TF-KO strains can be engineered with the help of TF information available from the regulatory model. The regulatory model used in this study contained a list of transcription factors that was involved in this regulatory model. It can be inferred that these TFs will also be a part of metabolites in this model. The S matrix which is automatically generated by the COBRA toolbox as soon as the model is imported to the Matlab can be utilized to create the TF-KO strains. The names from the list are taken one by one to match with the metabolites in the model. For every TF metabolite, there will be a bunch of reactions [columns] with non-zero positive stoichiometric coefficients representing the reactions that produces or activates the TF.

These reaction indices are retrieved and the upper bound of all these reactions is set to zero. This represents the particular TF-KO strain. The upper bounds of all the reactions should be set back to their original [wildtype] values before changing for the next TF. This process is carried out for all the transcription factors on a step by step basis available in the regulatory model. Then the 'simulation' function giving out vectors of expression state of all the genes is applied to each of the TF-KO strain. The similar procedure is also carried out for the other conditions to achieve expression states of all the genes in all the strains with respect to each of the condition.

3.3 Singular Value Decomposition and PCA

As a result of the application of the analysis process described previously to the available models, the matrix containing all the expression states of the genes in the model for every TF-KO strain in each condition is obtained. The first and the primary step after obtaining these matrices would be to find the number of expressions of differentially expressed genes between mutants and wildtypes. To achieve this, the expression state of each gene in every TF-KO strain is compared with the corresponding gene state of wildtype in that particular condition. The list of genes that changed expression in case of mutants to that of the wildtype was retrieved. The number of genes that changed expression was accounted as a measure of significance.

This method would help only in analyzing the difference between the wildtype and a mutant for a particular growth condition. Some other analytical method has to be implemented to study the effects of transcription factors and environmental conditions considering the results from all of the environmental conditions together. This is because of the reason that the variables to be taken into consideration are very large. Around 60 transcription factors who's KOs are studied for the expression states of around 350 genes in five different conditions in this project. Multi-variant analysis could help in studying these results for a better understanding of the behavior of *Saccharomyces cerevisiae* in different conditions.

The results from all of these experiments can be considered similar to that of a microarray analysis experiment where every row represents the expression state of a gene and every column represents the assay. Singular Value Decomposition [SVD] and Principal Component Analysis [PCA] has served as two major methods in dealing with the microarray analysis treating them as a dataset with extra-ordinarily large number of variables. These methods help in reducing the number of variables in the dataset to a significant level but sustaining most of the variance in the original dataset within the newly defined variables. This is the main object of any multi-variant analysis methods while SVD/PCA have been proved and applied successfully in the field of microarray analysis for a very long period (Raychaudhuri et al. 2000).

$$X = USV^T$$

SVD is defined as a method for representing a data matrix in the factorization form as shown above. Consider 'X' as a data matrix with expression states in a microarray analysis which is of the form m x n, with m number of genes and n number of assays. Application of SVD gives three other matrices which are the factorized forms of the data matrix 'X' with specific properties to each of them. The matrix U which is of the form m x n contains columns which are expressed as the left singular vectors which form the orthonormal basis for the assay based analysis. The rows of matrix VT or

the columns of the matrix V, which is an n x n square matrix, are expressed as the right singular vectors which forms orthonormal basis for the gene based analysis. The matrix S is a diagonal matrix with non-zero elements representing the singular values sorted from high to low order and the singular vectors are also ordered according to their singular values.

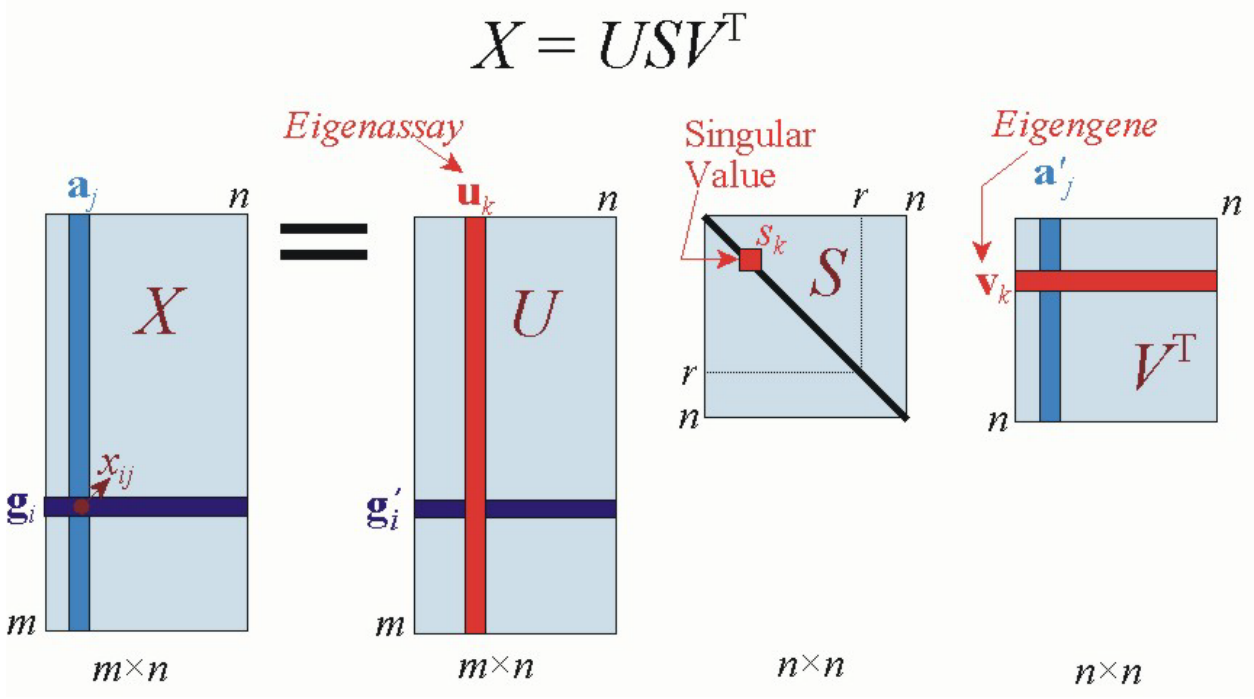


Figure 7: Application of SVD to Microarray Analysis Data (Wall et al. 2003).

In the perspective of a microarray data analysis which is similar to that of our study which was mentioned earlier, the n columns usually represent assays and the m rows represent genes. As explained earlier the right and left singular vectors span the gene transcriptional responses [Eigengene] and assay expression profiles [Eigenassay]. As it is the same case with the aspect of systems biology, the primary interest would be to analyze the gene transcriptional responses in the experiment.

The singular values are proportional relative variances that correspond to the singular vectors. Mostly the first and second singular vectors carry most of the variances in the whole data, which is similar to PCA. PCA is a method to visualize the data which have higher number of variables in lower number of variables known as components. The singular vectors from the SVD analysis can be treated as components. As, both components in case of PCA and singular vectors in case of SVD are orthogonal to themselves. It can be stated as SVD being a method to achieve PCA for visualization of the data (Wall et al. 2001).

So, it can be understood that the first two singular vectors which carry the greatest variance from the whole data could be sufficient to visualize any biological significance of the whole data. This can be done by means projection of the components in a scatter plot. As this study focuses on the

constraint based analysis of the GSMs [genome scale models] of *Saccharomyces cerevisiae*, the first two components could most probably be the conditions and genes in our models. This is because of the reason that these two components could hold the most variance of our data.

