STOICHIOMETRIC MODELS IN METABOLIC SYSTEMS BIOLOGY OF YEAST

by

Tunahan Çakır

B.S., Chemical Engineering, Boğaziçi University, 2001

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Graduate Program in Chemical Engineering
Boğaziçi University

2006

to My Mother, Melahat

and to My Niece Neslişah Sude...

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my thesis advisor, Prof. Kutlu Ülgen, for her everlasting support, encouragement and friendly attitude throughout my PhD study. Additionally, I sincerely thank Prof. Betül Kırdar for her support and her considerable contribution to all research chapters of the dissertation. I am thankful to Prof. Z. İlsen Önsan for her contribution regarding Chapters 4 and 5. I would like to acknowledge the other members of my thesis committee, Prof. Serpil Takaç and Assoc. Prof. Z. Petek Çakar Öztemel, for the time they devoted to read my thesis and their comments. Part of this study was carried out in Center for Microbial Biotechnology (CMB), at Technical University of Denmark (DTU), under the supervision of Prof. Jens Nielsen in August-September 2004 and June-November 2005. I gratefully acknowledge Prof. Nielsen for giving me the opportunity of visiting his center, as well as for his encouragement, never-ending optimism and substantial contribution to chapters 4 and 5. I also would like to thank him for the critical reading of the first part of chapter 3 and related helpful comments.

This research was supported by the Boğaziçi University Research Fund through projects 01M103, 03A504, 03S108 and 04HA502 and by DPT through project 03K120250. I gratefully acknowledge the doctoral scholarship provided by BAYG-TÜBITAK within the framework of the integrated PhD program, BDP, which included financial supports for three conferences (ECB-11 in Basel, ECB-12 in Lyngby, SYSBIOL-2006 in Murcia), one PhD course on Metabolic Engineering and Functional Genomics (Lyngby, 2004) as well as my stays in Lyngby in Summer 2004 and 2005. BÜVAK is acknowledged due to financial contributions regarding conferences.

I feel lucky to be involved in a very nice collaboration with Kiran Raosaheb Patil from CMB regarding the modeling of metabolome data presented in Chapter 5. This chapter would not be possible without him. Thanks Kiran for always ending up with very practical solutions whenever we encountered a problem during that work, and thanks for mostly-scientific chatting times. I also thank Isabel Rocha for her contribution to the initial phase of that work. I thank Joel F. Moxley (MIT) for providing me with a modified version of his GC-MS peak normalization software. Silas Villas-Bôas and Rasmus Devantier,

former PhD students in CMB, are gratefully acknowledged for their help in providing detailed information on their experimental metabolome data. I would like to thank two senior chemical engineering students, Elif Uzun and Ece Acuner, for their contribution to the last part of Chapter 4, regarding the alternative methods to generate flux distributions for CEF calculations. I am grateful to Steffen Klamt, for his immediate replies to my e-mails whenever I encountered problems regarding his FluxAnalyzer.

My special thanks are due to all current and former members of our Biosystems Engineering Research Group, especially Pınar Abla, Yalçın, Emrah, Hilal, Çağrı, Deniz, Betül, Duygu, Özde, Beste and Saliha, for providing a very-nice and very-friendly environment in our KB440, including scientific chats and gossips ☺. I thank all members of GenetikGeyikler -GeneticDeerings ☺-, the online version of such an environment, of which I was the 'top-deer' in terms of the number of sent messages. Thanks Yalçın & Emrah for beautiful times spent together in Europe from Basel to Copenhagen. I should not forget to thank Defne Abla, for enjoyable chatting moments full of joy and laughing.

I would like to deeply thank to Songsak, Mohammad, Ana and Birgitte (the secretary) for their unforgettable support and being always helpful during my stage in CMB, and all the members of Office 213 in Building 223, for making a beautiful environment and for feeling me home out there.

My family: I am deeply indebted for their never-ending support. I especially thank my mother for unbelievable support, for the coffees, teas, fruits and other stuff served directly to my desk while I was completely busy in front of the computer, and my 3-year-old princess Neslişah Sude, for enabling me to taste the feeling of being an uncle, and I dedicate this dissertation to these two ladies. Sadullah, I cannot underestimate your patience and tolerance during my 10 years in Boğaziçi University, especially during stressful moments, and I am sure one day you will better understand how much your brother loves you, cares you and thinks about you.

Now, I remember one phrase I had memorized in English-preparatory year in Fall 1996- my first year in Boğaziçi University- (10-years ago- meaning I am getting older ☹), "Time goes, you say? Ah no! Alas, time stays, we go". And, now is my turn to go ….

ABSTRACT

STOICHIOMETRIC MODELS IN METABOLIC SYSTEMS BIOLOGY OF YEAST

A system-level analysis of *Saccharomyces cerevisiae* metabolism was performed through integration of stoichiometric modeling and high-throughput 'omics' data. A bridge between metabolic networks and transcriptomics was built by employing the reactions involved in central carbon metabolism of the baker's yeast. The fold changes in control-effective fluxes (CEF), the weighted sum of calculated elementary modes passing through the reactions, were used for the prediction of the fold changes in mRNA transcripts of metabolic genes on different growth media (glucose-ethanol and galactose–ethanol). An acceptable correlation was obtained between the theoretical CEF-based flux ratios and experimental mRNA level ratios of 38 genes. Applicability of the approach to mammalian cell metabolism through analysis of red blood cell enzymopathies was also demonstrated. CEF approach was then employed to investigate the transcriptional regulation of fluxes in yeast metabolism for carbon shifts from fermentative (glucose) to nonfermentative (ethanol, acetate, lactate) substrates. An acceptable correlation was obtained for the analysis of such perturbation experiments, indicating that fluxes of yeast central metabolism are mainly transcriptionally regulated when there is a shift in carbon source.

An algorithm was developed to integrate metabolome data with metabolic network topology. The approach enables identification of reporter reactions, around which there are significant coordinated changes following a perturbation. Applicability of the algorithm was demonstrated for *S. cerevisiae.* Further combination of the results with transcriptome data enabled to infer whether the reactions are hierarchically or metabolically regulated.

Model-based structural robustness of yeast metabolism was analyzed to guide the research on phenomics. In silico lethality information of gene deletions on different carbon substrates indicated a more robust metabolism for *S. cerevisiae* than for *E. coli* bacterium.

ÖZET

MAYANIN METABOLİK SİSTEM BİYOLOJİSİ YAKLAŞIMIYLA İNCELENMESİNDE STOKİYOMETRİK MODELLERİN ROLÜ

Stokiyometrik modelleme ve hızlı-tarama yöntemleri sonucu elde edilen 'omics' verilerinin bütünleşik analizi yaklaşımıyla maya metabolizması sistem bazlı incelenmiştir. Mayanın merkezi karbon metabolizmasındaki tepkimeler baz alınarak hesaplanan temel akı yolları, metabolik ağyapıları ile transkriptom verilerinin bütünleştirilmesinde kullanılmıştır. Bu amaçla, bir reaksiyondan geçen akı modlarının ortalaması alınarak hesaplanan kontrol-etkili akıların (KEA) farklı besi ortamlarında gösterdiği değişiklikliğin (glikoz-etanol ve galaktoz-etanol), 38 genin mRNA transkriptlerindeki değişikliği başarılı ile öngördüğü görülmüştür. Bu yaklaşımın memeli metabolizmasına uygulanabilirliği, alyuvarlarda görülen enzim bozukluklarının analizi yoluyla gösterilmiştir. Daha sonra, KEA yaklaşımı ile, maya metabolizmasının akılarının karbon kaynağı değişikliği durumunda yazılımsal düzeyde kontrol edilebilirliği incelenmişir. Fermentatif (glikoz) karbon kaynağından fermentatif olmayan kaynağa (etanol, asetat, laktat) geçiş durumunda 5 farklı deneyde mRNA seviyelerinde görülen değişikliğin, karşılık gelen KEAlar ile uyumlu olduğu gözlemlenmiştir. Bu sonuçlar, karbon kaynağı değişikliği durumunda maya merkezi metabolizması akılarının yazılımsal düzeyde kontrol edildiğini göstermiştir.

Metabolom verilerinin metabolik ağyapılarıyla bütünleştirilmesini sağlayan bir algoritma geliştirilmiştir. Bu yaklaşım, maya metabolizmasında, etrafında en çok değişiklik görülen haberci tepkimelerin tesbitine olanak sağlamıştır. Sonuçların transkriptom verileriyle bütünleşik analizi sonucu, tepkimelerin hangi seviyede kontrol edildiği (hiyerarşik, metabolik) hakkında çıkarımlar yapılabilmiştir.

Fenomiks alanındaki araştırmalara öncül olması amacıyla, maya metabolizmasının model-tabanlı dayanıklılık analizi gerçekleştirilmiştir. Sonuçlar, *S. cerevisiae*'nin *E. coli*'ye göre daha dayanıklı olduğunu göstermektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS / ABBREVIATIONS

b	Vector representing net metabolite uptake by the cell
F	Network fragility score
k	The number of neighboring metabolites
R	Robustness score
R^2	Correlation coefficient
r_i	Flux of a particular reaction
S	Stoichiometric matrix
v	Metabolic reaction rate vector
v_i	Control effective flux of a reaction
Z	Standard normal distributed score for statistical significance
ε	Efficiency
μ	Mean
σ	Standard deviation
CE	Capillary electrophoresis
CEF	Control effective flux
CELLOBJ	Cellular objective
EFM	Elementary flux mode
EMP	Embden-Meyerhof-Parnas
ENZSUB	Enzyme subset
EP	Extreme pathway
FBA	Flux balance analysis
GC	Gas chromatography
HD	Hyperspherical direction
LC	Liquid chromatography
MCA	Metabolic control analysis
MCS	Minimal cut sets
MFA	Metabolic flux analysis

MNA	Metabolic network analysis
MPA	Metabolic pathway analysis
MS	Mass spectrometry
NMR	Nuclear magnetic resonance
PCA	Principal component analysis
PP	Pentose phosphate
SAM	Significance analysis of microarrays
TCA	Tricarboxylic acid
UNCOMP	Uncompartmented
UV	Ultraviolet
VHG	Very high gravity
YPD	Yeast peptone dextrose

1. INTRODUCTION

Systems biology has recently emerged as a promising field to interpret and understand the cell at systems level rather than studying the isolated parts of it. The advent of high-throughput experimental data collection technologies has been the driving force behind this new field. The transformation of biological research into a data-rich discipline was facilitated by the sequencing of the first genome in 1995. Today, genomes of more than 300 organisms have been sequenced, making the genomics one of the highly mature tools of systems biology. Availability of genome sequence of an organism has led to the birth of the other high-throughput technologies, namely transcriptomics (Lockhart and Winzeler, 2000) and proteomics (simultaneous screening of the levels of all expressed mRNA transcripts and proteins in the cell encoded by genome under certain conditions). Apart from the genetic material, simultaneous detection of the levels of small biochemical species, called metabolites, has been another focus, leading to a new field in data-rich biology, called metabolomics (Mendes, 2002). The conversion of these metabolites into each other through synthesis or decomposition is monitored by the enzymes, proteins with catalytic activities. The high-throughput quantification of the flux of the metabolites through enzymatic reactions is called fluxomics (Sanford *et al*., 2002). Phenomics, on the other hand, refers to high-throughput screening of cellular fitness or viability under certain conditions.

Systems biology deals with the analysis and integration of different types of 'omics' datasets through the use of computational tools to obtain overall quantitative description of cellular systems. Thereby, cells can be monitored *in silico*, facilitating more complete understanding of cellular functions. Computational modeling constitutes a crucial step in systems biology cycle, where dry *in silico* experiments such as simulations can be used to generate the assumptions and hypotheses. Wet experiments can be conducted to test the predictions of the models. Accordingly, models can be improved, and this cycle continues until an overall quantitative description of cellular functions is obtained (Kitano, 2002). Therefore, modeling tools are central to system-level understanding efforts.

Metabolism is a good starting point for systems biology research as it is studied in great detail and well annotated. Small-scale metabolic analysis of many organisms, from prokaryotes to mammalian cells and human, has been performed through static and dynamic modeling approaches. Genome-scale metabolic stoichiometric models have been developed for more than 10 microorganisms, facilitating the system-level understanding of cellular metabolism.

The aim of this study is integrated consideration of the metabolism of yeast, *Saccharomyces cerevisiae*, through the use of 'omics' data and computational tools. The background aspects of the conducted research are presented in the second chapter, entitled as "Background Aspects". The following chapters detail research under four main topics.

The first section of the third chapter introduces the application of a novel approach which links transcriptomics to fluxomics. Transcriptomic response of *S. cerevisiae* metabolism to carbon source perturbations was obtained from literature (Derisi *et al*., 1997; Griffin *et al*., 2002). Metabolic pathway analysis was used to identify a number of flux distributions, called elementary flux modes. The weighted combination of these distributions enabled assignment of a flux to each enzymatic reaction included in the stoichiometric small-scale model (53 reactions). Comparison of the fold change in these fluxes, called control-effective fluxes (CEF), with the fold change in corresponding mRNA transcripts led to an acceptable correlation. The second section in this chapter presents the application of this approach to a mammalian cell metabolism. The response of red blood cell metabolism (36 reactions) to enzymatic deficiencies, called enzymopathies, were investigated by calculating control-effective fluxes for non-deficient case and for the deficiency of five enzymes. Results reported in literature were used to identify relative importance of reactions in red blood cell metabolism, and, to verify the changes in CEFs in response to each deficiency by generating *in silico* deficiency profiles.

The fourth chapter focuses on the analysis of transcriptional regulation of metabolic fluxes of *S. cerevisiae*. Using a relatively larger model (83 reactions), the transcriptomic response for five carbon source perturbations (two in chemostat in respiratory mode, three in batch in respiro-fermentative mode) were compared with the flux response, obtained by

CEF calculation. Results indicated that transcriptome changes are hierarchically transmitted into flux level when the perturbation is in carbon source.

The fifth chapter gives the detailes of the research at metabolomic level. A new approach was developed to integrate metabolomic data with the metabolic stoichiometric model. A genome-scale yeast model was preprocessed and used as a scaffold to identify reporter reactions, the reactions which significantly respond to the environmental and genetic perturbations through changes in the levels of surrounding metabolites. A software package, named as RepRxn MX (Appendix G), was developed to automate the computational analysis. The approach was then systematically integrated with transcriptomic data, which enabled identification of regulation, i.e. whether a given reaction is regulated at the metabolic or at the transcriptional (hierarchical) level. The results indicate that although there are many metabolically regulated reactions in the metabolic network, regulation is predominantly hierarchical. This study can be regarded as one of the first steps towards the integration of different types of omics data by using metabolic networks as a scaffold in order to understand the architecture of metabolic regulatory circuits.

Theoretical investigation of structural metabolic robustness of *S. cerevisiae* is discussed in the sixth chapter. Different measures of structural robustness were compared for yeast metabolism. The *in silico* survival of the cell in response to single or multiple gene deletions was the basis for the approach. The advent in the phenomics area will enable the refinement and verification of the results presented in this chapter.

The summary of the main results and main contributions to the research field are given in the “Conclusions and Recommendations” chapter as well as the recommendations for future work.

2. BACKGROUND ASPECTS

2.1. Metabolic Engineering

Metabolism is the chemical engine that drives the living process (Edwards *et al.*, 1998). Metabolic engineering has emerged in the past decade as the interdisciplinary field aiming to improve cellular properties by using modern genetic tools to modify pathways (Stafford and Stephanopoulos, 2001). It was defined as the directed improvement of cellular activities through the modification of specific biochemical reaction(s) or the introduction of new one(s) with the use of recombinant DNA technology. (Figure 2.1) (Bailey, 1991; Stephanopoulos *et al.*, 1998). The rational approach of metabolic engineering seeks identification of the critical paths (or the metabolic bottlenecks) in a producing strain's metabolism and then directs the vast genetic engineering methodology to manipulate the identified enzyme systems (Vanrolleghem *et al.*, 1996).