3.4 Integration to Metabolic Model

It is necessary to study the interplay between the processes of regulation and metabolism to completely understand the effects of transcription factors over the growth and maintenance of the cells. There are several methods to study the impact of regulation over metabolism through computational analysis of genome scale models (Shlomi et al. 2007; Chandrasekaran and Price 2010). The rFBA [regulatory flux balance analysis] is a method where the simulation is carried out through the choice of dynamic flux profiles in different conditions. The main limitation of this method is the absence of an automated algorithm to relate between the regulator and the target.

It is considered that the transcriptional regulation can be carried out separately and then it can be integrated through the genes of the metabolic model which would be most suitable for this project with the implementation of Boolean formalism in the regulatory model.. This process is carried out for every assay [TF mutants] in every condition by looking for genes which are not expressed in each assay from the regulatory model. Then the upper bound values for the reactions in which those genes are involved are reduced to 10% of its original bound value in the respective metabolic model. By this elimination of some essential genes from the metabolic system is restricted. This process has been carried out for two of the most important conditions which are the minimal media with glucose and galactose as their carbon source.

For these conditions, the Boolean logic is applied to find the expression of each gene in each assay from each condition and they are retrieved. For example, if the expression state of a gene has a value greater than zero, then that gene is considered to be expressed in that assay. The list of genes that not expressed from each assay in each condition is retrieved. Their corresponding gene names in the metabolic model are matched with the list retrieved from the regulatory model. This was done, because the format of gene names used in the regulatory model and that of it in the metabolic model different. Then the corresponding gene index numbers and their related reactions in the imported metabolic model are retrieved. This can be achieved with the help of genes list and the gene-reaction matrix that are formed automatically by importing the SBML model. As, these information were retrieved, the upper bound values of those reactions are reduced to 10% to infer the impact of regulation in the metabolism.

4 Results

The simulations in the regulatory model of *Saccharomyces cerevisiae* are studied on the basis of comparative analysis between wildtypes and the TF-KO mutants in each condition. It has been noticed in the regulatory study that even for wildtypes in normal conditions such as minimal media with the presence of glucose as the carbon source, many genes don't seem to be expressed. In the above mentioned case, around 200 genes out of the 349 available genes in the model did not show any expression. This analysis seemed to be coherent with that of the published results from a similar experimental study that was mentioned in the above section [Section 2.3]. The primary aim with the results from the regulatory model is to study the number of genes that showed differential expression in the TF-KO assays from that of the wildtypes in each condition.

The expression state of each gene in every TF-KO is compared with that of the same in case of wildtype. The genes for which the expression states are same are eliminated and the genes for which it is different are retrieved. This number of retrieved genes is believed to be the controlled by the transcription factor in the process of regulating the metabolism. Similarly, this process has been carried out for all the assays [mutants] in every condition. Bar charts have been developed to analyze the different effects of different transcription factors in each condition. Later, the expression states of all the genes in each assay from all the conditions are concatenated together and SVD is applied. The application of SVD is analyzed through components from the singular vectors by PCA.

In case of the carbon limited condition, as there is no glucose they showed similarities towards the galactose condition in the published results based on experimental analysis (Fendt et al. 2010). The transcription factors like rgt1, leu3, rtg1, rtg3 and rox1 showed significant difference from the transcription factors from that of the wildtypes in the above mentioned research work. Also, it is interesting that the transcription factor bas1 which is stated as a novel transcription factor seemed to be involved in the process of regulation is also involved in this condition but which changes the expression of only one gene [Figure 8].

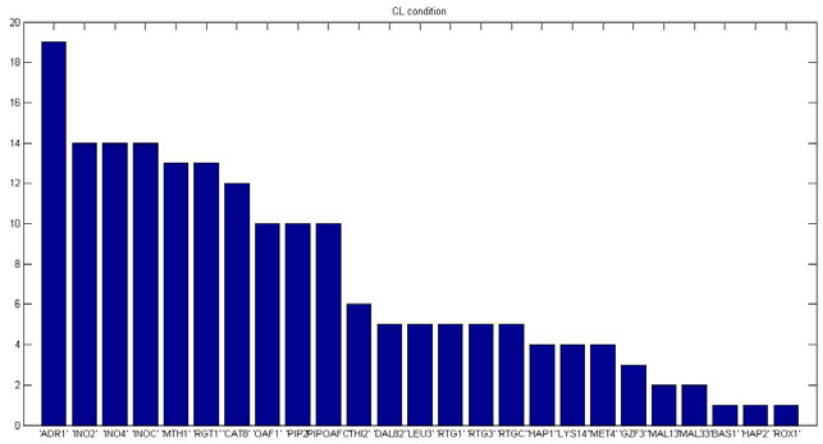


Figure 8: TF-KO analysis in Carbon Limited condition.

In case of Nitrogen limited condition, the transcription factor gcn4 seemed to have shown larger number of different expression genes from that of the wildtype. This transcription factor was reported to have played major role in the regulation process in most of their conditions (Fendt et al. 2010). This result seemed to be similar to that of the urea condition that was mentioned in the published result. It can also be noted that the transcription factor gcr2 has also shown some significant changes in case this condition. This is shown in the bar chart that explains the activity of each transcription factors in the nitrogen limited condition [Figure 9].

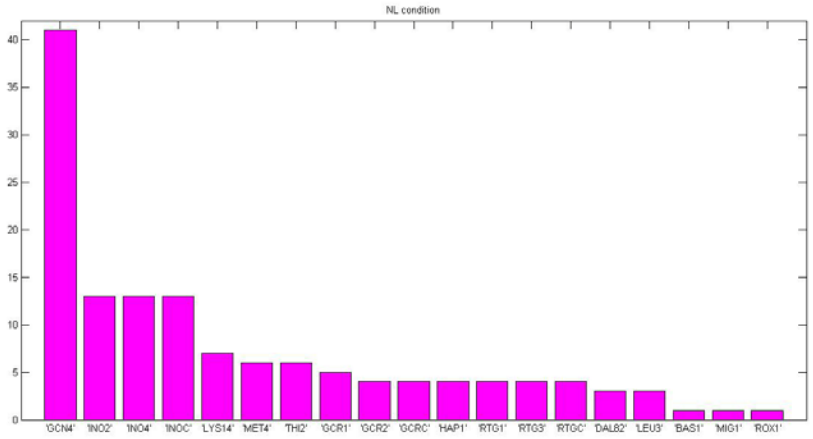


Figure 9: Transcriptional regulatory analysis in Nitrogen limited condition.

In case of the condition where Sorbitol is present in glucose minimal media, the simulation seemed to be very similar to that of the glucose minimal media condition. As the presence of sorbitol was only the difference between these two conditions, *Saccharomyces cerevisiae* seemed to behave very similar to each other. This was also the case with the experimental study of these two conditions (Fendt et al. 2010). The analysis of this condition can be explained from that of the bar chart shown below .