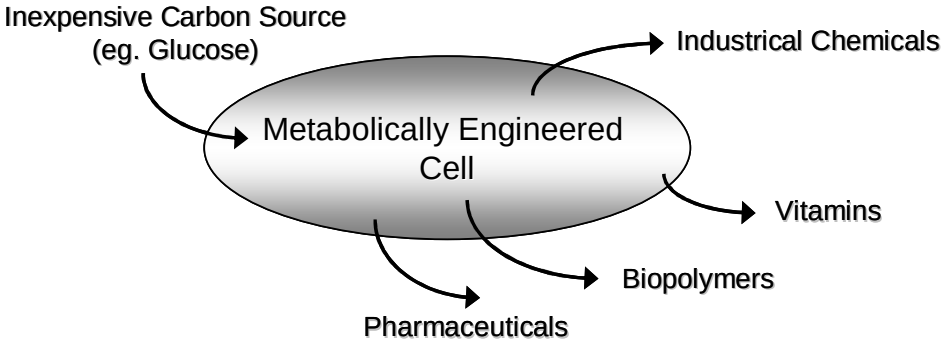


Figure 2.1. Metabolic engineering approach

On the experimental side, metabolic engineering applications have focused on pathway modifications in microorganisms to construct an improved strain of interest. There are a number of experimental tools utilized to genetically modify and analyze the cellular function for this aim (Nielsen, 2001). These tools are (Nielsen, 2001);

- availability of suitable strains and vectors that enable rapid transformation with reasonable transformation efficiency,
- access to promoters of varying strength,
- use of disruption cassettes to introduce specific genetic changes

- improved gene cloning techniques

The application of these tools led to the achievement of the following improvements in the metabolic engineering perspective (Stephanopoulos *et al.*, 1998; Ostergaard *et al.*, 2000; Nielsen, 2001),

- expression of heterologous genes for protein production
- extension of substrate range
- engineering pathways leading to new products
- engineering for the degradation of xenobiotics
- improvement of overall cellular physiology
- elimination or reduction of by-product formation
- improvement of yield or productivity.

On the modeling side, mathematical models have been used to elucidate the structure of metabolic pathways and distribution of kinetic control in metabolic pathways (Stafford and Stephanopoulos, 2001), which is discussed in detail in the following sections.

2.2. Models in Metabolic Engineering

The explosion in experimental data within biology has increased the attempts to develop mathematical models for description of cellular functions. One special focus within the biotechnology field is the cellular metabolism, since it has potential to be exploited for the production of compounds that might find industrial application as materials, pharmaceuticals, food additives, and so on. The metabolism of a living cell, however, is subject to regulatory mechanisms since it houses a number of interconnecting pathways that consist of hundreds of reactions. These regulatory mechanisms are difficult to identify since they are not completely elucidated yet. Thus, the establishment of whole-cell models to fully describe all aspects of cellular behavior in terms of its metabolism is, to date, not possible and all models are therefore based on simplifications. (Gombert and Nielsen, 2000).

Mathematical models that describe the cellular metabolism play a central role in the rapid developing field of metabolic engineering (Gombert and Nielsen, 2000). There are

two major classes of metabolic design tools, namely, stoichiometric models and flux analysis, and dynamic (kinetic) mathematical models of metabolic networks (Varner and Ramakrishna, 1999; Gombert and Nielsen, 2000). Dynamic modeling requires the knowledge of rate expressions and kinetic parameters, which is usually possible only for a small part of the cell metabolism, and mostly in *in vitro* conditions. Stoichiometric models, on the other hand, are based only on the reaction stoichiometry and reversibility, which are easy to obtain.

- Stoichiometric Modeling of Metabolism
    b. Metabolic Flux Analysis
    c. Flux Balance Analysis
    d. Metabolic Network Analysis with Isotope Labeling
    e. Metabolic Pathway Analysis
    f. Sampling Solution Spaces
    g. Graph Theoretical Methods

- Kinetic Modeling of Metabolism
    h. Dynamic Models
    i. Cybernetic Models
    j. Metabolic Control Analysis

2.3. Stoichiometric Modeling of Metabolism

Formulation of stoichiometric models requires two types of metabolic information. First, metabolic stoichiometry is required to compile all the chemical reactions that take place in the metabolic network of interest. The second needed information type is about the demands placed on the metabolic system, such as biomass synthesis, maintenance requirements and secretion of products. Afterwards, the collected metabolic information is put into the appropriate mathematical framework (Varma and Palsson, 1994). That is, mass balances are set up around intracellular metabolites, which lead to a stoichiometric matrix, **S**. This matrix puts a constraint on the cell such that all metabolic fluxes leading to the formation and degradation of any metabolite must balance, as represented in the following balance equation (Varma and Palsson, 1994, Stephanopoulos *et al.*, 1998):

$$\mathbf{S}.\mathbf{v} = \mathbf{b} \qquad (2.1)$$

Here, **v** is the metabolic reaction rate vector subject to the reversibility/ irreversibility constraints $\mathbf{v_{min}} < \mathbf{v} < \mathbf{v_{max}}$, and **b** is the vector representing the net metabolite uptake by the cell. The principles and related methods of stoichiometric modeling, which are discussed in the subsequent subsections, are summarized in Figure 2.2.

2.3.1. Metabolic Flux Analysis

Metabolic Flux Analysis (MFA) has been widely used for the quantification of the intracellular fluxes (steady state rates) in the metabolism of bacterial, yeast, filamentous fungi and animal cells. In MFA, mass balances over all the metabolites represented in matrix notation in Equation 2.1 are used to calculate the fluxes through the different branches of the network, enabling the identification of a snapshot of the metabolism under a particular condition. The fluxes are calculated by combining measurements of a few fluxes with linear algebra to render a determined (or overdetermined) equation system with zero degrees of freedom (Gombert and Nielsen, 2000). The directly measurable metabolic fluxes are the extracellular fluxes like product secretion, substrate uptake, oxygen uptake or carbon dioxide evolution and the growth rate (Stephanopoulos *et al.*, 1998, Wiechert, 2002). Stoichiometry-based MFA complements the stoichiometric relations by integrating them with the measured fluxes. If the flux measurements are nonredundant and they make the degrees of freedom of the system at least zero, then all intracellular metabolic fluxes can be estimated from the data. Thereby, (i) the structural identifiability of the fluxes can be decided, (ii) all fluxes, including the non-measured extracellular fluxes, can be computed efficiently, (iii) a confidence region for the estimated fluxes can be computed, (iv) the measured data can be derived explicitly, and (v) gross measurement errors can be detected. Thus, MFA is a mature tool for metabolic engineering (Wiechert, 2002).

2.3.2. Metabolic Network Analysis with Isotope Labeling

This is the extension of measured data set used for MFA by ^{13}C carbon labeling measurements (Wiechert, 2002). When the material balances used in traditional MFA are combined with balances of the labeling pattern of the metabolites, the models become non-

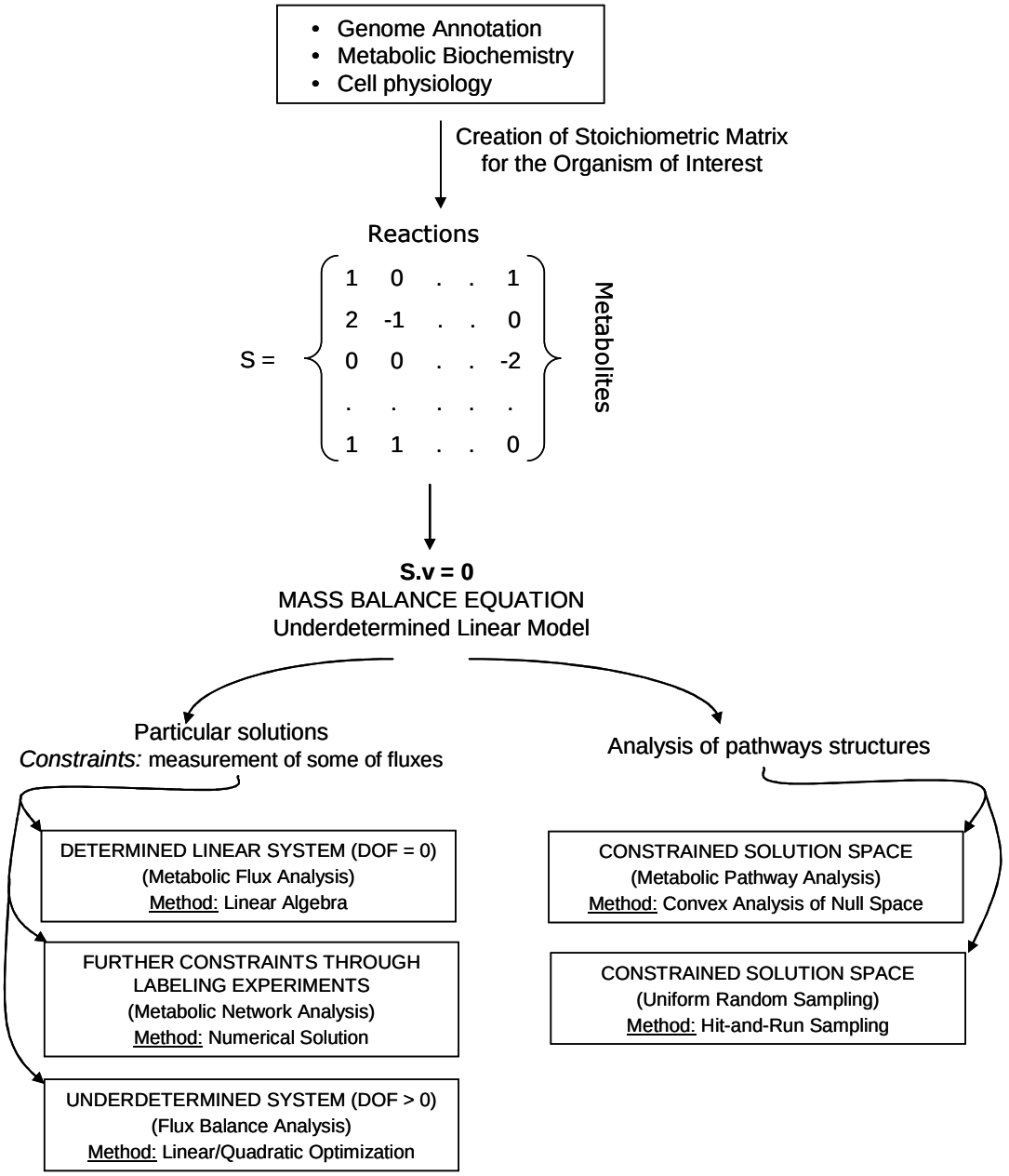


Figure 2.2. Principles and methods of stoichiometric modeling

linear. The additional information supplied by measurements of the labeling pattern of the metabolites do allow for more reliable quantification of the fluxes, as well as analysis of the pathway topology and possible reversibilities. Here, cells are fed with ^{13}C-labelled substrates, and the ^{13}C-enrichment in different carbon atoms of individual metabolites are measured using NMR or GC/MS. The enrichment patterns of individual carbon atoms give unique information about the activity of different operating pathways. This technique,

MNA, enables the identification of compartmentation of enzyme and metabolites within the cell as well as that of futile cycling and metabolic channeling (Gombert and Nielsen, 2000, Christiensen and Nielsen, 1999, Förster, 2003).

2.3.3. Flux Balance Analysis

Often it is not possible to determine an appropriate number of experimental constraints and hence to decrease the degrees of freedom to zero for estimating the metabolic fluxes in the resulting determined model by MFA. Since the system is underdetermined, meaning that the number of unknowns is greater than the number of equations, there is no unique solution. In this case, the metabolism can be investigated by formulating the stoichiometric model with an objective function, such as the optimization of growth (Förster, 2003). Hereby, the stoichiometric relations are complemented by a linear or quadratic metabolic optimization criterion. Several criteria like maximal growth rate, maximal product formation or minimal ATP production for a given substrate uptake can be investigated to get the corresponding optimal flux distribution. Such a formulation is a classical linear or quadratic programming problem since an optimality criterion is used to solve the metabolic flux balances to compute a unique solution due to the insufficiency of the number of measured fluxes (Wiechert, 2002). Detailed formulation of FBA approach is presented in Appendix A.

In underdetermined systems, a plurality of solution exists. The cell has an infinite number of choices on how to distribute its metabolic fluxes. This choice is constrained by the stoichiometric balance equation (Equation 2.1) and reaction reversibilities, which forms a domain of stoichiometrically allowable behavior of the microorganism (Figure 2.3). This allowable space defines the *metabolic genotype* of the strain since it describes the metabolic flux distributions that can be accomplished with the metabolic enzymes that the strain possesses (Varma and Palsson, 1994). Within this feasible domain, a single flux distribution is sought based on linear or quadratic optimization. Namely, microorganisms make their choices among several possibilities of the feasible domain such that their survivability is enhanced. The found single flux distribution represents the strain's *metabolic phenotype* under the particular conditions (Figure 2.3). In other words, metabolic genotype gives the domain of all possible flux distributions for a strain whereas metabolic

phenotype is the optimum flux distribution of the strain within this domain for a given environment. In this manner, FBA is an important tool to determine metabolic phenotype of the organisms quantitatively. (Varma and Palsson, 1994).

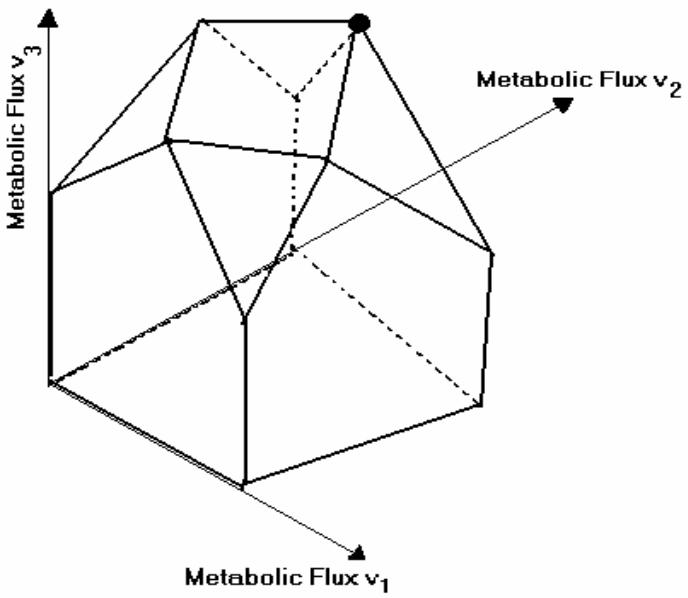


Figure 2.3. A hypothetical feasible domain defined by Equation 2.1, with the optimum point (metabolic phenotype) shown by a thick point. Adapted from Varma and Palsson (1994)

The linear optimization problems in FBA often exhibit multiple optimal solutions. This means that a number of flux distributions may exist that lead to the same optimal objective value (Förster, 2003). Methodologies are developed to calculate all these alternate optima (Lee *et al.*, 2000, Phalakornkule *et al.*, 2003, Mahadevan and Schilling, 2003, Zhu *et al.*, 2003).

2.3.4. Metabolic Pathway Analysis

The previous methods, MFA, MNA and FBA, allow the calculation of a particular solution, whereas metabolic pathway analysis (MPA) enables the screening for a number of different flux distributions or even the computation of all theoretical possible pathways within the defined metabolic network (Förster, 2003). Thus, MPA is one of the main approaches for the flux analyses of metabolic networks (Papin *et al.*, 2003; Schilling *et al.*,

1999). It is used to define the structure of the metabolic network and the overall metabolic capabilities of the microorganism. The method needs information only about the stoichiometry and the reversibility or irreversibility of reactions. Given the enzymatic reactions occurring in a particular microorganism, all possible routes are determined and analyzed. The possible paths from a substrate to products are depicted in Figure 2.4 for an example system.