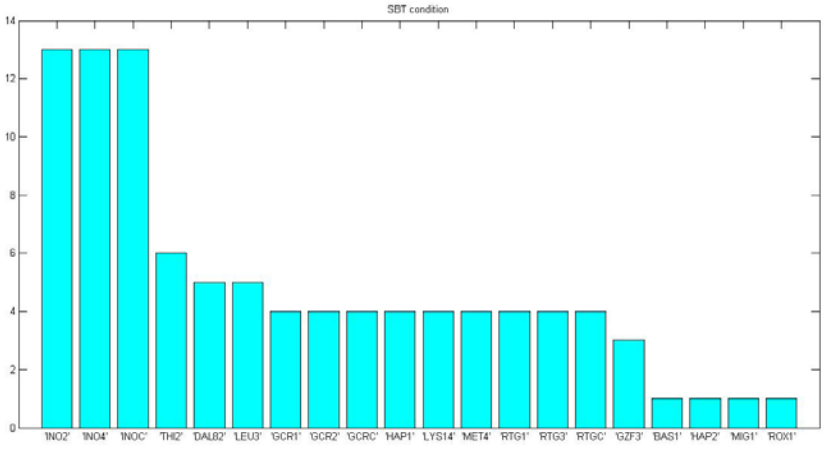


Figure 10: Transcriptional regulatory analysis in the presence of Sorbitol in minimal media.

In case of the presence of glucose, the simulation results were exactly the same as it was in the case of sorbitol. Even in the experimental results it can be inferred that the amount of respiration was also found to be very similar for each assay in these two conditions [Figure 5]. It seemed as the presence of sorbitol did not affect the transcriptional factors in any way different from that of the other condition.

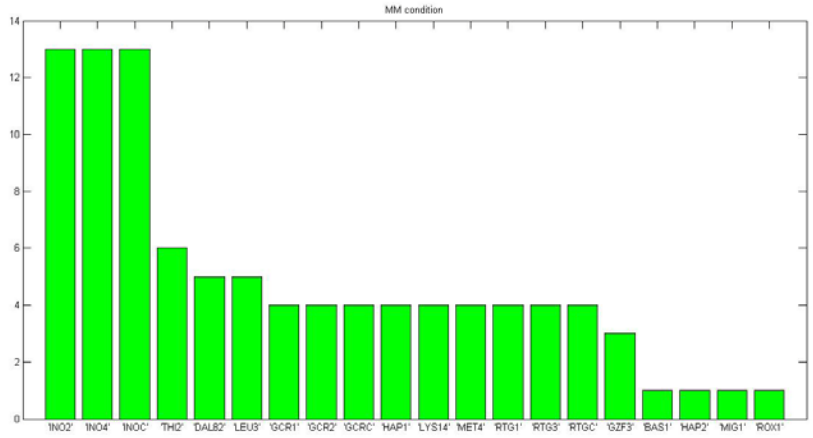


Figure 11: Transcriptional regulatory analysis in the minimal media with glucose as the carbon source.

The condition where galactose was the carbon source in a minimal media has given the most interesting simulation results among the others. With this case, the simulation results managed to cover almost all the transcription factors that were reported in the experimental study. Transcription factors like gcr2, leu3, rtg1, rtg3, rgt1 and rox1 were seen to be showed some considerable changes in the gene expression from that of the wildtype. Also, this condition seem to capture many number of transcription factors than any other condition that showed considerable difference from that of the wildtypes .

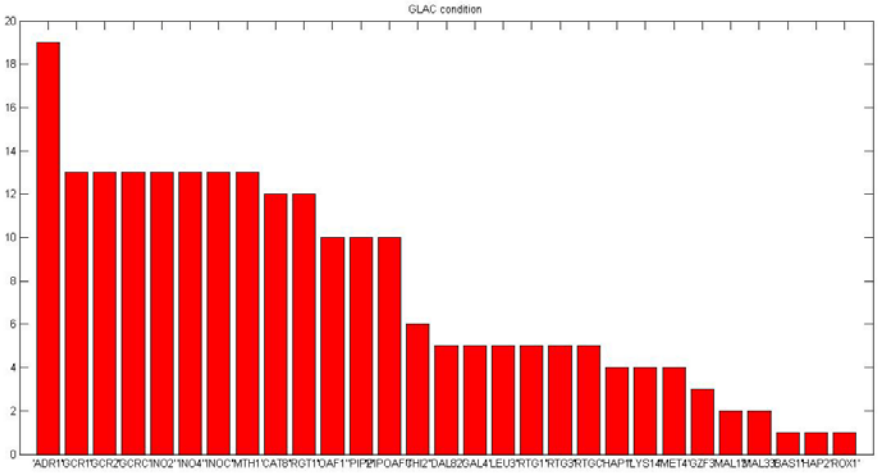


Figure 12: Transcriptional regulatory analysis where galactose was the carbon source in minimal media.

The analysis of transcriptional regulation was seen in each condition separately as above, but the SVD analysis has been applied to the whole gene expression data obtained from all the conditions together. This analysis helped in studying these assays by means of reducing the number of variables [genes, assays and conditions] in terms of components. As the decomposition of the entire expression state matrix was applied through diagonalization of its co-variance matrix. The method was explained in detail [Section 3.3]. Those right singular vectors can be considered as components in PCA. The relative variance of each component was calculated from their singular values, as they are proportional to each other and plotted for the visualization of the data[Figure 13] .

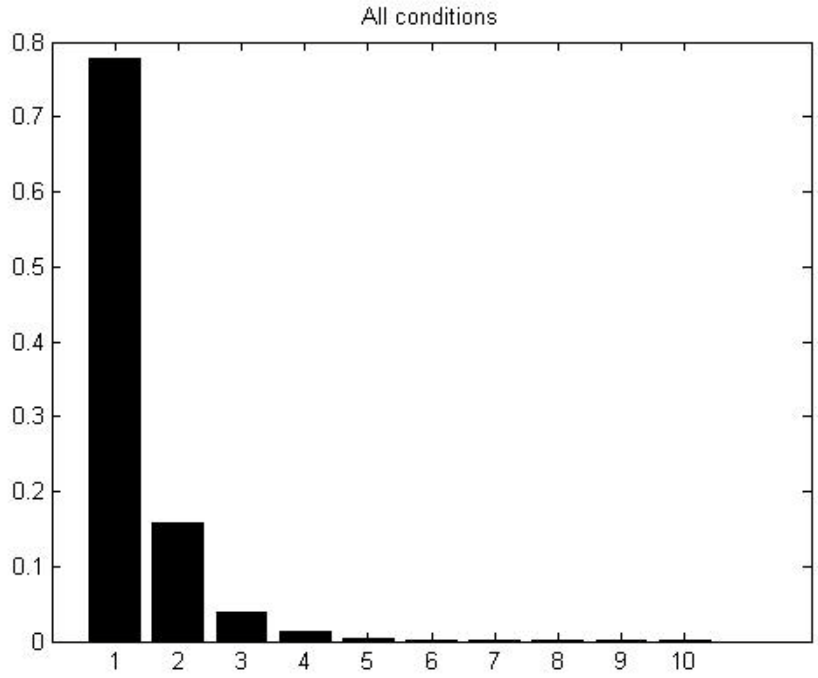


Figure 13: Principal components and their relative variance.

From the above figure [Figure 13], it can be inferred that almost 90% of the variance in the data has been achieved by the first two components or singular vectors. As, these singular vectors are the right singular or gene transcriptional responses represented in each of the assays, visualization of the data with the first two principal components could be enough to view the biological significance of the whole data.

These first two principal components or right singular vectors are plotted across each other similar to that of a scatter plot. The mutants or assays that show a significant difference from that of its wildtype in each condition is labeled. Also, values from each condition are plotted with different colors so as to see the difference and similarities of the behavior of *Saccharomyces cerevisiae* in different conditions. The wildtypes from each condition is plotted with different sign from that of the mutants and variation is shown by connecting the wildtypes to the mutants with a line [Figure 14].