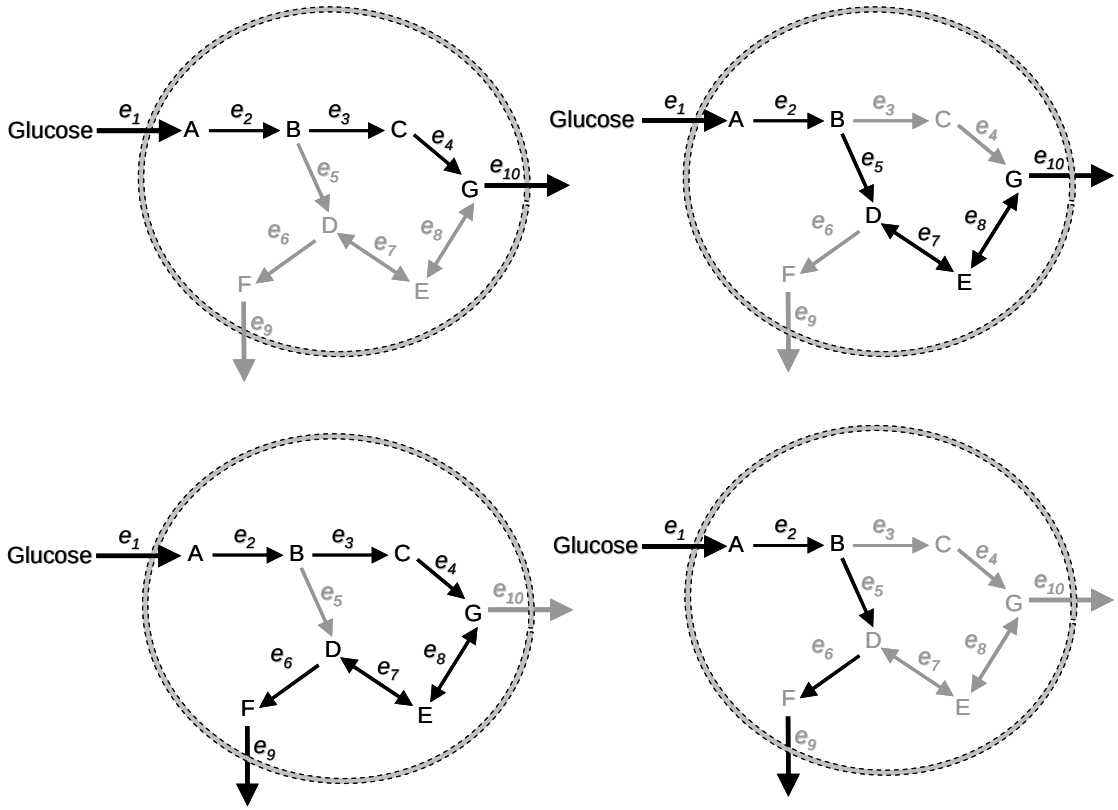


Figure 2.4. Graphical representation of 4 EFMs of an example network with 10 reactions and 8 metabolites

An important tool used in MPA is the detection of elementary flux modes (EFMs). An EFM is a minimal set of enzymes that could operate at steady state, with the enzymes weighted by the relative flux they need to carry for the mode to function (Schuster *et al.*, 1999; Schuster *et al.*, 2000; Schuster *et al.*, 2002a). EFM analysis allows the discovery and analysis of meaningful routes in metabolic networks. Control-effective flux (CEF) analysis is another tool in assessing metabolism by allowing the quantification of EFM results (Stelling *et al.*, 2002). The CEFs, which are directly determined from the set of EFMs,

represent the importance of each reaction in a metabolism for efficient and flexible operation of the entire metabolic network. Thereby, regulatory events of metabolism are implicitly incorporated. The method was successfully applied to *E. coli* for the functionality analysis based on the theoretical estimation of gene expression changes (Stelling *et al.*, 2002).

An alternative approach to the concept of EFMs is extreme pathways, EP, developed by Palsson's group (Schilling *et al.*, 2000, Palsson *et al.*, 2003, Klamt and Stelling, 2003a). Both approaches use convex analysis of the null space for the stoichiometric matrix in order to calculate a unique set of pathways. In EP analysis, a set of positively linear independent pathways mathematically equivalent to the edges of a convex solution cone are calculated, while in EFM analysis the method also includes some positively linear dependent pathways and this allows the direct calculation of all theoretically possible pathways within a metabolic network. Although EPs are mathematically more fundamental, it is more often easier to interpret the EFMs from a biochemical perspective (Förster, 2003).

Elementary modes are the smallest functioning subunits of a metabolic network. This motivates the hypothesis that they are also genetically regulated as a unit. Hence, this is a promising approach to the development of functional genomics tools (Wiechert, 2002, Schuster *et al.*, 2002b).

2.3.5. Sampling Solution Spaces

Metabolic pathway analysis enables investigation of overall metabolic capabilities of the organism of interest in defined environmental conditions, by simultaneous analysis of the calculated EFMs or EPs. However, it is not currently possible to apply MPA to the recent genome scale metabolic models, since the increase in the number of included reactions leads to an exponential increase in the number of EFMs, thereby making it impossible to calculate EFMs with current algorithms (Klamt and Stelling, 2002). An alternative approach therefore is the uniform random sampling of the flux solution space defined by the stoichiometric constraints and reaction reversibility information (Figure 2.3). One distinguishing feature of the collected flux distributions by random sampling

compared to EFMs is that they are not confined to the edges. That is, unlike EFMs, they mainly belong to the inner space. Sampling methods have been recently used for a genome scale metabolic model of *E. coli* (Almaas *et al.*, 2004); a model of human mitochondrial metabolism (Thiele *et al.*, 2005), and red blood cell metabolism (Price *et al.*, 2004a; Barrett *et al.*, 2006).

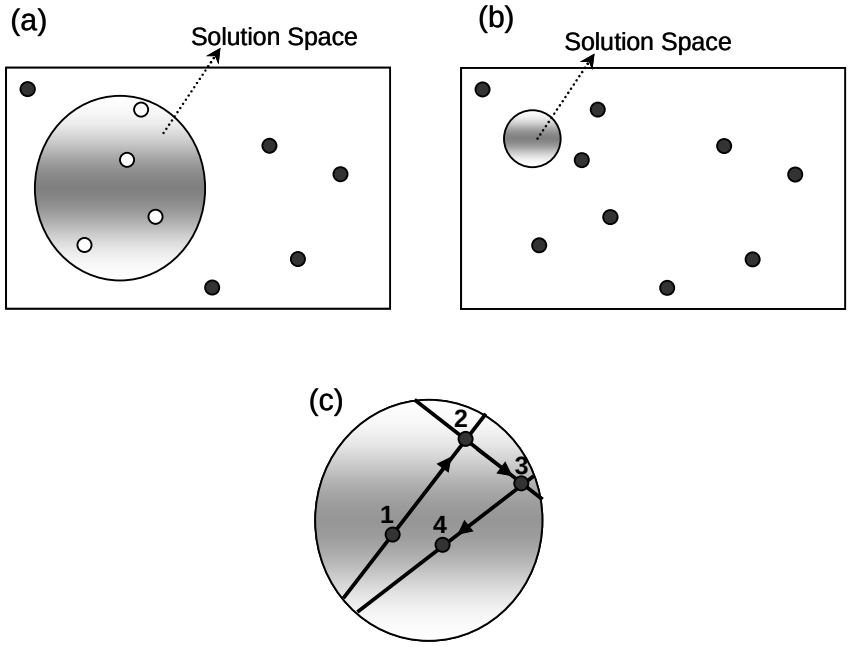


Figure 2.5. Sampling of solution spaces. (a) Monte-Carlo sampling by generating random points within an enclosing shape (b) Point collection can be impossible in some cases (c) Hit-and-Run sampling, starting from an initial point (1) inside the solution space

One fundamental approach for uniform random sampling is Monte-Carlo approach, where the solution space is enclosed with a multidimensional shape whose dimensions and volume are known (eg. Hypercube, hypersphere), and points which are guaranteed to fall into the enclosing space are randomly generated. Then, those which satisfy the inner solution space is identified and stored, for further analysis (Figure 2.5.a).

Although it is simple to implement Monte-Carlo approach for systems with small dimensions, the inner solution space becomes too small for bigger systems, eventually leading to a situation where none of the randomly generated samples falls within this space (Figure 2.5.b). Therefore, alternative approaches were developed, which does not need an

enclosing shape, and guarantees to remain always within the solution space of interest. One such approach is Hit-and-Run algorithm (Smith *et al.* 1984). Here, starting from a point (flux distribution) which is known to be inside the solution space satisfying stoichiometric constraints, a random line passing through this point is drawn, and its intersection points with the edges of solution space are determined. Then, random sampling within this line is performed, which ensures that obtained points are always within the feasible domain. This procedure is repeated until a desired number of points are collected (Figure 2.5.c). Those flux distributions can be used to calculate control effective fluxes.

2.3.6. Graph Theoretical Methods

Metabolic networks have potential to be analyzed by graph-theoretical tools after they are transformed into a graph-like representation. This method does not utilize the stoichiometric coefficients of reactions, but rather requires the stoichiometric reactions in order to identify and depict the interactions between enzymes (reactions) and metabolites. There are two main representation types, substrate graph and bipartite graph representations. Commonly used type is substrate (compound) graph (Figure 2.6) where metabolites correspond to *nodes*, and reactions correspond to connections (*links*) between these nodes (Ma and Zeng, 2003). The physical meaning of the link is the temporary educt-educt (substrate-substrate) complex itself, in which enzymes provide the catalytic scaffolds for the reactions yielding products, which in turn can become educts for subsequent reactions. This representation allows systematical investigation and quantification of the topologic properties of various metabolic networks using the tools of graph theory and statistical mechanics (Jeong *et al.*, 2000). The results obtained from a graph-theoretical perspective are helpful for understanding the general organization of metabolic networks (Klamt and Stelling, 2003b).

Two important characteristics of graphs are connectivity (degree) of nodes and path length between nodes. Connectivity of a node is the number of links (edges) attached to it. A path length is the number of edges connecting one node to another. Average shortest pathway is of particular interest and known also as the network diameter. If the network diameter is low, this means that the interaction between the nodes of the graph is high, and a perturbation in one part of the network can easily propagate and affect the other parts.

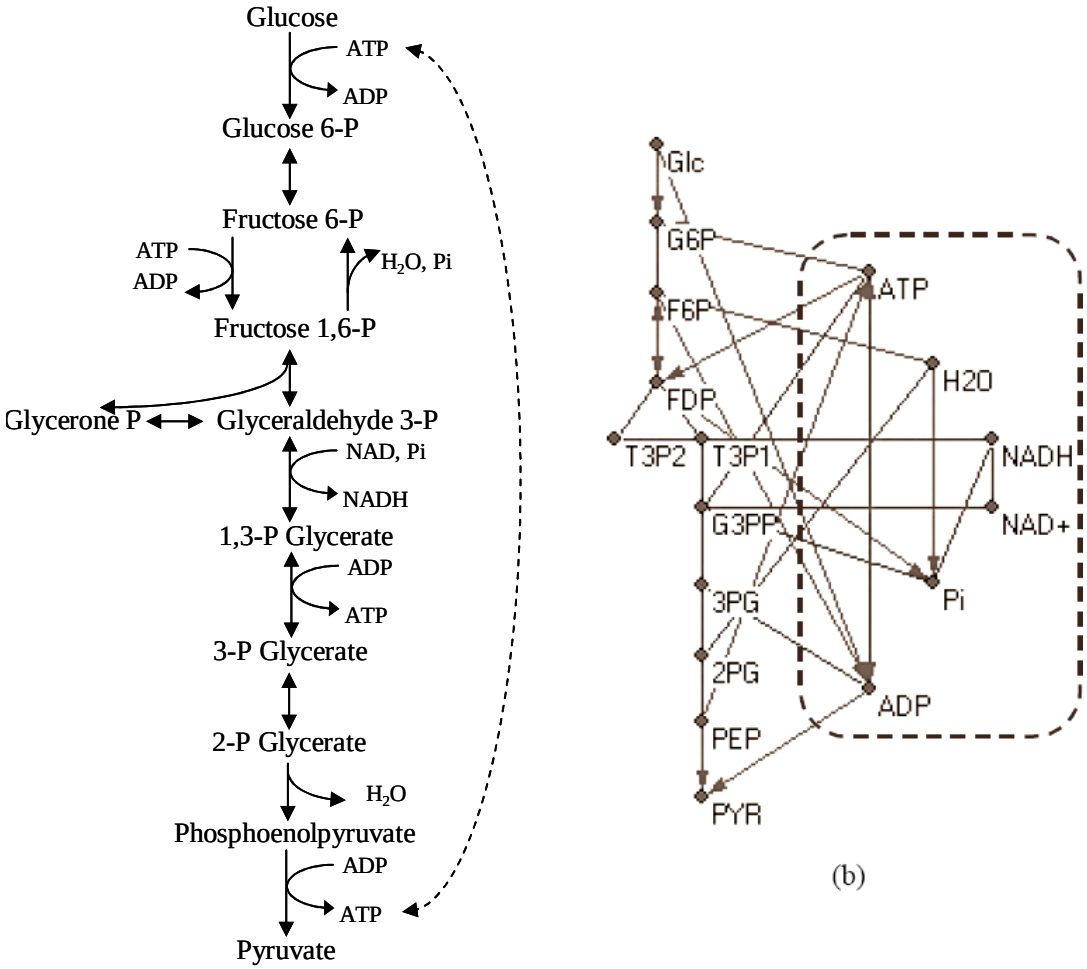


Figure 2.6. The glycolysis pathway as a part of metabolic network. (a) conventional presentation (b) the connection structure in a graphic representation . From Ma and Zeng, 2003

The structure of metabolic networks was studied using graph theory in detail (Jeong *et al.*, 2000; Wagner and Fell, 2001; Ma and Zeng, 2003; Lemke *et al.*, 2004). It was found that most of the nodes in metabolic networks have a low connection degree, while few nodes have a very high connection degree, as illustrated in Figure 2.7 (from Jeong *et al.*, 2000). Such networks are called scale-free networks, and their connectivity distribution follows power law. The high degree nodes dominate the network structure, and they are called hubs of the network (Ma and Zeng, 2003). Metabolic networks were found to have a small network diameter. Such networks are said to be small-world networks since to reach from any node to another is on average short, which further implies a high interconnection between the nodes.

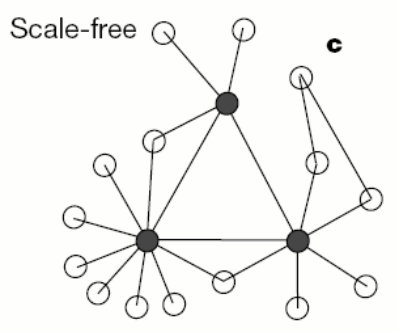


Figure 2.7. In the scale-free network most nodes have only a few links, but a few nodes, called hubs, have a very large number of links (hubs are shown as black nodes)

2.4. Kinetic Modeling of Metabolism

The availability of detailed information about the kinetics of specific cellular processes (eg. enzyme-catalyzed reactions, protein-protein interactions, or protein-DNA binding) enables analysis of the dynamics of these processes by combining kinetics with the known stoichiometry of metabolic pathways (Gombert and Nielsen, 2000). However, the main disadvantage of such models is the requirement of kinetic parameter values appearing in the rate expressions. Even if available to some extent, many of these parameters originate from *in vitro* experiments, rather than *in vivo* observations. In fact, cells exhibit different behaviours *in vivo*, which cannot be predicted by *in vitro* kinetic models (Teusink *et al.*, 2000).

2.4.1. Dynamic Models

By utilizing the available kinetic information about the enzymes of the selected pathways, one can simulate the cell behaviour in a dynamic manner in order to predict the levels of intracellular and extracellular metabolites in response to a disturbance to the cell. Most detailed kinetic models available in literature are for the two model microorganisms, *S. cerevisiae* and *E. coli* (Rizzi *et al.*, 1997; Chassagnole *et al.*, 2002). These models are limited with the reactions of central carbon metabolism (EMP pathway, PP pathway, TCA cycle, respiratory chain) since the rate expressions of other pathways, such as anabolic reactions leading to the synthesis of building blocks for the cell, cannot yet be represented by kinetic expressions. The validation of the models by the collected experimental data is

currently not promising since there are large discrepancies for a number of intracellular metabolite profiles. This can be attributed to the use of *in vitro* kinetic parameters. In addition, timescale used for the comparison (a few minutes) is quite limited to take into account the phenomena like enzyme synthesis and degradation (Gombert and Nielsen, 2000).

2.4.2. Cybernetic Models

Regulatory aspects of cell metabolism can be accounted by applying cybernetic principles. Here, kinetic models are integrated with cybernetic variables, which are thought to represent cell regulatory architecture, especially gene expression (Gombert and Nielsen, 2000, Varner and Ramkrishna, 1999). The principle is that physiology operates to optimally satisfy nutritional objectives. The resultant model is postulated to be capable of predicting modification of enzyme expression and activity profiles in response to genetic or environmental perturbation (Varner and Ramkrishna, 1999). Although the approach has a number of applications (Altıntaş *et al.*, 2002), it bears the stated drawbacks of kinetic models as stated above.

2.4.3. Metabolic Control Analysis

Metabolic Control Analysis (MCA) replaces the qualitative terms of ‘rate-limiting’ and ‘not rate-limiting’ by a quantitative scale, and the effect of an enzyme on a metabolic flux is then represented by flux control coefficient (Fell, 1997). Thus, MCA allows the quantification of flux control within a pathway. Flux control coefficients represent the relative increase in a given flux within the pathway of interest in response to an increase in the activity of an enzyme of the pathway (Nielsen, 2001; Stephanopoulos *et al.*, 1998). The most important result of MCA for metabolic engineering is that the control of flux through a pathway is usually distributed over many different enzymatic steps constituting the pathway. This fact points out that a single genetic modification will hardly result in a large change in the flux distribution (Wiechert., 2002).

There are different approaches for the calculation of control coefficients. Direct approach utilizes experimental methods such as genetic manipulation, enzyme titration and

inhibitor titration to construct a curve of the change in the pathway flux in response to the alterations in the enzymatic activity. The slope of this curve at any enzyme activity gives the control coefficient of that enzyme at that point (activity). (Fell, 1996). Although the experimental *in vivo* determination of control coefficients is a helpful approach for modeling and model validation, it is difficult to construct such a curve, since a number of perturbations are required to induce alterations in the activity of the enzymes of interest (Wiechert, 2002).

Indirect approach for the calculation of control coefficients requires information on enzyme kinetics. Kinetic rate expressions are necessary to calculate elasticity coefficients, a local property reflecting the response of steady-state reaction rate to a perturbation in the metabolite concentrations. Elasticity is calculated for each reaction in the system by introducing an infinitesimal change in a metabolite concentration and calculating the resultant effect on the rate of that reaction by using the kinetic rate expression. Calculated elasticity coefficients are combined with the theorems of MCA to obtain a quantitative estimation of control coefficients (Stephanopoulos *et al.*, 1998; Heinrich and Schuster, 1996; Fell, 1997). This approach is of importance when it is difficult to use direct methods. However, the calculated values of control coefficients strictly depend on the way the kinetic model constructed, reaction reversibilities, and the regulatory events accounted in the rate expressions.

2.5. Systems Biology Approach and *in silico* Biology

A metabolite, an enzyme or a gene in cellular systems do not function on their own, independently from the rest of the system, but they are highly interconnected to the other entities. This means that cell must be analyzed as a network rather than analyzing its constituents in an isolated manner. This view has led to the birth of a new discipline, called systems biology, aiming at understanding biology at system level. Systems biology can be defined as (Klipp *et al*., 2005) “the coordinated study of biological systems by (1) investigating the components of cellular networks and their interactions, (2) applying experimental high-throughput and whole-genome techniques, and (3) integrating computational methods with experimental efforts”. The driving force behind this new field is the achievability of high-throughput data collection, which allows the simultaneous

measurement of gene expression (transcriptomics), protein abundance (proteomics), metabolite levels (metabolomics) or fluxes (fluxomics) at genome level. All these –omes are closely interconnected within the cell, constituting a hierarchical structure (Figure 2.8). Holistic analysis of this structure at system level forms the field of systems biology.

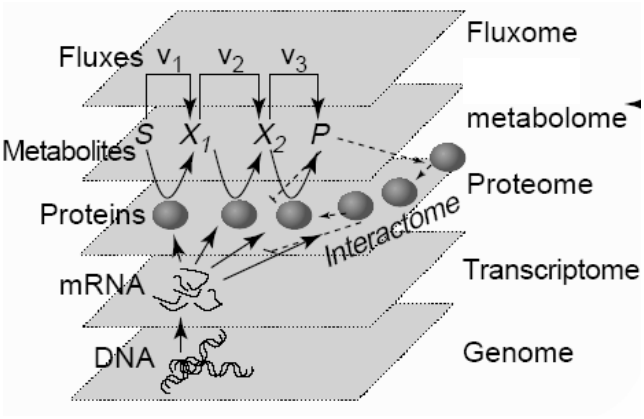


Figure 2.8. Interaction of different –omes in a cell, constituting a highly interconnected system. From Nielsen and Oliver, 2005

2.6. Transcriptomics: DNA chips

Earlier methods for transcription analysis such as northern blots allowed analysis of only a few genes at a time. High-throughput technologies have been recently developed enabling simultaneous analysis of mRNA levels of all genes in a genome. Knowing when, where and to what extent a gene is expressed is central to understanding the function of genes. Hence, use of DNA arrays is promising (Lockhart and Winzeler, 2000, Holloway *et al.*, 2002).

The use of DNA arrays for expression profiling is based on the fundamental process of hybridization. A DNA array is simply a surface of a solid support (e.g. glass) with about 1 cm x 1 cm dimensions, containing large sets of immobilized nucleic acid probe sequences at addressable locations available for hybridization. This is referred as downloading the genome onto a chip. Then, RNA is extracted from the biological sample of interest, and labeled RNA or DNA target sequences are produced by a number of sample preparation steps. The mixture of labeled targets is then applied to the DNA array under controlled conditions to allow hybridization with complementary immobilized

probes located on the array surface (Bro, 2003). Thereby, individual cDNAs/cRNAs from the target mixture hybridize (bind) with the corresponding probe on array, proportionally to their representation in a sample (Oliver *et al.*, 2002). In this design, the probes on the array act as immobile substrates whereas the samples (targets) applied onto the array are mobile substrates. After the hybridization step, fluorescent labeling is used to locate and quantify the binding of applied target sequences to their complementary probes on the array by imaging with a light scanner. The resulting scan is further processed using specialized computer softwares to calculate an intensity value for each gene represented on the array (Bro, 2003). The data can then be further analyzed to identify expression patterns and variations that correlate with cellular physiology and function. The resultant information can be helpful in assigning function to unknown genes, expanding our understanding of cellular processes, identification of potential drug targets and generation of genome-wide snapshots of transcriptional activity in response to any stimulus or perturbation (Harrington *et al.*, 2000).

Two basic types of arrays are available to; spotted microarrays and high-density oligonucleotide arrays (Bro, 2003, Harrington *et al.*, 2000, Oliver *et al.*, 2002, Lockhart and Winzeler, 2000). In spotted arrays, pre-synthesized single or double-strand DNAs are printed onto glass slides whereas in oligonucleotide arrays, developed by Affymetrix (Lipshutz *et al.*, 1999), sets of oligomers are synthesized in situ on glass wafers using a photolithographic manufacturing process. On spotted arrays, genes are represented by single cDNA fragments, greater than several hundred base pairs in length. On oligonucleotide arrays, a given gene is represented by a probe containing 15-20 different 25-mer oligonucleotides that serve as unique, sequence specific detectors. This representation minimizes the cross-hybridization between similar sequences (Harrington *et al.*, 2000, Oliver *et al.*, 2002). Additionally, there is only one probe per transcript in spotted arrays whereas Affymetrix oligonucleotide arrays contain 22-40 probes per transcript increasing the credibility.

2.6.1. Sample Preparation and Hybridization

Total RNA is extracted from the harvested cell samples, from which mRNA is isolated by exploiting the poly-A tail contained in most eukaryotic mRNAs. Next, mRNA

is converted to cDNA. The single-stranded cDNA is employed for further synthesis of double-stranded cDNA in the procedure of Affymetrix chips. Isolated double-stranded cDNA is used to synthesize cRNA via *in vitro* transcription reaction. Afterwards, the cRNA is fractionated to fragments of length 35-200 nucleotides. Finally, the fragmented cRNA is hybridized to DNA array, washed to remove non-hybridized material and then stained with a fluorescent dye. Labeling of the applied target is required to detect hybridization on an array (Bro, 2003).

2.6.2. Data Analysis

Affymetrix GeneChips contain more than 22 probes per transcript, to enable an improved estimate of intensity detection for each transcript. Intensity of each gene is calculated from the corresponding probe sets.

There are three fundamental steps required for efficient and effective data analysis: data normalization, data filtering, and pattern identification (Harrington *et al.*, 2000).

2.6.2.1. Normalization. Direct comparison of expression values (gene intensities) usually requires normalization of data, either between paired samples or across a set of experiments (Harrington *et al.*, 2000). Namely, normalization is necessary to make the gene expression levels on the arrays comparable. A common approach for this is based on the assumption that the total amount of mRNA from each sample (array) is constant. Thus, all expression levels from an array can be multiplied by a scaling factor to make the sum of all expression levels the same for all arrays (Bro, 2003), but there are also more sophisticated methods like non-linear scaling (Bolstad *et al.*, 2003).

2.6.2.2. Data Filtering and Statistical Analysis. DNA expression array studies usually aims at identification of differently expressed genes in one or several samples compared to a control. Thus, following the normalization step, data should be reduced by filtering out uninformative genes; for example, genes that are expressed below a user defined threshold or genes that show no variation in their expression level during the course of experiment (Harrington *et al.*, 2000).

The easiest way to identify the magnitude of change in gene expression levels is to calculate how many folds the expression change. However, in order to gain confidence in that the observed changes are a result of true changes of mRNA levels in the cell rather than due to experimental errors, it is of importance to have replicates and to use statistical tests. Typically, widely known statistical tests such as t-test (to compare two conditions) and ANOVA (to compare multiple conditions) can be applied to associate the change with a probability, called p-value. Alternative methods are also continuously developed such as the Significance Analysis of Microarrays (SAM) (Tusher *et al.*, 2001) and RankProduct (Breitling *et al.*, 2004). When, as in the case of DNA array data, huge amounts of data are simultaneously tested using statistical tests, correction for multiple testing must be employed using Bonferroni or Benjamini-Hochberg corrections (Kaminski and Friedman, 2002; Bro, 2003).

2.6.2.3. Pattern Identification. The next step is to identify patterns and groups in the data that can be used to assign biological meaning to the expression profiles (transcriptional data). Clustering programs such as hierarchical clustering and k-means clustering can be used for this purpose (Harrington *et al.*, 2000, Halloway *et al.*, 2002). Clustering methods attempt to identify genes that behave similarly across a range of conditions or samples. Genes that demonstrate similar patterns of expression are hypothesized to share common regulatory elements or common functions (Kaminski and Friedman, 2002). The main advantage of clustering tools is that they eliminate the inherent difficulty in becoming familiar with the results, allowing the investigator to analyze the data in an automated and practical fashion. In general, it is recommended to apply more than one clustering method to a dataset. The results of several methods and of several gene filtering schemes should be compared to decide on what is the true signal in the dataset of interest. The decision should lie on the reproducibility of the clusters by various methods. Thereby, independent of the chosen analysis method, one can confidently conclude about whether the observed patterns represent true biological phenomena (Kaminski and Friedman, 2002).

2.7. Metabolomics: Metabolome Analysis and Metabolite Profiling

There are three main approaches in metabolic data analysis: metabolite profiling; metabolomics, and metabolic fingerprinting (Fiehn, 2001; Fiehn and Weckwerth, 2002).

Metabolite profiling is the analysis and quantification of pre-defined metabolites belonging to a certain biochemical pathway or a set of chemically related compounds. Metabolomics, on the other hand, has the ultimate goal of unbiased identification and quantification of all the metabolites present in a certain biological sample irrespective of their chemical similarity or their co-occurrence in a pathway. The third approach, metabolic fingerprinting gives a rapid classification of samples without aiming at identification and quantification of metabolites. Fingerprinting approach can be used in genotype discrimination. Here, only the data patterns in the form of the whole spectra are compared without any knowledge of which peaks belong to which metabolites. It requires little or no sample preparation, compared to the other approaches. (Fiehn, 2001; Fiehn and Weckwerth, 2002). There is also a closely related term, metabolic footprinting, which depends on the generation of spectra for only extracellular metabolites, without again any quantification (Allen *et al.*, 2003).

Metabolomics is the most recent face of functional genomics. Like other 'omics', it is a non-targeted approach to study biological phenomena, by simultaneous profiling large number of small organic molecules (metabolites) encountered in an organism (Mendes, 2002). Whereas there has been significant progress in high-throughput profiling of mRNAs and proteins, comparably less effort has been put into profiling the end products of gene expression, metabolites (Figure 2.9; Fiehn, 2001, Sumner *et al.*, 2003). The major underlying reason is the current inability to comprehensively profile all of the metabolome because of limitations such as chemical complexity (Sumner *et al.*, 2003). The genome and transcriptome are made up of linear polymers of four nucleotides with very similar chemical properties, which render high throughput analytical approaches possible. This is also valid for proteome, which, although being substantially more complex, is still based on a limited set of 22 primary amino acids. When one surveys metabolome, however, the chemical complexity is considerably greater. Consequently, the chemical diversity and complexity of the metabolome is a challenging limiting factor to profile all of the metabolome simultaneously (Sumner *et al.*, 2003).

Changes in transcriptome or proteome do not always correlate with alterations in biochemical (metabolic) profiles. On the other hand, the components of the metabolome, metabolites, define the biochemical phenotype of a cell, and can be regarded as the final

output of gene expression (Figure 2.9). Hence, metabolome profiling can provide the most functional information among the omics technologies (Sumner *et al.*, 2003).

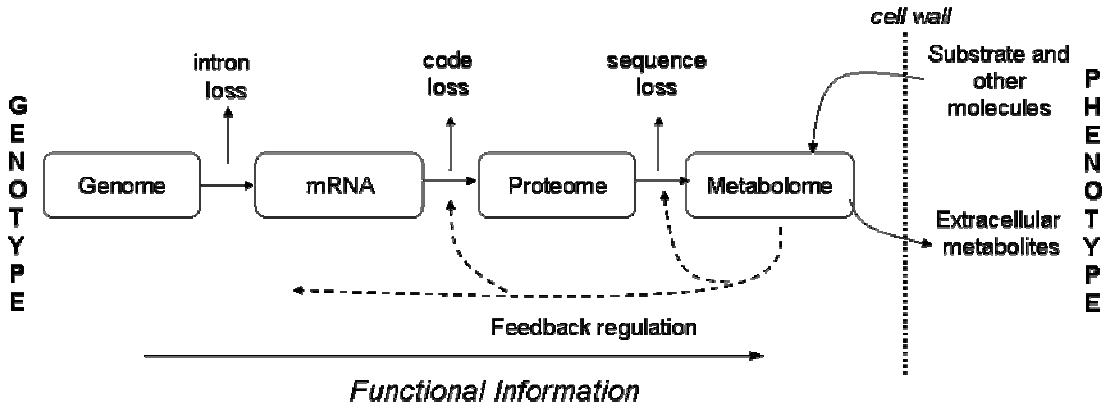


Figure 2.9. Metabolomics is the end product of genome, with, among the omes, most direct relation to phenotype. Adapted from Nielsen *et al.*, 2004

2.7.1. Metabolite Identification

A variety of analytical methods is available to generate metabolite profiles, with emphasis in mass spectrometry (MS) and nuclear magnetic resonance (NMR) (Mendes, 2002). Among the other methods to profile metabolome are LC (liquid chromatography)/UV (ultraviolet), GC (gas chromatography)/MS, LC/MS, LC/LIF (laser induced fluorescence) and CE (capillary electrophoresis)/LIF. The selection of the most suitable method is generally a reconciliation between speed, selectivity, and sensitivity. For example, NMR is rapid and selective, but has relatively low sensitivity. CE/LIF detection is highly sensitive, but it lacks selectivity (Sumner *et al.*, 2003). That is, each analytical detection method itself has a certain bias or advantage (Fiehn and Weckwerth, 2002). For example, GC requires that samples be volatile. Hence, derivatization is required for nonvolatile samples. LC, on the other hand, allows the analysis of nonvolatile species without the need for derivatization (Sumner *et al.*, 2003). MS requires that metabolites be ionizable, ultraviolet absorption (UV) assumes that biochemical compounds have moieties with excitable electrons. Therefore, no single best metabolomic technique exists (Fiehn and Weckwerth, 2002). Also, a single analytical technique will not provide sufficient visualization of the metabolome due to chemical differences.

Metabolomics aims at the identification of metabolite components of a mixture. This requires much deeper analysis compared to the fingerprinting approach. First, deconvolution of the fingerprint (spectrum, chromatogram etc.) into entities that represent the component biochemical species is required. Afterwards, a reference database (metabolite library) must be looked up to identify the biochemical species (Mendes, 2002). Also, automated metabolite identification requires reliable information on both retention time and mass spectra (Fiehn, 2001).