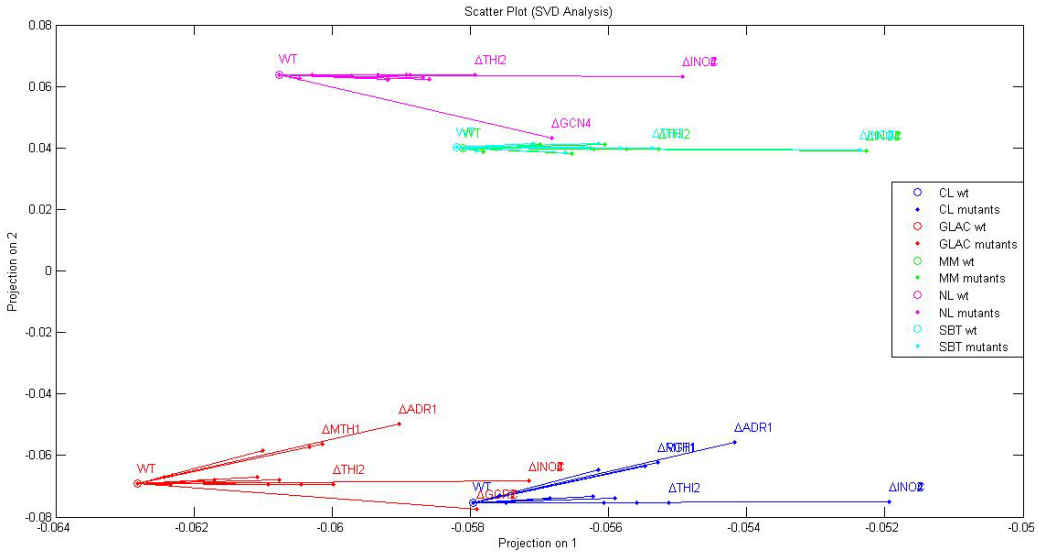


Figure 14: PCA of the whole expression data. Principal component 1 over principal component 2.

It can be seen from above [Figure 14], the projection of principal components have shown that the conditions of Minimal medium with glucose in presence and absence of sorbitol has overlapped one over the other in a similar fashion. It was also clear that the component 1 has acquired the variance between the assays and component 2 has acquired the variance from the conditions. It can be seen from the CL and GLAC conditions where they both lack the presence of glucose are grouped below the origin and all the other conditions where there is presence of glucose are grouped above the origin. The transcription factors which show significant difference were also plotted in this scatter plot. The genes that were modulated by the transcription factors in each condition were given in detail in the appendix section.

These regulatory gene-expression profiles of each assay in each condition were integrated with that of the respective metabolic model. The integration of regulatory predictions with that of the metabolic model is to analyze the effects of genetic perturbations [TF-KO mutants] over the fluxes of metabolic reactions. The analysis of these mutants on a flux level would be helpful in understanding the relation between regulation and metabolism. It has also been proved earlier that the introduction of regulatory rules in the metabolic model has increased its ability to predict the process of metabolism more accurately.

It has been understood from the results of regulatory predictions that most of the genes are not expressed in most of the conditions/assays. It has been shown that even for the wildtype in a normal minimal media condition; around 200 genes were predicted to be not expressed. Through the integration process the metabolic model would not be able to predict a proper solution if all the reactions related to these genes was deleted from the model. Hence a strategy of really low expression i.e. about only 10% of the flux were allowed for the all the reactions related to the non-expressed genes in the corresponding metabolic model for each of those assays. These reactions can be retrieved with the help of gene-reaction matrix from COBRA toolbox.

The ability of the metabolic model to simulate the crab-tree effect was checked for all the wild-types in each of those different environmental conditions. The ability to simulate crab-tree effect which is to witness the excretion of ethanol by the wild-type was unsuccessful for most of the conditions. The regulatory rules applied to the corresponding metabolic model of wildtype in the minimal media with galactose condition were able to simulate the crab-tree effect. The further analysis of each of those TF-KO mutants were also analyzed for this condition and looked for the amount of ethanol excretion when compared to that of the wild-type.

The flux of ethanol excretion reaction for each of the TF-KO mutants was calculated by the application of corresponding regulatory predictions from the regulatory model in this particular galactose minimal media condition. The amount of ethanol excretion by the TF-KO mutants is plotted against the same in wild-type [Figure 15]. The mutants that predict lower amounts of ethanol excretion are considered to be involved in changing the cellular behavior towards fermentation. Alternatively, the mutants that predict more ethanol prediction are considered to be involved in controlling the cellular behavior towards respiration.

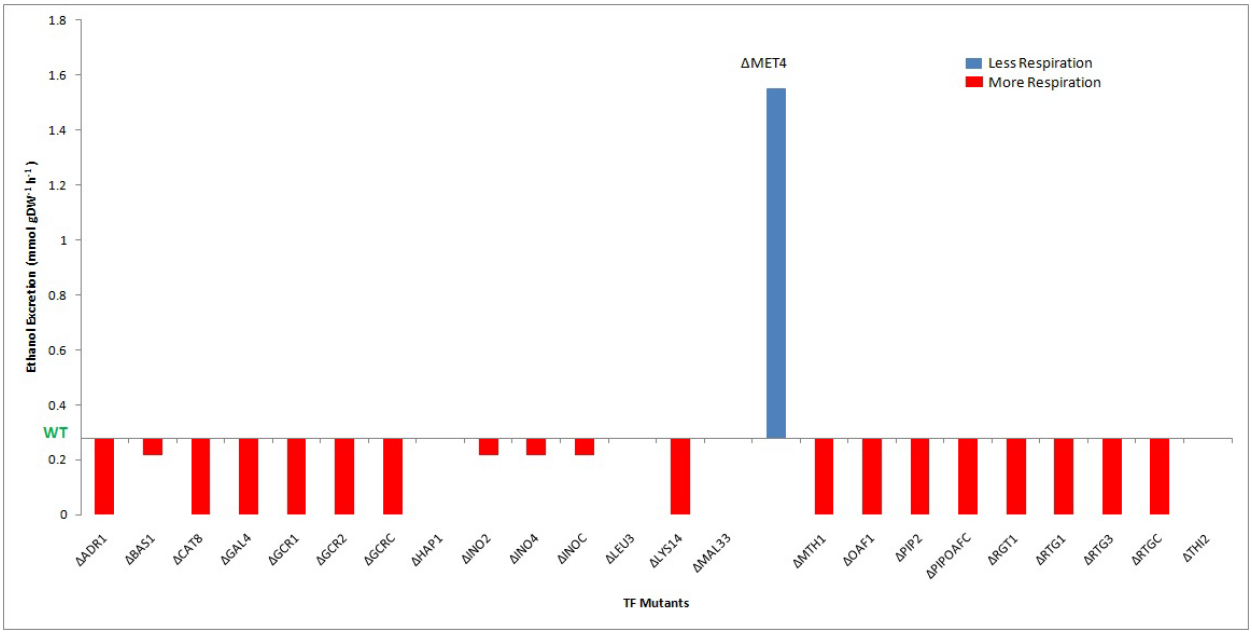


Figure 15: Integration of regulatory and metabolic model for GLAC condition and their corresponding fluxes of ethanol excretion in their respective TF-KO mutants.

It can be inferred from the above plot [Figure 15] that ΔMET4 mutant predicts very high excretion of ethanol compared to the wildtype in galactose condition. This shows that this transcription factor involves in controlling the behavior of the cell towards respiration. Also, it can be seen that all the other transcription factors control the behavior of the cell towards fermentation.