2.7.2. Data Interpretation

Metabolomics generates multivariate datasets like transcriptomics. Hence, statistical tests discussed in section 2.6.2.2 can also be applied to metabolome data to identify metabolites which significantly change between any two conditions. Additionally, unsupervised methods such as principal components analysis (PCA) and clustering; or supervised methods such as machine learning algorithms can be used to analyze the emerging metabolome data. Such statistical analyses allow the classification of experimental data. Thus, they are preferred when the main objective is to classify samples based on their metabolite profiles, which is usually held by the fingerprinting approach (Mendes, 2002). One other novel approach is the framework of metabolic control analysis with co-response approach (Raamsdonk *et al.*, 2001). In this approach, the co-response of metabolite concentrations of a mutant against the deletion of a functionally unknown gene is to be determined. A function can be assigned to the deleted gene based on the similarity to co-responses of concentrations of other mutants whose functionally known genes are deleted. Other novel approaches are being developed in parallel with the increasing attention towards metabolome data collection (Weckwerth *et al.*,2004; Kümmel *et al.*, 2006)

2.8. Fluxomics: Stoichiometric Models and Isotope Labeling

Determination of metabolism-wide fluxes by the use of organism-specific stoichiometric models is termed as fluxome. Fluxome represents the functional output of the combined transcriptome, proteome, and metabolome changes as it forms the top level of the hierarchical structure (Figure 2.8).

Approaches discussed in sections 2.3.1., 2.3.2 and 2.3.3 form the basis of fluxome analysis. FBA is the easiest way of determination of fluxes since it requires only measurement of a few exchanged fluxes. The easy determination of fluxes through FBA led to a number of related applications. Recently reconstructed stoichiometric genome-scale models of a number of microorganisms, among which are *Escherichia coli* (904 genes, 931 reactions and 625 metabolites, Reed *et al*., 2003), and *Saccharomyces cerevisiae* (708 genes, 842 reactions and 584 metabolites, Förster *et al*., 2003a), have been successfully analyzed by optimization (FBA). FBA is used as an alternative to the relatively difficult task of measuring the internal fluxes in the metabolic network since such genome-scale networks are underdetermined in nature. Major applications have been to determine exchange fluxes of metabolic products (Edwards *et al.*, 2001; Famili *et al.*, 2003) and to determine phenotypic differences in metabolic behaviour of organisms through phase plane analysis (Edwards *et al*., 2002; Duarte *et al.*, 2004). Additionally, prediction of the outcomes of genetic manipulation has been a major focus. Qualitative viability comparison for knockout strains with the FBA-based prediction was more than 80 per cent for *E. coli* (Edwards and Palsson, 2000) and yeast (Förster *et al*., 2003b). FBA also helped to propose a reasoning for the dispensability and essentiality of yeast genes (Papp *et al*., 2004). Another approach has been to identify the genes whose deletions will lead to a desired phenotype. FBA through mixed-integer linear programming, called OptKnock procedure (Burgard *et al*. 2003), and through integration of genetic algorithm, called OptGene procedure (Patil *et al*. 2005) were developed to serve this aim.

One major challenge for metabolic engineering in post-genomic era is integration of genome-scale biological data into mathematical and experimental tools of the discipline. Therefore, the recent trend has been to incorporate additional biological knowledge such as regulatory constraints in order to refine the resulting flux values. It was demonstrated that the prediction capability of genome-scale yeast model was improved when regulatory information from gene expression data in Boolean logic formulation was incorporated (Akesson *et al.*, 2004). In a recently updated *E. coli* metabolic model integrated with transcriptional regulatory model, growth and gene expression simulations were done by regulatory FBA, which combines linear optimization to determine a growth-optimized metabolic flux distribution with logic statements to simulate the effects of regulatory

processes over time. The model was able to predict high throughput phenotyping and gene expression experiments (Covert *et al.*, 2004).

Although FBA allows a quick estimate of metabolic fluxes, solid estimate of fluxes requires additional experimental constraints ^{13}C labeled glucose feeding experiments (Section 2.3.3). Of the two approaches, NMR and MS analysis, the latter has potential for high-throughput analysis at miniaturized scale. High-throughput flux profiling for the central carbon metabolism was recently achieved by GC-MS analysis (Fischer *et al.*, 2004). This was followed by the screening of a number of deletion mutants at the fluxome level (Fischer and Sauer, 2005; Blank *et al.*, 2005). The analysis is based on the use of a small-scale stoichiometric matrix. Resulting fluxes from small model can be used as constraints to get an estimate of genome-wide metabolic fluxes through the use of genome-scale models (Blank *et al.*, 2005). On the other hand, intracellular fluxome in the form of fingerprint, as in the case of metabolome fingerprints, can be used for metabolic variant discrimination (Zamboni and Sauer, 2004).

3. METABOLIC PATHWAY ANALYSIS OF YEAST AS A TOOL FOR SYSTEMS BIOLOGY AND EXTENSION OF THE APPROACH FOR HUMAN METABOLISM

3. 1. Metabolic Pathway Analysis of Yeast Strengthens the Bridge between Transcriptomics and Metabolic Networks

Recent developments in the field of "omics" have resulted in the accumulation of huge number of experimental data to be analyzed. The challenge is to develop powerful methods for the integrated analysis of system properties. Therefore, the need for system level understanding of living organisms, which explains the relationship between structure, function and regulation in complex cellular networks by combining experimental and theoretical approaches, is growing in the post-genomic era.

One of the recent challenges in the area of transcriptomics is to relate gene expression levels to the fluxes carried through the enzymes encoded by these genes (Oh and Liao, 2000; Oh *et al.*, 2002). A novel theoretical approach was devised to simultaneously predict key aspects of network functionality, robustness and gene regulation in *Escherichia coli* from stoichiometric network structure alone (Stelling *et al.*, 2002). Elementary flux mode analysis, one of the tools of metabolic pathway analysis (Klamt and Stelling, 2003a; Palsson *et al.*, 2003), allows implicit integration of regulatory events into the stoichiometric metabolic analysis. Compared to the flux balance analysis (FBA), which only considers efficient operation of the cell (Bonarius *et al.*, 1996; Edwards *et al.*, 2001; Pramanik and Keasling, 1997; Shi *et al.*, 1999), elementary mode analysis takes also flexibility of the metabolism into consideration (Schuster *et al.*; 1999; Stelling *et al.*, 2002). Introduction of a parameter characterizing flexibility and efficiency derived from metabolic network structure, called control effective flux (CEF), establishes an indirect relation between transcriptomics and fluxomics.

In this chapter, central carbon metabolism of the yeast *Saccharomyces cerevisiae* was analyzed using metabolic pathway analysis tools. Elementary flux modes for growth

on three carbon substrates (glucose, galactose and ethanol) were determined using the catabolic reactions occurring in yeast. Resultant elementary modes were used for gene deletion phenotype analysis and for the analysis of robustness of the central metabolism and network functionality. Control effective fluxes, determined by calculating the efficiency of each mode, were used for the prediction of transcript ratios of metabolic genes in different growth media (glucose-ethanol, galactose-ethanol). A high correlation was obtained between the theoretical and experimental expression levels of 38 genes when ethanol and glucose media are considered. Such analysis was shown to be a bridge between transcriptomics and fluxomics through this study. Control effective flux distribution was found to be promising in the *in silico* predictions by incorporating functionality and regulation into the metabolic network structure. Thus, theoretical transcript ratio analysis strengthens the relationship between transcriptomics and metabolic networks.

3.1.1. Modelling Aspects

53 reactions constituting the central metabolism of *Saccharomyces cerevisiae* (glycolysis/gluconeogenesis, pentose phosphate pathway, citric acid cycle, glyoxylate shunt and oxidative phosphorylation) were considered in the construction of an *in silico* model (Appendix B). The reaction set described by Förster *et al*. (2002) was extended by the complementary reactions presented in literature (Förster *et al.*, 2003a; Gombert *et al.*, 2001; Granström *et al.*, 2000; Nissen *et al.*, 1997; Stückrath *et al.*, 2002; van Gulik and Heijnen, 1995). Compartmentation of cofactors (NADH and NADPH) and acetyl-coenzyme-A was taken into account by treating these compounds as if two distinct species existed in mitochondrial and cytosolic reactions. Enzymes of peroxisomes were treated to be cytosolic (Förster *et al.*, 2003a).

The biomass equation is given in terms of the biosynthetic precursors (r_{53}). This representation requires molar monomer composition of proteins, nucleotides, lipids and carbohydrates. The number of required precursors for the synthesis of each monomer is known. This information is combined with the information of utilized macromolecular mass composition to obtain the equation (Cortassa *et al.*, 1995). The reported monomer and macromolecule composition values differ noticeably in the literature for

Saccharomyces cerevisiae (Cortassa *et al.*, 1995; Stückrath *et al.*, 2002). However, it was found that a 20 per cent change in the composition of any of biomass precursors had only minute effect on the resultant flux patterns (Daae and Ison, 1999). Based on these facts, same stoichiometric coefficients were used in biomass equation in ethanol and glucose media. The fact that the determined elementary modes and their routes were found to be insensitive to the biomass composition indicated the validity of this assumption.

For glyoxylate shunt enzymes, ICL1 is found to be localized in the cytosol (Taylor *et al.*, 1996). MLS1 enzyme is also primarily cytosolic (Kunze *et al.*, 2002). CIT and MDH have different forms functioning in cytosol, peroxisome or mitochondria. Cytosolic forms (CIT2, MDH2) fuel glyoxylate pathway whereas mitochondrial isoenzymes (CIT1, MDH1) function in the TCA cycle (Kispal *et al.*, 1988, 1989; Roth and Schüller, 2001). Thus, all glyoxylate shunt enzymes were considered to be cytosolic (r_{46}-r_{49}). The reversible conversion of citrate to isocitrate catalyzed by ACO1 (r_{33}) is both cytosolic and mitochondrial, hence is important for both TCA cycle and glyoxylate shunt.

The gluconeogenic enzymes (FBP1, PCK1) are essential when growth is on ethanol and hence considered in the present stoichiometric model (r_6 and r_{23}). Pyruvate branch enzymes PDC and PDA (r_{16} and r_{21}) were also considered (Boubekeur *et al.*, 1999; Flikweert *et al.*, 1996). Malic enzyme (MAE1) was reported to be active in *Saccharomyces cerevisiae* (Boles *et al.*, 1998, Cortassa *et al.*, 1995). Thus, reaction 43 (r_{43}) was included into the stoichiometric model.

Reaction 45 (r_{45}) represents the shuttle of acetaldehyde and ethanol for the net transfer of NADH from cytosol to mitochondria (Bakker *et al.*, 2001; Förster *et al.*, 2002; Overkamp *et al.*, 2000). The shuttle reactions are catalyzed by the mitochondrial alcohol dehydrogenase enzyme (ADH3) and cytosolic isoenzymes (ADH1,2).

In conventional flux analysis methods, from microorganisms to mammalian cells (Lee *et al.*, 2000), only the reactions belonging to the corresponding active pathways are used to analyze a given physiological condition. That is, inactive pathways are not included in models. However, in this analysis, although, for example, IDP2 and ALD4 enzymes are thought to be inactive when growth is on glucose medium (Haselbeck *et al.*,

1993; Haurie *et al.*, 2001; Kurita and Nishida, 1998; Loftus *et al.*, 1994; Minard *et al.*, 1998; Tessier *et al.*, 1998), these reactions were included to be able to predict transcript ratios of corresponding genes in two different conditions. Following the same logic, all of the gluconeogenesis and glyoxylate shunt enzymes can be safely included whereas these are not accounted in MFA studies due to their inactivity in glucose medium (Jorgensen *et al.*, 1995; Nissen *et al.*, 1997; van Gulik and Heijnen, 1995). This approach allowed improved predictions of corresponding mRNA level ratios.

3.1.2. Computational Methods

Elementary flux modes were calculated using METATOOL 3.5.2 and 4.3 (Pfeiffer *et al.*, 1999; Schuster *et al.*, 2002a) and FluxAnalyzer 4.0 (Klamt *et al.*, 2003). Control effective flux calculations were performed using Microsoft Excel and using a script written in MATLAB (MathWorks Inc.).

CEFs were determined directly from the set of elementary modes. After each mode was normalized with respect to the substrate flux, an efficiency value was assigned to each elementary mode for each cellular objective, $\varepsilon_{j,CELLOBJ}$, using the ratio of mode's outputs (reactions related to cellular objectives) to the investment required to establish the mode (the sum of absolute fluxes of each mode) as described (Stelling *et al*., 2002). *j* is the index for EFMs, and *i* is the index for fluxes.

$$\varepsilon_{j,CELLOBJ} = \frac{r^{j}_{CELLOBJ}}{\sum_{i} \left| r_i^{j} \right|} \tag{3.1}$$

In efficiency calculations, biomass flux as well as ATP maintenance flux were taken as the mode outputs since these reactions are basis for the cellular objectives of the yeast.

Then, the flux of a particular reaction, r_i^j, in all determined elementary modes was weighted with the efficiency of the corresponding modes to calculate the control effective flux (CEF), v_i, of the reaction.

$$v_i = \sum_{CELLOBJ} \frac{1}{r_{CELLOBJ}^{\max}} \frac{\sum_j \varepsilon_{j,CELLOBJ} \left| r_i^j \right|}{\sum_j \varepsilon_{j,CELLOBJ}} \qquad (3.2)$$

The ratio of CEFs of reactions in two different media is used to predict the expression ratios of metabolic genes responsible for the enzymes of the reactions in *Saccharomyces cerevisiae.*

3.1.3. Elementary Flux Mode Calculations for Yeast Metabolism

Elementary flux modes were calculated for yeast grown in three different substrates, namely glucose, ethanol and galactose (Table 3.1). In cases 1.a, 2.a and 3, all possible reactions, whether they are active or not on these specific substrates were taken into consideration. Cases 1.b and 2.b are environment-specific. PCK1, ALD4, IDP2, FBP1 enzymes and glyoxylate shunt enzymes were excluded in case 1.b and only the reactions that are known to be active in a medium when the only available carbon substrate is glucose were considered. PYK and PFK enzymes were excluded in case 2.b and the reactions that are known to be active in a medium where the only available carbon source is ethanol were considered. The reason to analyze cases 1.b and 2.b is to show that the theoretical gene level ratios have to be calculated by considering all functional genes in the organism, regardless of their activity in the specified condition considered (eg. different substrates).

The number of modes was calculated as 8726 and 1308 in glucose and ethanol media respectively. Since the number of available modes is considered to be a measure of the flexibility of the central metabolism on alternative carbon sources (Stelling *et al*., 2002) this finding indicates a better flexibility of the microorganism in the glucose medium as expected.

The same number of modes was determined in case 1.a and 3. The only difference between glucose system and galactose system in terms of the considered reactions is the replacement of hexokinase reaction by those catalyzed by GAL1, GAL7 and GAL5. Since these three enzymes constitute an enzyme subset, this replacement does not alter the

number of resultant elementary flux modes (case 1.a and case 3). Enzyme subsets are the enzymes which always operate together in fixed flux proportions at steady state (Pfeiffer *et al.*, 1999; Schuster *et al.*, 2002a).

Table 3.1. Different cases analyzed

	Substrate	Number of EFMs	Excluded Reactions
1.a	Glucose	8726	-ADH2 –GAL1,5,7
1.b	Glucose	199	-ADH2 –GAL1,5,7 -PCK1 -ALD4 -IDP2 -FBP1 -Glyoxylate Shunt
2.a	Ethanol	1308	-GLK1(HXK) -GAL1,5,7
2.b	Ethanol	641	-GLK1(HXK) -GAL1,5,7 -PYK -PFK
3	Galactose	8726	-ADH2

3.1.4. Deletion Phenotype Analysis

Since elementary modes indicate all possible routes that the microorganism can choose to maintain its growth, this information can be utilized in deletion phenotype analysis. If a gene is deleted, all the routes which include the corresponding reaction are disrupted. A gene is considered to be essential if it is involved in all EFMs enabling growth. That is, its deletion will destroy all routes leaving no way for the survival of the deletion mutant. In the present study, the EFM analysis was used to predict the deletion phenotypes and the results were compared with the experimentally determined lethal deletions reported in literature.

3.1.4.1. Glucose as Substrate. In glucose medium, 13 lethal *in silico* gene deletions (*acs, aco1, fba, pgi, rki, pgk, eno, gpm, tdh, pyk, pfk, gpd, pdc*) were identified by EFM analysis. EFM results indicate that the omission of the reaction catalyzed by PDC enzyme leaves no EFM out of 8726 (case 1.a) pointing to the lethality of this deletion. It was reported that PDC deficient cells could not grow on mineral medium when glucose was the only available substrate (Flikweert *et al.*, 1996). EFM analysis could also successfully predict the lethality of *acs* and *pfk* deletions (i.e. *acs1-acs2* and *pfk1-pfk2* double deletions). All elementary modes include these reactions in their routes assigning them as essential genes. It was reported that the *acs1-acs2* double deletion was not viable in media containing

glucose or ethanol (van den Berg and Steensma, 1995). Also, a glucose-negative phenotype was reported for *pfk1-pfk2* double deletions mutants (Avranitidis and Heinisch, 1994).

fba, pgi, rki, pgk, eno, gpm, tdh and *pyk* deletions were stated as lethal mutations in MIPS database (http://mips.gsf.de/genre/proj/yeast/index.jsp). *In silico* predictions for only 2 genes (*aco1, gpd*) were not in agreement with the information available in MIPS, in which these deletions were stated to be viable.

3.1.4.2. Ethanol as Substrate. *In silico* single gene deletion analysis resulted in the identification of 18 lethal cases (*acs, aco1, fba, pgi, rki, pgk, eno, gpm, tdh, fum, tpi, ald4, sdh, fbp, icl1, mls1, pck1, nadhx-fadhx*) in ethanol containing medium.

EFM analysis shows that none of the 1308 modes in case 2.a have zero flux through ALD4 enzyme. This result suggests that the deletion of *ald4* gene in this organism is lethal. *ald4* deletion in *S. cerevisiae* is reported to be detrimental for growth on ethanol (Tessier *et al.*, 1998). Similarly, the lethal effect of *mls1* gene deletions in *S. cerevisiae* (Kunze *et al.*, 2002) when grown in ethanol medium was also confirmed by EFM analysis. All of the 1308 EFMs in case 2.a house the enzymatic reaction catalyzed by the product of this gene. In the present EFM analysis, all of the calculated modes utilize oxygen. This is an expected result because growth on nonfermentable substrates is known to be respiratory. *fbp* and *pck1* deletions are also reported to be lethal in ethanol medium (Stückrath *et al.*, 2002), which is in agreement with here-made predictions.

The lethality of *in silico* deletions was also checked from the online database, MIPS. *acs, aco1, fba, tdh, fum, sdh* and *icl1* deletions are also given as lethal in MIPS database, in parallel with the EFM simulation. As for the remaining 6 cases, the *in silico* prediction cannot be checked for 4 mutations and the prediction was incorrect for 2 deletions (*pgk, pgi*).

In case of double deletions of pyruvate kinase and malic enzyme genes, zero biomass yields were obtained for all the pathways in EFM analysis. In *Saccharomyces cerevisiae*,

the strains lacking malic enzyme gene (*mae1*) and both genes of pyruvate kinase (*pyk1*, *pyk2*) were found to be unable to grow in ethanol (Boles *et al.*, 1999).

When elementary flux modes with maximum biomass yield are examined, all of them are found to have a zero flux through PDC enzyme. This suggests that the deletion of the corresponding gene does not alter the biomass yield of the strain. Flikweert *et al.*, (1996) showed that the biomass yields of *pdc* deficient and wild-type *S. cerevisiae* were identical in ethanol-limited chemostat cultures. Thus, the phenotype of a silent mutation could successfully be predicted by EFM analysis.

3.1.5. Biomass Yield and Robustness Analysis

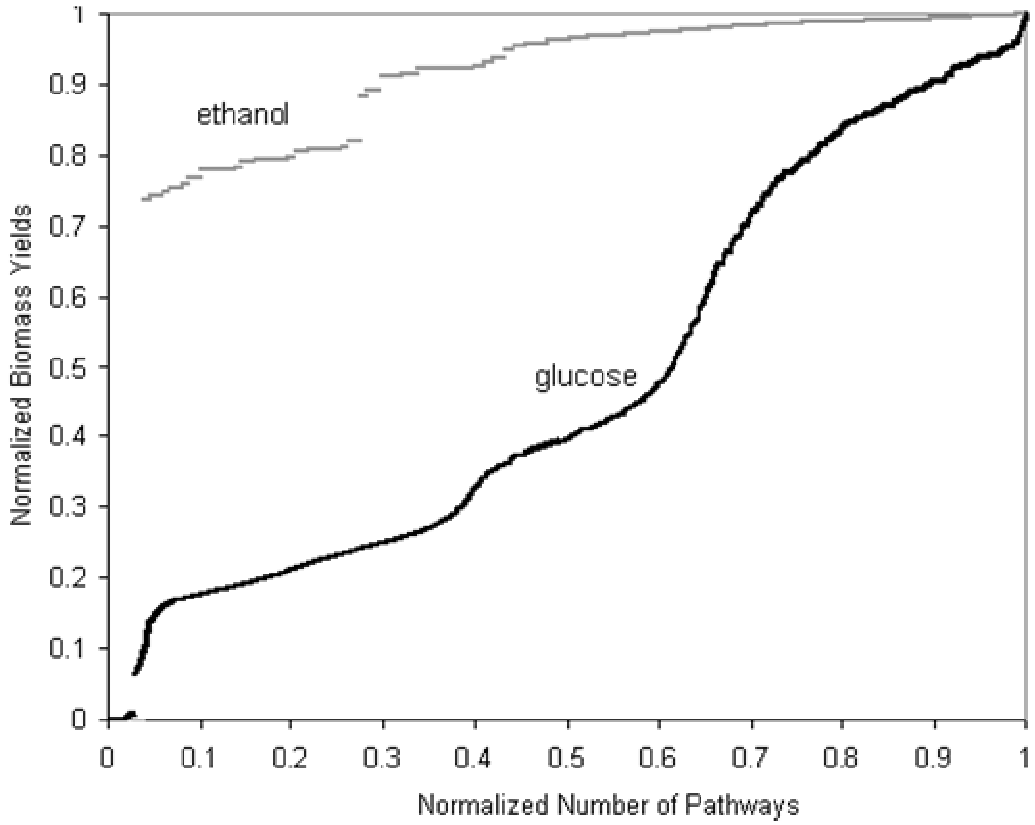
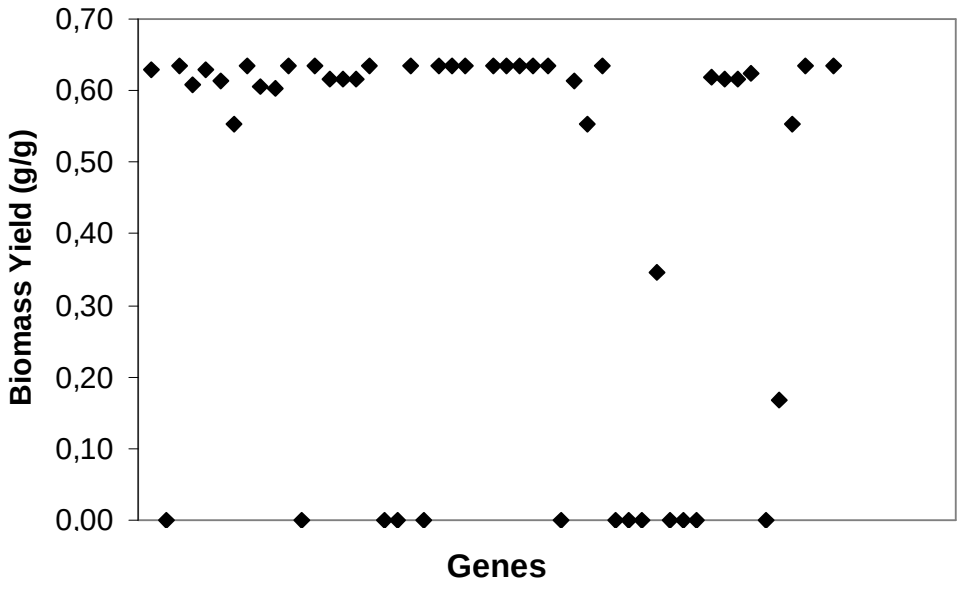


Figure 3.1. Biomass yield profiles of EFMs for growth on glucose and ethanol. The pathways are arranged in the order of increasing biomass yields. The axes are in normalized units

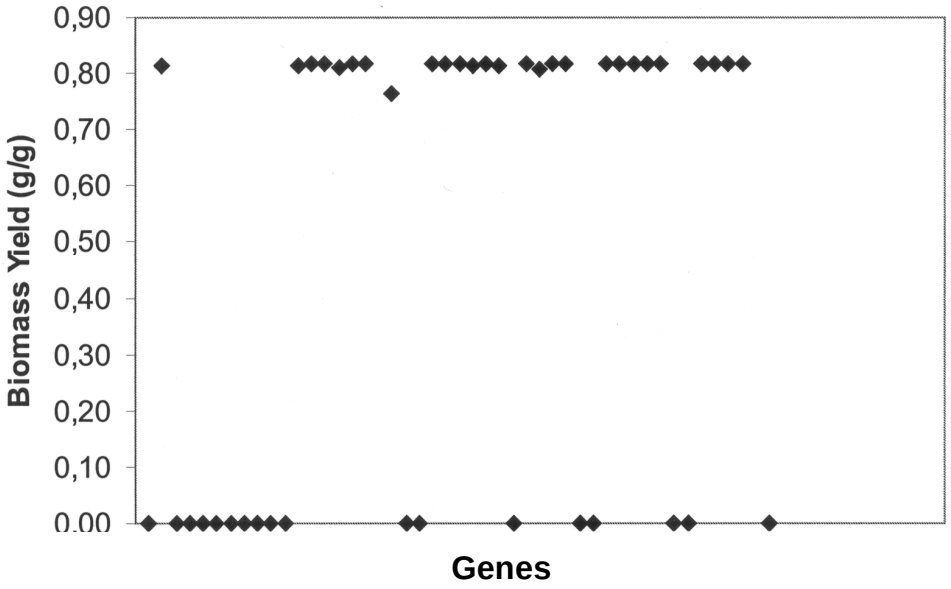
For each of the calculated elementary flux modes, a theoretical biomass yield was calculated by dividing the relative flux towards biomass production (r_{53}) to the relative flux of substrate uptake reaction (r_1 /r_3). Resultant yield profiles of elementary pathways for glucose and ethanol media are compared in Figure 3.1. The gradual change in biomass yield on ethanol, in the normalized pathway range of 0.3-1.0, indicates that 70 per cent of all pathways are in the range of high biomass yields (i.e. between 0.9-1.0). Such high yields could only be observed in 10 per cent of the pathways in glucose medium. 30 per cent of glucose pathways (in the normalized pathway range of 0.1-0.4) were found to result in normalized biomass yields of 0.20 ± 0.05. These results agree well with the fact that high biomass yields can be obtained during respiratory growth on nonfermentable substrates.

Biomass yield can be regarded as a measure of the robustness of microorganisms towards disturbances. It was found that although deletion mutants usually results in reduced growth rates compared to the wild type strain, the maximum biomass yields of viable deletions were almost the same with that of the wild type strain. Maximum biomass yields on glucose and ethanol were calculated using EFM analysis for each deletion mutant (Figures 3.2.a, 3.2.b). When the maximum biomass yields are analyzed for the deletion cases of each enzyme, yeast cells are found to have a robust central metabolism supporting the literature findings (Ebenhöh and Heinrich, 2003; Wagner, 2000). In Figures 3.2.a and 3.2.b, there are a number of *in silico* mutations having zero biomass yields. That is, the deletion of these genes was found to be lethal for the cell theoretically. There are 13 and 18 such deletion mutants having zero biomass for growth on glucose and ethanol respectively (details are given in previous section).

In glucose medium, the maximum biomass yields of 5 viable deletions (*tpi, nadhx, fadhx, fum1, sdh* deletions) are found to be less than that of the wild type strain (Figure 3.2.a). Among these, inhibition of oxidative phosphorylation reactions (NADHX, FADHX) results in damage in the respiration of the organism. Particularly inhibition of NADH oxidation renders the cell almost anaerobic with maximum biomass yield of 0.17 g/g (26.5 per cent of parental strain). In the case of *tpi* deletion, mutant strains can only have a maximum biomass yield of 0.35 g/g (54.7 per cent of parental strain).



(a)



(b)

Figure 3.2. Maximum biomass yields of mutant strains calculated from EFM analysis. Each point is for one *in silico* deletion strain (a) glucose is the substrate (b) ethanol is the substrate

In ethanol medium, the maximum biomass yields of viable deletions were almost the same with that of the wild type strain (Figure 3.2.b). The only noticeable exception for

ethanol growth was *ald6* deletion whose biomass yield was 93.6 per cent of the parental strain. The main cause in the decrease of biomass yield in *in silico ald6* mutants is the decrease in NADPH supply of metabolism. Since NADPH is required in biosynthetic reactions, its level highly controls biomass growth. Although there are other reactions supplying NADPH (IDP2, MAE1, ZWF1, GND), the carried control effective flux through these enzymes are noticeably lower than that of ALD6 being far from compensatory effect.

3.1.6. Theoretical Transcript Ratio Analysis

Calculated CEF ratios of each gene for ethanol and glucose media were presented in Table 3.2 and the corresponding transcript ratios for ethanol-glucose media were taken from DeRisi *et al.*, 1997. Of the 41 genes considered which have functions in the central metabolism of the yeast, calculated CEF ratios for 40 genes showed acceptable correlation ($R^2 = 0.65$) with the transcript data (Figure 3.3). In a similar study, calculated CEF ratios of 47 genes in *Escherichia coli* was also found to correlate with $R^2 = 0.60$ (Stelling *et al*., 2002). The gene *pdc* was identified as outlier as shown by dark points in Figure 3.3 and thus not included in the regression analysis. Figure 3.3 is on logarithmic scale with y axis showing CEF ratios and x axis showing Experimental data ratios. Dark point in the figure belongs to pdc, which is a statistical outlier and thus not included in the regression analysis.

Region I in Figure 3.3. shows the genes whose expression has increased when ethanol is the substrate (20 genes). The points in region III, on the other hand, correspond to the genes whose expression levels have decreased in ethanol medium (9 genes). Direction of the change in the gene expression levels of 5 genes in Region II and 7 genes in Region IV might be falsely predicted upon medium change. ‘False prediction’ within the text mainly refers to the fact that although the experimentally determined expression levels have been found to increase in the ethanol /glucose medium, estimated CEF ratios indicates a decrease in gene expression in ethanol /glucose medium. Most of these false points have the values close to 1 and they may possibly arise from experimental errors. A small modification in their values will shift these points to the reasonable regions (I and III). Derisi *et al.*, 1997, have reported a decrease in the expression level of *ald6* gene in ethanol. *ald6* gene was found to be activated during the diauxic shift together with the

other cat8 dependent genes (*fbp1, pck1, idp2, acs1, icl1, mls1, mdh2*) (Haurie *et al.*, 2001). Based on this fact, one may also expect an increase in the transcription of *ald6* in parallel to the increase in the activity of all cat8 dependent genes as predicted in this study. The activation of this gene upon medium change from galactose to ethanol was also reported (Griffin *et al.*, 2002). The false predictions may also arise from the usage of unweighted averages for some of the reactions.

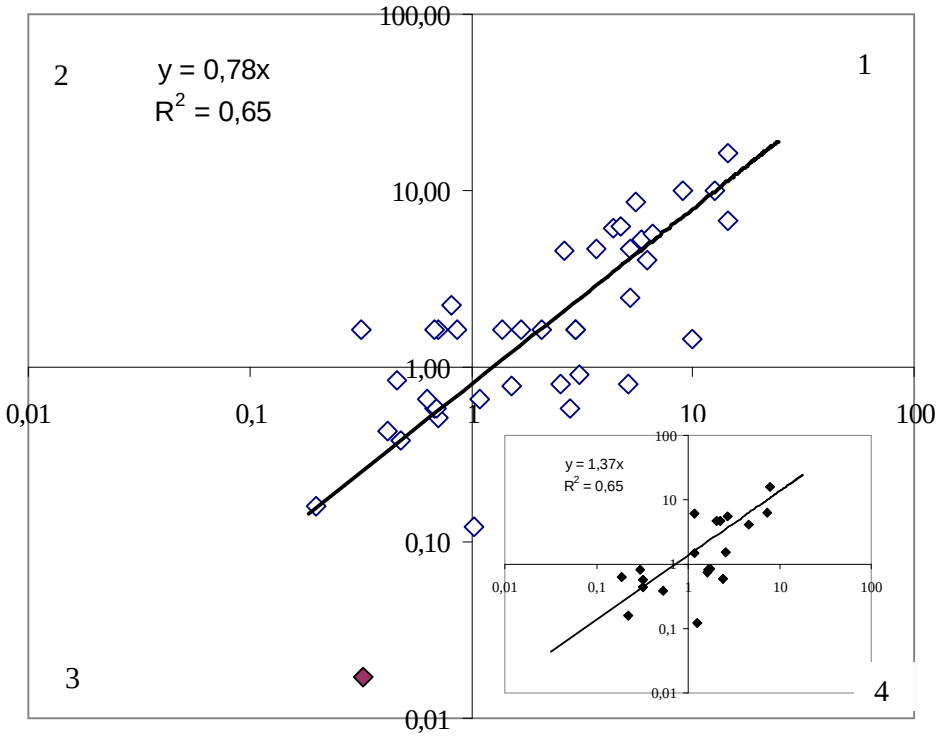


Figure 3.3. Experimental versus theoretical ratios of gene expression levels in glucose - ethanol media. The values are the ratios of ethanol medium to the glucose medium. The inset shows experimental versus theoretical ratios in galactose- ethanol media

The calculated CEF and experimental transcript ratios (Griffin *et al.*, 2002) of the genes in ethanol and galactose media are presented in Table 3.2. The only difference between glucose system and galactose system in terms of the considered reactions are the replacement of hexokinase reaction by those catalyzed by GAL1, GAL7 and GAL5, which constitute an enzyme subset. For the 20 genes considered which have functions in the central metabolism of the yeast, calculated CEF ratios showed acceptable correlation (R^2 = 0.66) with the transcript data (Figure 3.3, inset).