5 Discussion

Based on the comparison between the experimental analysis and the computational simulations, the transcription factor BAS1 which was found in all the conditions with significant changes in case of experimental analysis was also found in the computational simulation but with only modulating one gene in each condition. The transcription factor GCR2 was also found in all of the conditions of the computational simulation but with comparatively making less number of significant changes to that of the wildtype in each condition. The transcription factor GCN4 which seemed to have affected in all three conditions in the experimental results which are similar to the conditions in the computational simulation was found only in the NL condition. The transcription factor PHO2 [Figure 5] was not found to effect any of the conditions in the analyzed computational simulations.

This might be due to the reason that the regulatory model used in the computational analysis could be incomplete. It can also be seen that some of the transcription factors reported [Figure 5] to have made some significant impact in one or more of the studied conditions like RPN4, HAP3, HAP5, MAC1, MDS3, MSN2 and STB5 were not at all present in the studied model. This can also be inferred mainly with the number of transcription factor mutants analyzed in the experimental study which is 119 TF-KO mutants to only 57 transcription factors available in the studied model.

The integration of the regulatory model to the metabolic model was studied mainly for the minimal media with galactose [GLAC] condition. This is because of the reason that on a gene expression level comparison with that of the experimental results, this condition showed more consistency with the experimental results than other conditions. Also, this was the only condition that emulated crab-tree effect for the wildtype by the application of corresponding regulatory predictions. As mentioned in the results section the plot [Figure 15] represents the involvement of transcription factors in this condition in controlling *Saccharomyces cerevisiae* between respiration and fermentation.

By comparing these simulations of TF-KO mutants in the metabolic model to the experimental results, it can be seen that a significant similarity was not observed. According to the computational simulation only ΔMET4 strain seemed to behave towards higher fermentation whereas this mutant did not seem to affect the behavior of the cell in any of the studied conditions. Also, mutants like ΔGCR2 and ΔRGT1 were the only mutants that showed similar predictions to that of the experimental results. The mutants such as ΔRTG1, ΔRTG3 and ΔLEU3 showed predictions of more respiration whereas they were reported to be involved in shifting towards more fermentation.

In general, the results obtained from the integration of regulatory and metabolic models stresses on the fact that the regulatory model used in this project was incomplete. This paves a way for a more intense study on the transcriptional regulation and a more detailed reconstruction of the regulatory genome-sale model is absolutely necessary. A regulatory model containing interactions between more numbers of transcription factors and genes could provide a better simulation of real-time experiment. This research work could be considered as a starting point towards the development of a new broader and more comprehensive regulatory model. This could further help in processes like random sampling to infer any new biologically significant information regarding the regulation of metabolism in *Saccharomyces cerevisiae.*

It would also be interesting to study the other transcription factors like 'INO' family transcription factors [INO2, INO4, and INOC] and THI2 which was predicted to have some significant roles in the process of regulation. The mechanisms behind ΔMET4 mutant becoming more fermentative in case of galactose condition could be studied more intensively. This might provide new information about the behavior of *Saccharomyces cerevisiae* which can be used in modulating this organism towards specific perspectives in industrial or medicinal applications.

It is also a considerable fact that the gene expression profiles obtained from the Boolean regulatory model can only issue ON/OFF state of a gene, where in reality the increase or decrease in expression of a given gene could prove to be more significant. This is one of the main disadvantages in analyzing a Boolean regulatory model apart from its ability to increase the efficiency of computational prediction of metabolic processes. The usage of analysis through new technologies like PROM (Chandrasekaran and Price 2010) could possibly give better results. PROM calculates the probability of a gene getting expressed into protein which could be related to the amount of expression of that particular gene in that particular condition. This would give a better integration of regulation and metabolism compared to the normal Boolean integration.

It has also been proposed that through mixed-integer linear programs converted from Boolean and other formalisms, the biological processes inside a cell are described in more accurate manner (Jensen et al. 2011). With all these proposed technologies, the integration of regulation and metabolism could provide a better understanding of the behavior of the cells. These technologies could also help in making new hypothesis and more accurate simulations that could pave wave for an improvement in the efficiency of the systems biology field.

Appendix A: Linear Optimization

Assume that the linear programming deals with a data containing '*n*' variables and *'m'* linearly independent equality constraints and we use this method in a matrix-vector form for the following (Nash and Sofer 1996),

$$\text{Maximize } z = c^T x$$

Subject to the constraints $Ax=b;\ x \geq 0$

Here, *'x'* is a basic feasible solution; 'z' is the objective function; *'A'* is a matrix with the stoichiometric values along with the constraints; *'c^T'* is a co-efficient matrix of '*x'.*

Let *'x'* be a feasible solution vector with variables are ordered so that

$$x = \begin{pmatrix} x_B \\ x_N \end{pmatrix}$$

Where, 'x_B' and 'x_N' are the basic and the non-basic variables respectively. Now, the objective function can be written as

$$z = c_B^T x_B + c_N^T x_N$$

Here c_B^T and c_N^T are the coefficients of the basic and the non-basic variables respectively. Similarly the constraints can be written as

$$Bx_B + Nx_N = b \text{ and}$$

$$x_B = B^{-1}b - B^{-1}Nx_N$$

By changing different values for the non-basic variables, it is possible to obtain all reasonable solutions. When x_B is substituted in the objective function

$$z = c_B^T B^{-1}b + [c_N^T - c_B^T B^{-1}N]x_N$$

By setting the non-basic variable $x_N = 0$, one of the possible solution for this problem can be obtained.

$$z = c_B^T B^{-1}b$$

When there are some exceptions in the non-basic variable x_B which is not zero, then x_B can be written as

$$x_B = \hat{b} - \hat{A}_t x_t$$

Where the vector $\hat{b} = B^{-1}b$,

The vector $\hat{A}_t = B^{-1}A_t$

And A_t is the t^{th} column of matrix *A.*

The examination of the x_B equation was done for each component

$$(x_B)_i = \hat{b}_i - \hat{a}_{i,t} x_t$$

$$\bar{x}_t = \min_{1 \le i \le m} \left\{ \frac{\hat{b}_i}{\hat{a}_{i,t}} : \hat{a}_{i,t} > 0 \right\}$$

This minimum value from the above equation identifies the new non-basic variable and finds the solution with x_t as the new basic variable.

Thus the above objective function becomes

$$x_B \leftarrow x_B - \hat{A}_t x_t \; and \; \hat{z} \leftarrow \hat{z} + \hat{c}_t x_t$$

Appendix B: SVD and PCA

Application of SVD to a data matrix S gives,

$$X^{(l)} = \sum_{k=1}^{l} u_k s_k v_k^T$$

Where, l is the rank of the matrix X. The above expression is considered to be one of the main results from the SVD analysis. The matrix $X^{(l)}$ takes the minimized sum of squares of difference between each of its elements from *X (Wall et al. 2003).* This can be represented as,

$$\sum_{ij} \left| x_{ij} - x_{ij}^{(l)} \right|^2$$

The matrices V^T *and* S can be calculated from diagonalizing X^TX,

$$X^T X = VS^2V^T$$

U can be obtained as follows,

$$U = XVS^{-1}$$

The gene transcriptional response in a micro-array analysis $\mathbf{g_i}$, can be obtained with the help of the right singular matrix V. This can be expressed as,

$$\boldsymbol{g_i} = \sum_{k=1}^{\boldsymbol{l}} u_{ik}\, s_k \boldsymbol{v_k}$$

Here *i= 1…m*, 'm' represents the number of genes. Similarly to analyze the assay expression profile $\mathbf{a_j}$, it can be achieved from the left singular matrix U and it can be expressed as,

$$\boldsymbol{a_j} = v_{jk}\, s_k \boldsymbol{u_k}$$

Here *j=1…n,* 'n' represents the number of assays analyzed in the micro-array experiment.