Table 3.2. Theoretical (CEF-based) and Experimental (mRNA-based) transcript ratios. The given values are the ratios of ethanol medium to the glucose (galactose) medium.

Gene	CEF ratio	DeRisi *et al*.	Griffin *et al*.	Gene	CEF ratio	DeRisi *et al*.	Griffin *et al*.
pgi1	0.51	0.71		*pyk**	0.16	0.20	0.22
fba1	0.44	0.42	0.32	*zwf1*	1.54	1.37	
tpi1	0.81	0.46	0.30	*sol**	1.54	2.10	
*tdh**	0.57	0.68	0.32	*gnd12**	1.54	1.67	
pgk1	0.57	0.70		*pdc**	0.02	0.33	
*gpm**	0.64	1.10		*ald6*	1.51	0.68	2.57
*eno**	0.64	0.63	0.19	*ald4*	8.62	5.56	
rki1	1.53	0.87		*acs**	5.58	6.55	2.71
rpe1	1.56	0.71		*pda**	0.12	1.05	1.27
*tkl**	1.54	2.94		*pyc**	0.85	3.08	1.73
tal1	1.54	0.32		*cit13**	2.47	5.25	
*tki**	1.54	2.94		*idp1*	0.75	1.52	1.62
aco1	4.05	6.25	4.57	*idh12**	0.59	2.78	2.40
idp2	1.48	10.00	1.17	*kgd12**	0.82	5.16	1.66
*lsc12**	0.82	2.53		*sdh**	4.72	5.21	2.24
fum1	4.72	3.70	2.04	*mae1*	2.10	0.81	
mdh1	5.27	5.88		*pck1*	15.64	14.29	7.76
*pfk12**	0.38	0.49	0.54	*cat2*	6.05	4.35	1.17
fbp1	6.36	14.29	7.24	*mls1*	9.90	9.09	
icl1	9.90	12.5		*cit2*	6.20	4.76	
mdh2	4.57	2.63					

* For the reactions which are governed by multiple genes, an unweighted average of expression ratios of responsible genes were used.

In conclusion, the genes whose theoretically calculated transcript ratios are widely different from that of the experimental ones mainly belong to the tricarboxylic acid cycle and pyruvate branch point. This may be due to the restricted reaction set which does not include the anabolic reactions such as fatty acid synthesis, amino acid synthesis and nucleotide synthesis. These reactions mainly utilize AKG, OAC, AcCoA, PEP, PYR, RL5P, E4P as precursors. Although these metabolites are included into the biomass equation with suitable stoichiometries in concordance with their utilization in biosynthesis reactions, this probably restricts the flexibility of the reactions that include these metabolites and thus affects the theoretically calculated CEF ratios. The main reason of

considering a restricted number of biochemical reactions is that the inclusion of further reactions leads to a combinatorial explosion in the number of EFMs (Förster *et al.*, 2002; Klamt and Stelling, 2002; Klamt *et al.*, 2003). On the other hand, reproducibility of microarray data in transcriptome analysis may also be the cause of false predictions and the experimental errors in this type of analysis could not be excluded. (Piper *et al.*, 2002).

Modification of the method used in the calculation of CEF ratios by omitting (i) the flux of ATP generation due to maintenance, (ii) the flux modes with zero biomass yield or (iii) using the number of nonzero fluxes instead of the sum of absolute fluxes in each mode, did not result in any further improvement.

3.1.7. CEF Analysis and Functionality

The new approach of control effective flux analysis permits to relate transcriptomics to fluxomics unlike other stoichiometric approaches (FBA, MFA). CEF is a sign of relative importance of reactions, and absolute CEF values are given in Figure 3.4. Importance of specific reactions under special conditions could be predicted using these absolute CEF values.

For example, CEF values of ALD4 and ALD6 are found to be higher in ethanol medium compared to those in glucose medium. ALD4 (YOR374W) is a mitochondrial NADH dependent enzyme, which is experimentally found to be active and necessary in ethanol medium (Boubekeur *et al.*,1999; Tessier *et al.*, 1998). The other isozyme, ALD6, is mainly functional in glucose medium by contributing to the NADPH levels in the cytosol (Meaden *et al.*, 1997) and is also active in ethanol medium (Haurie e*t al.*, 2001). Comparison of CEFs of these genes indicates that *ald4* gene is much more important in ethanol medium. Its CEF is 218.0 whereas that of *ald6* is 49.5 in ethanol medium (Figure 3.4).

A very low effective flux carried by PDC enzyme in ethanol medium is also in accordance with the literature (Flikweert *et al.*, 1996) since the mutants of this gene does not differ very much from the wild type strain in terms of growth rate phenotype.

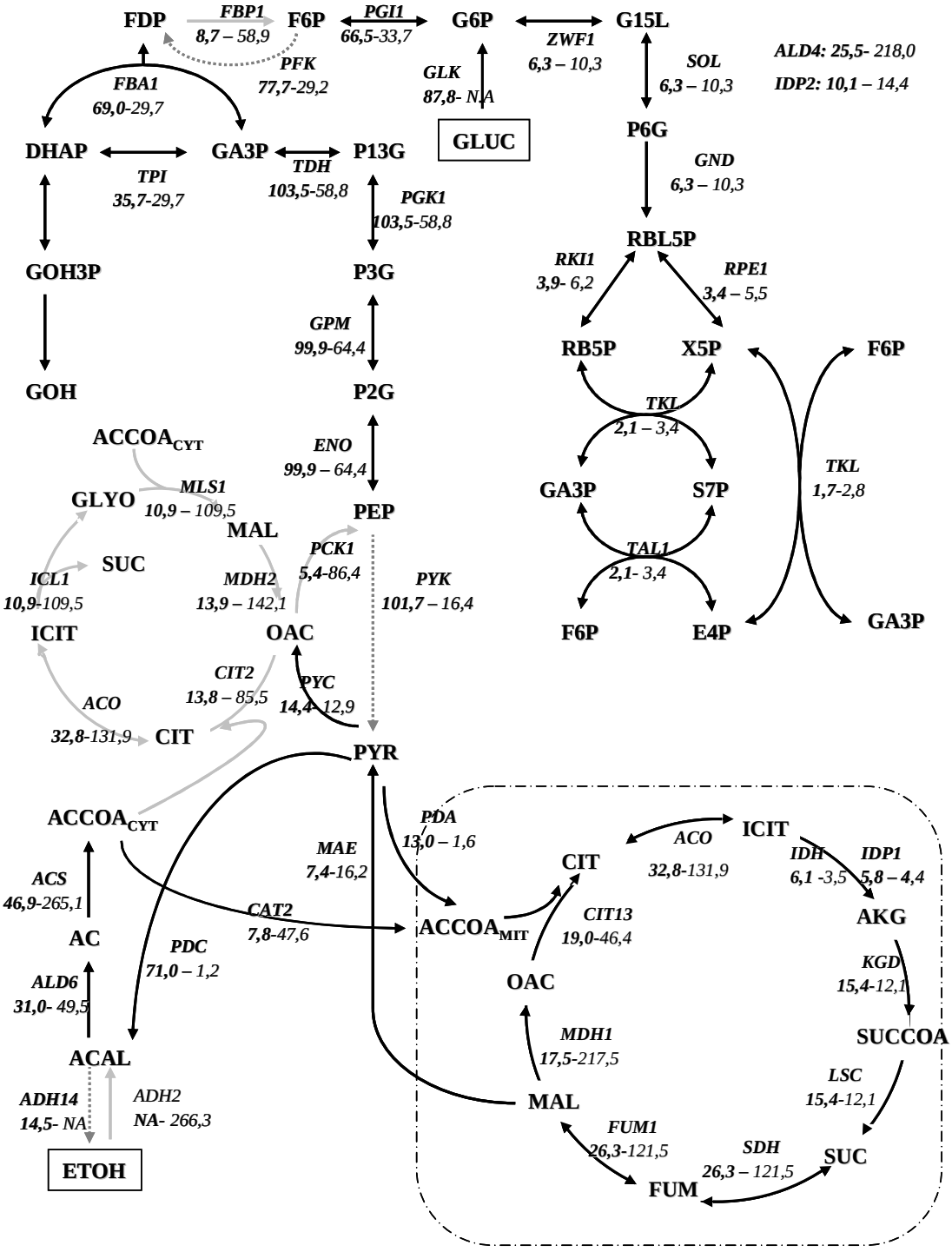


Figure 3.4. CEF distribution map for *S. cerevisiae* cells. First number indicates CEF for glucose growth, the second is CEF for ethanol growth. Gray reaction lines are only active in ethanol growth; dotted reaction lines are only active in glucose growth

By examining solely the outputs of CEF analysis, one can conclude that there is an increase in the sum of CEF of NADPH producing enzymes in ethanol medium compared to the glucose medium. This is observed for both cytosolic (ZWF1, GND, IDP2, ALD6) and mitochondrial (MAE1, IDP1) enzymes with corresponding ratios of (84.5/53.5) and (20.6/13.2). NADPH production in ethanol medium is important due to aerobic growth and leads to higher biomass production. CEF ratio of biomass production (r_{53}) was calculated as 1.56 (0.94/0.60) confirming this statement.

Effective fluxes carried by glycolysis enzymes (PYK and PFK) in ethanol medium were calculated to be lower than those in glucose medium as expected. Moreover, the flux of gluconeogenic counterpart of PFK, that is FBP1, is about two times higher in ethanol medium. Similarly, the gluconeogenic PCK1 enzyme has a flux considerably dominant to the flux of PYK in this medium (Figure 3.4). Thus, examination of CEF values can be used in predicting the relative importance of glycolytic and gluconeogenic enzymes in different media.

One should note that as the absolute values of CEFs will change with respect to the number of considered reactions, only CEF values calculated for the same network structure can be compared.

3.1.8. Concluding Remarks

A novel approach which links transcriptomics to fluxomics was applied to *Saccharomyces cerevisiae*. This approach was previously successfully applied to *Escherichia coli* (Stelling *et al*., 2002). The applicability of the method to *Saccharomyces cerevisiae*, which has different characteristics such as being a eukaryote and having cellular compartments, reinforces the universality of the approach.

The study reflects the power of CEF analysis over FBA and the importance of flexibility. When the most efficient EFMs are omitted from the system, the resultant control effective flux distribution remains almost unchanged. Hence, the flexibility seems to be more important than the efficient operation of the cell. There is no objective in the CEF analysis unlike FBA which can force the system to behave in a particular manner.

Same behaviour of the cell can be predicted by CEF analysis even if the most efficient behaviour of the cell (predicted by FBA analysis) is not considered. Metabolic reaction system is allowed to be flexible since it is free to choose all possible routes.

Although a very small number of reactions (a total of 53) are considered compared to the recently published set of metabolic reactions in *Saccharomyces cerevisiae*, (1175 reactions, Förster *et al.*, 2003a), the analysis presented in this study exhibits reasonable correlation with the experimental growth and transcriptome data. Analysis of genome-wide networks which will be facilitated by the development of novel computational tools and equipments, will give rise to construction of better models in the future.

3.2. Metabolic Pathway Analysis of Enzyme-Deficient Human Red Blood Cells

In this section, five enzymopathies (G6PDH, TPI, PGI, DPGM and PGK deficiencies) in the human red blood cells are investigated using a stoichiometric modeling approach, i.e., metabolic pathway analysis. The overall aim is to extend the approach detailed in section 3.1 to the human metabolism.

The primary physiological objective of the red blood cell is gas transport and exchange. Beyond this, it must perform several metabolic functions for its own survival. The major metabolic function of the erythrocyte is to produce the necessary cofactors (ATP, NADPH, and NADH) by energy and redox metabolisms for maintaining its osmotic balance and electroneutrality and fighting oxidative stresses (Bossi and Giardina, 1996; Joshi and Palsson, 1989; Wiback and Palsson, 2002). These cofactors are also necessary for the bioconcave shape of the cell as well as for the specific intracellular cation concentrations.

Enzymopathy can be described as a biochemical disorder in which a specific enzyme undergoes alterations in its activity that may have pathological consequences. For the erythrocytes, deficiencies of about 20 enzymes, associated with widely different degrees of severity and complexity, have been identified so far (Jacobasch and Rapoport, 1996; Jacobasch, 2000; Schuster and Holzhutter, 1995). Since enzymopathies in human red blood

cells can cause serious diseases including anemia, the analyses of erythroenzymopathies are of basic importance in handling with such diseases.

Modelling studies on erythroenzymopathies and erythrocyte metabolism have been mainly by kinetic models (Holzhütter *et al.*, 1985; Joshi and Palsson, 1989; Martinov *et al.*, 2000; Mulquiney and Kuchel, 1999; Schuster and Holzhütter, 1995; Schuster *et al.*, 1989), which are based on detailed information on kinetic properties of the enzymes. Few attempts have been made to model red blood cells stoichiometrically (Schuster *et al.*, 1998; Price *et al.*, 2003; Wiback and Palsson, 2002), which has the advantage of relying solely on the stoichiometries of the considered reactions.

In a metabolic network consisting of cellular reactions, the analysis of the fluxes allows one to establish a relationship between cell genotype and phenotype. One of the main approaches for the flux analyses of metabolic networks is Metabolic Pathway Analysis (MPA) (Papin *et al.*, 2003; Schilling *et al.*, 1999), which is used to define the structure of the metabolic network and the overall metabolic capabilities of the microorganism. The method only uses information about the stoichiometry and the reversibility or irreversibility of reactions. Given the enzymatic reactions occurring in a particular microorganism, all possible routes are determined and analyzed.

An important tool used in MPA is the detection of elementary flux modes (EFMs). An EFM is a minimal set of enzymes that could operate at steady state, with the enzymes weighted by the relative flux they need to carry for the mode to function (Schuster *et al.*, 1999; Schuster *et al.*, 2000; Schuster *et al.*, 2002a). EFM analysis allows the discovery and analysis of meaningful routes in metabolic networks. Control-effective flux (CEF) analysis is another tool in assessing a metabolism (Stelling *et al.*, 2002). The CEFs, which are directly determined from the set of EFMs, represent the importance of each reaction of a metabolism for efficient and flexible operation of the entire metabolic network. Thereby, regulatory events of metabolism are implicitly incorporated. The method was successfully applied to *E. coli* and *S. cerevisiae* for the functionality analysis based on the theoretical estimation of gene expression changes (Section 3.1; Çakır *et al.*, 2004; Stelling *et al.*, 2002).

In the present section, elementary flux mode (EFM) detection was performed for the erythrocyte metabolic network via the simulation tools, METATOOL 4.3 (Dandekar *et al.*, 2003; Pfeiffer *et al.*, 1999) and FluxAnalyzer 4.0 (Klamt *et al.*, 2003). The following enzymes, which are clinically essential, were taken as model systems in the metabolic investigation of enzyme deficiencies of erythrocytes; glucose-6-phosphate dehydrogenase (G6PD), triosephosphate isomerase (TPI), phosphoglucose isomerase (PGI), disphosphoglycerate mutase (DPGM), and phosphoglycerate kinase (PGK). CEF analysis was performed to analyze EFM results quantitatively and to make comparisons for the activities of pathways upon different degrees of deficiencies. When available, experimental findings reported in literature related to metabolic behavior of the human red blood cells were compared with the results of EFM and CEF analyses. This study differs from the previous models in that regulatory events are also accounted by the help of CEF analysis, with an emphasis on enzyme deficiencies. The ultimate goal herein is the target identification for drug design for the treatment of patients with enzymopathies.