Appendix C: Simulation Results.

i) **CL Condition:** The genes that showed differential expression from that of the wildtype for the transcription factor mutants in the carbon limited condition are given in the table below.

TF	ADR1'	BAS1'	CAT8'	DAL82'	GZF3'	HAP1'	HAP2'	INO2'	INO4'	INOC'	LEU3'	LYS14'
No. of genes	19	1	12	5	3	4	1	14	14	14	5	4
Genes	ACS1'	SHM2'	ACS1'	CAR2'	DAL2'	CTT1'	ASN1'	ACC1'	ACC1'	ACC1'	GDH2'	LYS12'
	ADH2'		ADH2'	DAL2'	DAL4'	CYT1'		ACS2'	ACS2'	ACS2'	ILV2'	LYS20'
	ALD4'		ALD6'	DAL4'	DAL7'	DLD1'		CKI1'	CKI1'	CKI1'	ILV5'	LYS21'
	ANT1'		FBP1'	DAL7'		QCR2'		CHO1'	CHO1'	CHO1'	LEU1'	LYS4'
	CIT3'		ICL1'	DUR1,2'				CHO2'	CHO2'	CHO2'	LEU4'	
	CRC1'		IDP2'					DPP1'	DPP1'	DPP1'		
	CTA1'		JEN1'					FAS1'	FAS1'	FAS1'		
	CYB2'		MDH2'					FAS2'	FAS2'	FAS2'		
	FAA2'		MLS1'					GUT1'	GUT1'	GUT1'		
	FDH2'		PCK1'					HNM1'	HNM1'	HNM1'		
	FOX2'		SFC1'					INO1'	INO1'	INO1'		
	GUT1'		YAT1'					ITR1'	ITR1'	ITR1'		
	ICL2'							OPI3'	OPI3'	OPI3'		
	IDP3'							PSD1'	PSD1'	PSD1'		
	MDH3'											
	POT1'											
	POX1'											
	PXA1'											
	PXA2'											

TF	MAL13'	MAL33'	MET4'	MTH1'	OAF1'	PIP2'	PIPOAFC'	RGT1'	ROX1'	RTG1'	RTG3'	RTGC'	THI2'
No. of genes	2	2	4	13	10	10	10	13	1	5	5	5	6
Genes	MAL11'	MAL31'	MET17'	ACS1'	ANT1'	ANT1'	ANT1'	ACS1'	COX5B'	ACO1'	ACO1'	ACO1'	PHO3'
	MAL12'	MAL32'	MET2'	ADH2'	CRC1'	CRC1'	CRC1'	ADH2'		CIT2'	CIT2'	CIT2'	THI20'
			MET3'	ALD6'	FAA2'	FAA2'	FAA2'	ALD6'		IDH1'	IDH1'	IDH1'	THI21'
			SAM1'	FBP1'	FOX2'	FOX2'	FOX2'	FBP1'		IDH2'	IDH2'	IDH2'	THI22'
				ICL1'	IDP3'	IDP3'	IDP3'	ICL1'		PYC1'	PYC1'	PYC1'	THI6'
				IDP2'	MDH3'	MDH3'	MDH3'	IDP2'					THI80'
				JEN1'	POT1'	POT1'	POT1'	JEN1'					
				MDH2'	POX1'	POX1'	POX1'	MDH2'					
				MLS1'	PXA1'	PXA1'	PXA1'	MLS1'					
				PCK1'	PXA2'	PXA2'	PXA2'	PCK1'					
				SFC1'				SFC1'					
				SUC2'				SUC2'					
				YAT1'				YAT1'					

ii) **GLAC condition:** The genes that show differential expression from that of the wildtype for the transcription factor mutants for the condition with Galactose as the carbon source are given in the table below.

TF	ADR1'	BAS1'	CAT8'	DAL82'	GAL4'	GCR1'	GCR2'	GCRC'	GZF3'	HAP1'	HAP2'	INO2'	INO4'	INOC'
No. of genes	19	1	12	5	5	13	13	13	3	4	1	13	13	13
Genes	ACS1'	SHM2'	ACS1'	CAR2'	FUR4'	ADH1'	ADH1'	ADH1'	DAL2'	CTT1'	ASN1'	ACS2'	ACS2'	ACS2'
	ADH2'		ADH2'	DAL2'	GAL1'	CDC19'	CDC19'	CDC19'	DAL4'	CYT1'		CKI1'	CKI1'	CKI1'
	ALD4'		ALD6'	DAL4'	GAL10'	ENO1'	ENO1'	ENO1'	DAL7'	DLD1'		CHO1'	CHO1'	CHO1'
	ANT1'		FBP1'	DAL7'	GAL2'	ENO2'	ENO2'	ENO2'		QCR2'		CHO2'	CHO2'	CHO2'
	CIT3'		ICL1'	DUR1,2'	GAL7'	FBA1'	FBA1'	FBA1'				DPP1'	DPP1'	DPP1'
	CRC1'		IDP2'			GPM1'	GPM1'	GPM1'				FAS1'	FAS1'	FAS1'
	CTA1'		JEN1'			PGI1'	PGI1'	PGI1'				FAS2'	FAS2'	FAS2'
	CYB2'		MDH2'			PGK1'	PGK1'	PGK1'				GUT1'	GUT1'	GUT1'
	FAA2'		MLS1'			PGM1'	PGM1'	PGM1'				HNM1'	HNM1'	HNM1'
	FDH2'		PCK1'			TDH1'	TDH1'	TDH1'				INO1'	INO1'	INO1'
	FOX2'		SFC1'			TDH2'	TDH2'	TDH2'				ITR1'	ITR1'	ITR1'
	GUT1'		YAT1'			TDH3'	TDH3'	TDH3'				OPI3'	OPI3'	OPI3'
	ICL2'					TPI1'	TPI1'	TPI1'				PSD1'	PSD1'	PSD1'
	IDP3'													
	MDH3'													
	POT1'													
	POX1'													
	PXA1'													
	PXA2'													

TF	LEU3'	LYS14'	MAL13'	MAL33'	MET4'	MTH1'	OAF1'	PIP2'	PIPOAFC'	RGT1'	ROX1'	RTG1'	RTG3'	RTGC'	THI2'
No. of genes	5	4	2	2	4	13	10	10	10	12	1	5	5	5	6
Genes	GDH2'	LYS12'	MAL11'	MAL31'	MET17'	ACS1'	ANT1'	ANT1'	ANT1'	ACS1'	COX5B'	ACO1'	ACO1'	ACO1'	PHO3'
	ILV2'	LYS20'	MAL12'	MAL32'	MET2'	ADH2'	CRC1'	CRC1'	CRC1'	ADH2'		CIT2'	CIT2'	CIT2'	THI20'
	ILV5'	LYS21'			MET3'	ALD6'	FAA2'	FAA2'	FAA2'	ALD6'		IDH1'	IDH1'	IDH1'	THI21'
	LEU1'	LYS4'			SAM1'	FBP1'	FOX2'	FOX2'	FOX2'	FBP1'		IDH2'	IDH2'	IDH2'	THI22'
	LEU4'					ICL1'	IDP3'	IDP3'	IDP3'	ICL1'		PYC1'	PYC1'	PYC1'	THI6'
						IDP2'	MDH3'	MDH3'	MDH3'	IDP2'					THI80'
						JEN1'	POT1'	POT1'	POT1'	JEN1'					
						MDH2'	POX1'	POX1'	POX1'	MDH2'					
						MLS1'	PXA1'	PXA1'	PXA1'	MLS1'					
						PCK1'	PXA2'	PXA2'	PXA2'	PCK1'					
						SFC1'				SFC1'					
						SUC2'				YAT1'					
						YAT1'									

iii) **MM Condition:** The genes that show differential expression from that of the wildtype for the transcription factor mutants in minimal media condition with glucose as the carbon source.

TF	BAS1'	DAL82'	GCR1'	GCR2'	GCRC'	GZF3'	HAP1'	HAP2'	INO2'	INO4'
No. of genes	1	5	4	4	4	3	4	1	13	13
Genes	SHM2'	CAR2'	FBA1'	FBA1'	FBA1'	DAL2'	CTT1'	ASN1'	ACC1'	ACC1'
		DAL2'	GLK1'	GLK1'	GLK1'	DAL4'	CYT1'		ACS2'	ACS2'
		DAL4'	GPM1'	GPM1'	GPM1'	DAL7'	DLD1'		CKI1'	CKI1'
		DAL7'	PGI1'	PGI1'	PGI1'		QCR2'		CHO1'	CHO1'
		DUR1,2'							CHO2'	CHO2'
									DPP1'	DPP1'
									FAS1'	FAS1'
									FAS2'	FAS2'
									HNM1'	HNM1'
									INO1'	INO1'
									ITR1'	ITR1'
									OPI3'	OPI3'
									PSD1'	PSD1'

TF	INOC'	LEU3'	LYS14'	MET4'	MIG1'	ROX1'	RTG1'	RTG3'	RTGC'	THI2'
No. of genes	13	5	4	4	1	1	4	4	4	6
Genes	ACC1'	GDH2'	LYS12'	MET17'	GSC2'	COX5B'	ACO1'	ACO1'	ACO1'	PHO3'
	ACS2'	ILV2'	LYS20'	MET2'			CIT1'	CIT1'	CIT1'	THI20'
	CKI1'	ILV5'	LYS21'	MET3'			IDH2'	IDH2'	IDH2'	THI21'
	CHO1'	LEU1'	LYS4'	SAM1'			PYC1'	PYC1'	PYC1'	THI22'
	CHO2'	LEU4'								THI6'
	DPP1'									THI80'
	FAS1'									
	FAS2'									
	HNM1'									
	INO1'									
	ITR1'									
	OPI3'									
	PSD1'									

iv) SBT Condition: The genes that show differential expression from that of the wildtype for the transcription factor mutants in the minimal media condition with the presence of sorbitol and glucose.

TF	BAS1'	DAL82'	GCR1'	GCR2'	GCRC'	GZF3'	HAP1'	HAP2'	INO2'	INO4'
No. of genes	1	5	4	4	4	3	4	1	13	13
Genes	SHM2'	CAR2'	FBA1'	FBA1'	FBA1'	DAL2'	CTT1'	ASN1'	ACC1'	ACC1'
		DAL2'	GLK1'	GLK1'	GLK1'	DAL4'	CYT1'		ACS2'	ACS2'
		DAL4'	GPM1'	GPM1'	GPM1'	DAL7'	DLD1'		CKI1'	CKI1'
		DAL7'	PGI1'	PGI1'	PGI1'		QCR2'		CHO1'	CHO1'
		DUR1,2'							CHO2'	CHO2'
									DPP1'	DPP1'
									FAS1'	FAS1'
									FAS2'	FAS2'
									HNM1'	HNM1'
									INO1'	INO1'
									ITR1'	ITR1'
									OPI3'	OPI3'
									PSD1'	PSD1'

TF	INOC'	LEU3'	LYS14'	MET4'	MIG1'	ROX1'	RTG1'	RTG3'	RTGC'	THI2'
No. of genes	13	5	4	4	1	1	4	4	4	6
Genes	ACC1'	GDH2'	LYS12'	MET17'	GSC2'	COX5B'	ACO1'	ACO1'	ACO1'	PHO3'
	ACS2'	ILV2'	LYS20'	MET2'			CIT1'	CIT1'	CIT1'	THI20'
	CKI1'	ILV5'	LYS21'	MET3'			IDH2'	IDH2'	IDH2'	THI21'
	CHO1'	LEU1'	LYS4'	SAM1'			PYC1'	PYC1'	PYC1'	THI22'
	CHO2'	LEU4'								THI6'
	DPP1'									THI80'
	FAS1'									
	FAS2'									
	HNM1'									
	INO1'									
	ITR1'									
	OPI3'									
	PSD1'									

v) **NL Condition:** The genes that show differential expression to that of the wildtype for the mutants in nitrogen limited condition with glucose as its carbon source are listed in the table below.

TF	BAS1'	DAL82'	GCN4'	GCR1'	GCR2'	GCRC'	HAP1'	INO2'	INO4'	INOC'
No. of genes	1	3	41	5	4	4	4	13	13	13
Genes	SHM2'	CAR2'	ADE1'	ECM38'	FBA1'	FBA1'	CTT1'	ACC1'	ACC1'	ACC1'
		DUR1,2'	ADE2'	FBA1'	GLK1'	GLK1'	CYT1'	ACS2'	ACS2'	ACS2'
		DUR3'	ADE3'	GLK1'	GPM1'	GPM1'	DLD1'	CKI1'	CKI1'	CKI1'
			ADE4'	GPM1'	PGI1'	PGI1'	QCR2'	CHO1'	CHO1'	CHO1'
			ADE8'	PGI1'				CHO2'	CHO2'	CHO2'
			ARG4'					DPP1'	DPP1'	DPP1'
			ARO1'					FAS1'	FAS1'	FAS1'
			ARO2'					FAS2'	FAS2'	FAS2'
			ARO3'					HNM1'	HNM1'	HNM1'
			ARO4'					INO1'	INO1'	INO1'
			ARO8'					ITR1'	ITR1'	ITR1'
			ASN2'					OPI3'	OPI3'	OPI3'
			BAT1'					PSD1'	PSD1'	PSD1'
			CPA2'							
			DAL5'							
			ECM40'							
			GCV3'							
			GLN1'							
			GLT1'							
			HIS1'							
			HIS2'							
			HIS3'							
			HIS4'							
			HIS5'							
			HIS7'							
			HOM2'							
			HOM3'							
			ILV1'							
			ILV3'							
			KRS1'							
			LYS1'							
			LYS2'							
			LYS9'							
			MET16'							
			MET6'							
			PRO1'							
			THR1'							
			TRP2'							
			TRP3'							
			TRP4'							
			TRP5'							

TF	LEU3'	LYS14'	MET4'	MIG1'	ROX1'	RTG1'	RTG3'	RTGC'	THI2'
No. of genes	3	7	6	1	1	4	4	4	6
Genes	BAT1'	LYS1'	MET16'	GSC2'	COX5B'	ACO1'	ACO1'	ACO1'	PHO3'
	GDH2'	LYS12'	MET17'			CIT1'	CIT1'	CIT1'	THI20'
	LEU1'	LYS2'	MET2'			IDH2'	IDH2'	IDH2'	THI21'
		LYS20'	MET3'			PYC1'	PYC1'	PYC1'	THI22'
		LYS21'	MET6'						THI6'
		LYS4'	SAM1'						THI80'
		LYS9'							

References

Bailey JE. 1998. Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. *Biotechnology progress* **14**(1): 8-20.

Bailey JE, Birnbaum S, Galazzo JL, Khosla C, Shanks JV. 1990. Strategies and challenges in metabolic engineering. *Annals of the New York Academy of Sciences* **589**: 1-15.

Barnett JA. 2003. Beginnings of microbiology and biochemistry: the contribution of yeast research. *Microbiology* **149**(Pt 3): 557-567.

Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, Herrgard MJ. 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nature protocols* **2**(3): 727-738.

Bonneau R. 2008. Learning biological networks: from modules to dynamics. *Nature chemical biology* **4**(11): 658-664.

C.Guthrie, Fink GR. 1991. Guide to Yeast genetics and Moleular Biology. In *Methods Enzymology*, Vol 169.

Cereghino GP, Cregg JM. 1999. Applications of yeast in biotechnology: protein production and genetic analysis. *Current opinion in biotechnology* **10**(5): 422-427.

Chandrasekaran S, Price ND. 2010. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America* **107**(41): 17845-17850.

Chen HC, Lee HC, Lin TY, Li WH, Chen BS. 2004. Quantitative characterization of the transcriptional regulatory network in the yeast cell cycle. *Bioinformatics* **20**(12): 1914-1927.

Covert MW, Schilling CH, Famili I, Edwards JS, Goryanin, II, Selkov E, Palsson BO. 2001. Metabolic modeling of microbial strains in silico. *Trends in biochemical sciences* **26**(3): 179-186.

Csete ME, Doyle JC. 2002. Reverse engineering of biological complexity. *Science* **295**(5560): 1664-1669.

Desvergne B, Michalik L, Wahli W. 2006. Transcriptional regulation of metabolism. *Physiological reviews* **86**(2): 465-514.

Edwards JS, Covert M, Palsson B. 2002. Metabolic modelling of microbes: the flux-balance approach. *Environmental microbiology* **4**(3): 133-140.

Edwards JS, Ibarra RU, Palsson BO. 2001. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nature biotechnology* **19**(2): 125-130.

Fell DA, Thomas S. 1995. Physiological control of metabolic flux: the requirement for multisite modulation. *The Biochemical journal* **311 (Pt 1)**: 35-39.

Fendt SM, Oliveira AP, Christen S, Picotti P, Dechant RC, Sauer U. 2010. Unraveling condition-dependent networks of transcription factors that control metabolic pathway activity in yeast. *Molecular systems biology* **6**: 432.

Friedman N, Linial M, Nachman I, Pe'er D. 2000. Using Bayesian networks to analyze expression data. *Journal of computational biology : a journal of computational molecular cell biology* **7**(3-4): 601-620.

Gianchandani EP, Papin JA, Price ND, Joyce AR, Palsson BO. 2006. Matrix formalism to describe functional states of transcriptional regulatory systems. *PLoS computational biology* **2**(8): e101.

Gietz RD, Woods RA. 2001. Genetic transformation of yeast. *BioTechniques* **30**(4): 816-820, 822-816, 828 passim.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M et al. 1996. Life with 6000 genes. *Science* **274**(5287): 546, 563-547.

Gueldener U, Heinisch J, Koehler GJ, Voss D, Hegemann JH. 2002. A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. *Nucleic acids research* **30**(6): e23.

Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J et al. 2004. Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**(7004): 99-104.

Haynes KA, Silver PA. 2009. Eukaryotic systems broaden the scope of synthetic biology. *The Journal of cell biology* **187**(5): 589-596.

J.R. Broach JRPaEWj. 1991. The Molecular and Cell Biology of Yeast Saccharomyces. In *Cold Spring Harbor Laboratory Press*, (ed. CS Harbor), New York.

Jensen PA, Lutz KA, Papin JA. 2011. TIGER: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. *BMC systems biology* **5**: 147.

Karin M. 1990. Too many transcription factors: positive and negative interactions. *The New biologist* **2**(2): 126-131.

Karlebach G, Shamir R. 2008. Modelling and analysis of gene regulatory networks. *Nature reviews Molecular cell biology* **9**(10): 770-780.

Kauffman S, Peterson C, Samuelsson B, Troein C. 2003. Random Boolean network models and the yeast transcriptional network. *Proceedings of the National Academy of Sciences of the United States of America* **100**(25): 14796-14799.

Kitano H. 2002. Computational systems biology. *Nature* **420**(6912): 206-210.

Krappmann S, Lipscomb WN, Braus GH. 2000. Coevolution of transcriptional and allosteric regulation at the chorismate metabolic branch point of Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* **97**(25): 13585-13590.

Kreeger PK, Lauffenburger DA. 2010. Cancer systems biology: a network modeling perspective. *Carcinogenesis* **31**(1): 2-8.

Latchman DS. 1997. Transcription factors: an overview. *The international journal of biochemistry & cell biology* **29**(12): 1305-1312.

Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I et al. 2002. Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science* **298**(5594): 799-804.

McAdams HH, Arkin A. 1997. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **94**(3): 814-819.

Nash S, Sofer A. 1996. *Linear and nonlinear programming*. McGraw-Hill, New York.

Nevoigt E. 2008. Progress in metabolic engineering of Saccharomyces cerevisiae. *Microbiology and molecular biology reviews : MMBR* **72**(3): 379-412.

Olson EN. 1990. MyoD family: a paradigm for development? *Genes & development* **4**(9): 1454-1461.

Ostergaard S, Olsson L, Nielsen J. 2000. Metabolic engineering of Saccharomyces cerevisiae. *Microbiology and molecular biology reviews : MMBR* **64**(1): 34-50.

Patil KR, Rocha I, Forster J, Nielsen J. 2005. Evolutionary programming as a platform for in silico metabolic engineering. *BMC bioinformatics* **6**: 308.

Price ND, Reed JL, Palsson BO. 2004. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature reviews Microbiology* **2**(11): 886-897.

Radovick S, Nations M, Du Y, Berg LA, Weintraub BD, Wondisford FE. 1992. A mutation in the POU-homeodomain of Pit-1 responsible for combined pituitary hormone deficiency. *Science* **257**(5073): 1115-1118.

Raychaudhuri S, Stuart JM, Altman RB. 2000. Principal components analysis to summarize microarray experiments: application to sporulation time series. *Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing*: 455-466.

Sauer U. 2006. Metabolic networks in motion: 13C-based flux analysis. *Molecular systems biology* **2**: 62.

Schilling CH, Letscher D, Palsson BO. 2000. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of theoretical biology* **203**(3): 229-248.

Shlomi T, Eisenberg Y, Sharan R, Ruppin E. 2007. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Molecular systems biology* **3**: 101.

Snyder M, Gallagher JE. 2009. Systems biology from a yeast omics perspective. *FEBS letters* **583**(24): 3895-3899.

Varma A, Palsson BO. 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Applied and environmental microbiology* **60**(10): 3724-3731.

Wall ME, Dyck PA, Brettin TS. 2001. SVDMAN--singular value decomposition analysis of microarray data. *Bioinformatics* **17**(6): 566-568.

Wall ME, R.; A, R. LM. 2003. "Singular value decomposition and prinicipal component analysis". in *A Practical Approach to Microarray Data Analysis. Los Almos national Laboratory* **02**(4001): 19.

Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H et al. 1999. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* **285**(5429): 901-906.

Yang MQ, Athey BD, Arabnia HR, Sung AH, Liu Q, Yang JY, Mao J, Deng Y. 2009. High-throughput next-generation sequencing technologies foster new cutting-edge computing techniques in bioinformatics. *BMC genomics* **10 Suppl 1**: I1.