3.2.1. Metabolism of the Human Red Blood Cell

Main reactions in human red blood cells are depicted in Figure 3.5 and given in Table 3.3 (Joshi and Palsson, 1989, Mulquiney and Kuchel, 1999; Schuster *et al.*, 1998; Wiback and Palsson, 2002). There are 39 reactions and 44 metabolites in the network.

Red cells lack nuclei and other intracellular organelles; they are incapable of protein and lipid synthesis and of oxidative phosphorylation (Baynes and Dominiczak, 1999; Bossi and Giardina, 1996). Unlike most metabolic networks, the red cell does not generate biomass (Wiback and Palsson, 2002). The main red cell energy source is glucose that is metabolized through the Embden-Meyerhof pathway (r_1-r_{11}) with the production of 2 moles of ATP and lactate as end products per mol of glucose. The theoretical net gain of 2 mol of ATP for every 1 mol of glucose metabolized through anaerobic glycolysis is modified by the Rapoport-Luebering shunt (r_{12}, r_{13}) which is controlled by diphosphoglycerate mutase (DPGM) and generates 2,3-diphosphoglycerate (D23PG) (Bossi and Giardina, 1996; Monsen and Vestergaard-Bogind, 1978). The pentose phosphate shunt (r_{14}-r_{21}) contributes to the redox status of the cell by producing 2 moles of NADPH per mol of glucose entering the cycle.

Table 3.3. Reactions in red blood cell metabolism and corresponding catalyzing enzymes[a]

	Reactions	Enzymes
	1. Glycolysis	
(1)	1 GLC + 1 ATP $\rightarrow$ 1 G6P + 1 ADP	*HK*
(2)	1 G6P $\leftrightarrow$ 1 F6P	*PGI*
(3)	1 F6P + 1 ATP $\rightarrow$ 1 FDP + 1 ADP	*PFK*
(4)	1 FDP $\leftrightarrow$ 1 GA3P + 1 DHAP	*ALD*
(5)	1 DHAP $\leftrightarrow$ 1 GA3P	*TPI*
(6)	1 GA3P + 1 NAD $\leftrightarrow$ 1 D13PG + 1 NADH	*GAPDH*
(7)	1 D13PG + ADP $\leftrightarrow$ 1 P3G + 1 ATP	*PGK*
(8)	1 P3G $\leftrightarrow$ 1 P2G	*PGM*
(9)	1 P2G $\leftrightarrow$ 1 PEP	*EN*
(10)	1 PEP + 1 ADP $\rightarrow$ 1 PYR + 1 ATP	*PK*
(11)	1 PYR + 1 NADH $\rightarrow$ 1 LAC + 1 NAD	*LDH*
	2. Rapoport-Luebering Shunt	
(12)	1 D13PG $\rightarrow$ 1 D23PG	*DPGM*
(13)	1 D23PG $\rightarrow$ 1 P3G	*DPGase, (DPGM)*
	3. Pentose Phosphate Pathway	
(14)	1 G6P + 1 NADP $\rightarrow$ 1 GL6P + 1 NADPH	*G6PDH*
(15)	1 GL6P $\leftrightarrow$ 1 GO6P	*PGLase*
(16)	1 GO6P + 1 NADP $\rightarrow$ 1 RL5P + 1 NADPH + 1 CO_2	*GL6PDH*
(17)	1 RL5P $\leftrightarrow$ 1 XYL5P	*XPI*
(18)	1 RL5P $\leftrightarrow$ 1 R5P	*RPI*
(19)	1 R5P + 1 XYL5P $\leftrightarrow$ 1 SED7P + 1 GA3P	*TK*
(20)	1 SED7P + 1 GA3P $\leftrightarrow$ 1 F6P + 1 ERY4P	*TA*
(21)	1 XYL5P + 1 ERY4P $\leftrightarrow$ 1 F6P + 1 GA3P	*TK-I*
	4. Nucleotide Metabolism	
(22)	1 R5P + 1 ATP $\rightarrow$ 1 PRPP + 1 AMP	*PRPPsyn*
(23)	1 R1P $\leftrightarrow$ 1 R5P	*PRM*
(24)	1 PRPP + 1 ADE $\rightarrow$ 1 AMP	*ADPRT*
(25)	1 INO $\leftrightarrow$ 1 HYPX + R1P	*PNPase*
(26)	1 HYPX + 1 PRPP $\rightarrow$ 1 IMP	*HGPRT*

Table 3.3. continued

(27)	1 IMP $\rightarrow$ 1 INO	*IMPase*
(28)	1 AMP $\rightarrow$ 1 ADO	*AMPase*
(29)	1 AMP $\rightarrow$ 1 IMP	*AMPDA*
(30)	1 ADO $\rightarrow$ 1 INO	*ADA*
(31)	1 ADO + 1 ATP $\rightarrow$ 1 ADP + 1 AMP	*AK*
(32)	2 ADP $\leftrightarrow$ 1 ATP + 1 AMP	*AdylK*
	5. Cellular Functions	
(33)	1 GSSG + 1 NADPH $\leftrightarrow$ 2 GSH + 1 NADP	*GSSG-R*
(34)	2 GSH + 1 H_2O_2 $\rightarrow$ 1 GSSG	*GSHpox*
(35)	1 ATP + 3 Na^+ + 2 K_{ext}^+ $\rightarrow$ 1 ADP + 3 Na_{ext}^+ + 2 K^+	*Na-K-ATPase*
(36)	ATP $\rightarrow$ ADP	*MemPhos*
(37)	MetHb + NADH $\rightarrow$ Hb + NAD	*MetHbRed*
(38)	D23PG + Hb $\rightarrow$ D23PG:Hb	*D23PGdrain*
(39)	2 GSH + 1 O_2 $\rightarrow$ 1 GSSG + 1 H_2O_2	*GSHox*

[a]Arrow type, $\rightarrow$ or $\leftrightarrow$, indicates whether the reaction is considered irreversible or reversible in the analysis.

The red cell requires energy, (i) to replenish its adenine nucleotide pool using salvage pathways (r_{22-32}), (ii) to protect the cell against oxidative stress (r_{33}, r_{34}), (iii) to control its volume through membrane Na-K ATPase (cation pump) (r_{35}), (iv) to maintain the plasticity of its membrane (r_{36}), (v) to prevent the accumulation of methemoglobin (r_{37}), and (vi) to modulate oxyhemoglobin (r_{38}) (Bossi and Giardina, 1996; Schilling and Palsson, 1998).

3.2.2. Analyzed Enzyme Deficiencies

Glucose-6-Phosphate Dehydrogenase (G6PD) enzyme fuels the first reaction of pentose phosphate pathway (PPP) in which G6P is oxidized and NADP is reduced resulting in NADPH production (r_{14}). The biological functions of the PPP in the human red blood cell are to synthesize ribose as sugar components of the nucleotides and to maintain the continuous supply of NADPH as an obligatory substrate for the glutathione system

protecting the cell against oxidative stress (Jacobasch and Rapoport, 1996). Normal red cells can defend themselves to a considerable extent against such changes by reducing GSSG to GSH through the glutathione reductase reaction (r_{33}). This requires a source of NADPH. As G6PD-deficient red cells are unable to reduce $NADP^+$ to NADPH at a normal rate, they are unable to remove hydrogen peroxide through peroxidase activity (r_{34}) (Bronk, 1999), which requires reduced glutathione (GSH) to protect the cells against oxidative damage (Bossi and Giardina, 1996). Namely, G6PD enzyme is indispensable to protect cells against even mild oxidative stress. G6PD defects belong to the most widespread enzymopathies in man. It is estimated to affect 400 million people worldwide (Miwa, 1996; Weatherall, 2000).

Triose Phosphate Isomerase (TPI) catalyzes the interconversion of dihydroxyacetone phosphate (DHAP) and glyceraldehyde-3-phosphate (GA3P) (r_5) and plays an important role in several crucial metabolic pathways. The metabolic pattern of TPI deficient erythrocytes is characterized by high levels of DHAP and a relatively minute decrease of ATP. DHAP accumulation has been reported to be toxic for cellular functions and responsible for the severity of TPI enzymopathies but the mechanism of DHAP toxicity is not well understood (Jacobasch and Rapoport, 1996, Orosz *et al*., 1996; Schneider, 2000). The defect leads to hemolytic anemia coupled with neurological dysfunction (Olah *et al*., 2002).

The enzyme phosphoglucose isomerase (PGI) catalyzes the reversible isomerization from G6P to F6P, an equilibrium reaction of glycolysis (r_2). Glucose turnover reacts, therefore, only on deficiency below a very low critical residual activity of PGI but then with a drastic decline of lactate formation, i.e., decrease in glycolytic flux (Jacobasch and Rapoport, 1996). The consequence of a limitation by the PGI reaction is an increase of the G6P level which causes a feedback inhibition of HK resulting both in a lower rate of glycolysis and increased PPP activity associated, in turn, with the recombination of F6P formed in PPP with glycolytic pathway (Kanno *et al*., 1998). With the effect of HK inhibition, ATP, D23PG and GSH regeneration decreases (Jacobasch and Rapoport, 1996). This is the third most common enzymopathy in the world (Kanno *et al*., 1998).

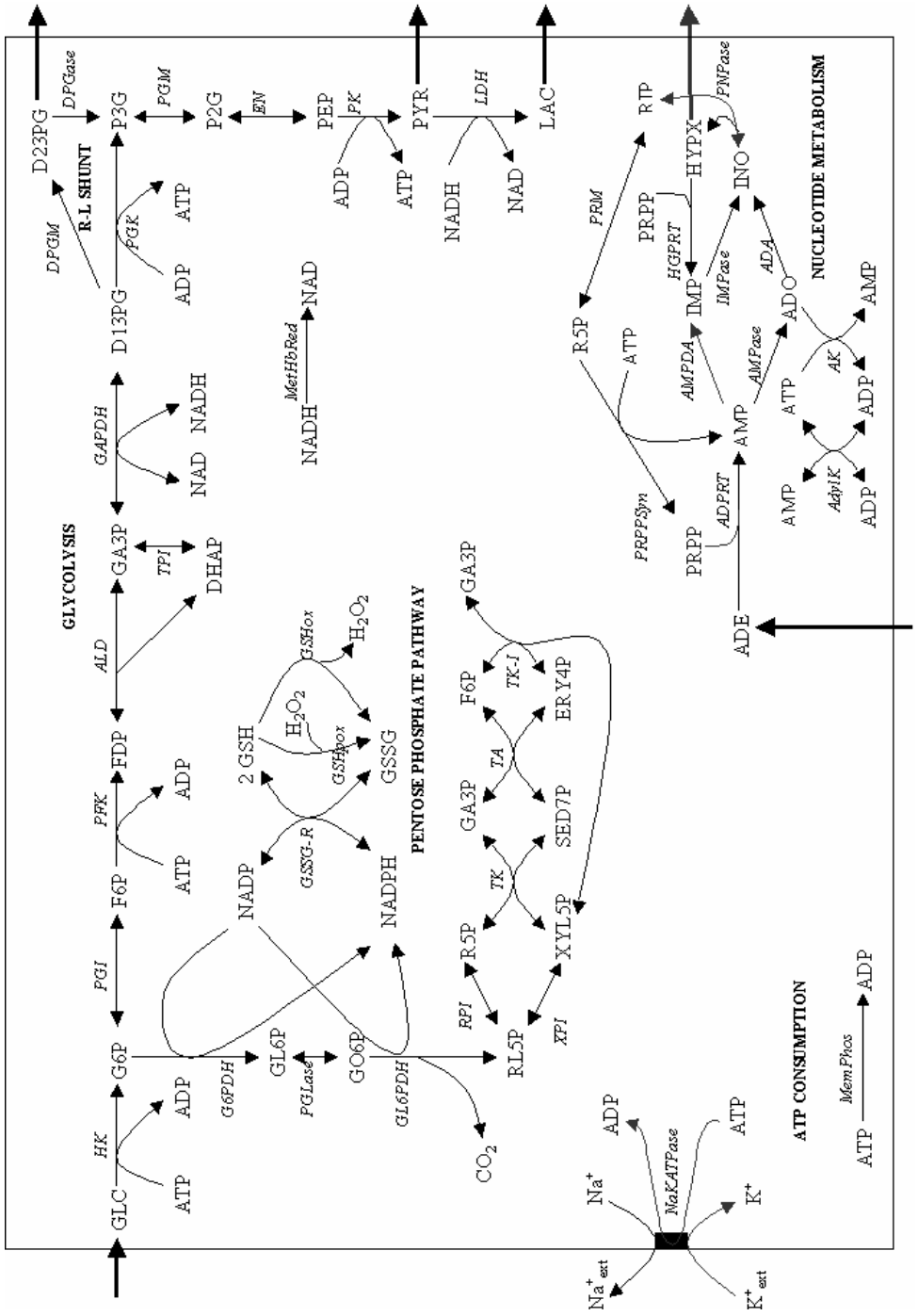


Figure 3.5. Human red blood cell metabolism. Adapted from Schuster *et al.*, 1998. Metabolites considered as external in EFM calculation are GLC, LAC, PYR, HYPX, ADE, CO_2, D23PG, K^+_{ext} and Na^+_{ext}

Disphosphoglyceromutase (DPGM) is a multifunctional enzyme which catalyzes both the synthesis and dephosphorylation of D23PG in human red blood cells (r_{12}) (Jacobasch and Rapoport, 1996; Mulquiney and Kuchel, 1999). With lowering of DPGM, the turnover via D23PG declines in favor of substrate phosphorylation catalyzed by PGK and PK leading to changes of the metabolic pattern. ATP, FDP, triose phosphates, 3PG, 2PG, PEP are enhanced, ADP, D23PG, F6P, G6P are diminished (Jacobasch and Rapoport, 1996).

Phosphoglycerate Kinase (PGK) is a key enzyme for ATP generation in the glycolytic pathway, catalyzing the conversion of D13PG (1,3-diphosphoglycerate) to 3PG (3-phosphoglycerate) (r_7) bypassing the Rapoport-Luebering shunt. A significant accumulation of D23PG, and a decreased concentration of ATP were observed in patients with PGK deficiency (Fujii and Miwa, 2000). Also, diminished glucose consumption was reported (Jacobasch, 2000).

3.2.3. Computational Methods

The same simulation tools as discussed in previous section (3.1.2) were used in the EFM analysis of red blood cell metabolism. CEF calculations were made using Equations 3.1 and 3.2. In efficiency calculations (equation 3.1), *GSSGR, GSHpox, MemPhos, NaKATPase, MetHbRed* and *23DPGdrain* (r_{33}- r_{38}) were taken as the mode outputs since these reactions are basis for the cellular objectives, i.e., main functions of erythrocytes. *HK* reaction (r_1), which consumes glucose, was taken as substrate uptake reaction.

The ratio of CEFs of reactions at two different conditions (healthy vs. enzyme-deficient) was used to predict the efficiencies of the reactions in human red blood cell metabolism.

For the analysis of enzyme deficiencies in terms of the degree of deficiency, a modification was made on the formulation of CEF calculation. The fluxes of the reactions of EFMs that included the deficient enzyme in their routes were multiplied by a constant, d_j, representing the degree of enzyme deficiency. d_j takes values between 1 (healthy case) and 0 (complete deficiency). The efficiencies of these modes were also weighted by d_j.

Some of the cellular objectives disappear upon complete deficiency (eg. NaKATPase, MemPhos and D23PGdrain reactions were not functioning in complete PGK deficiency). To account for the decreased contribution of these objectives on CEFs, the related CEF terms, $v_{i,CELLOBJ}$ were weighted by d_j. Otherwise, the analysis led to incorrect results in the boundary of complete deficiency since the ratio of very small numbers resulted in values which had significant effects on the calculated CEFs. For G6PD and TPI deficiencies, where R5P and DHAP were considered as external metabolite respectively, the modes including these metabolites in their routes as external were weighted by $(1 - d_j)$. This implies that these modes vanish for health case and gradually gain importance as the degree of deficiency increases.

3.2.4. Analyses of EFMs, CEFs, and *in silico* Deficiency Profiles

Biochemical pathways in the human red blood cell are investigated as a model for the ultimate goal of target identification in drug design. Possible enzyme deficiencies inherited in the red blood cell are analyzed on the basis of elementary flux mode. The effect of enzymopathies on the behaviour of metabolic network and on the regulatory events is studied using CEF concept, which allows the quantification of EFM analysis. The number of EFMs for each enzymopathy is given in Table 3.4 where non-functioning (disrupted) EFMs and indispensable enzymes (the enzymes whose absence lead to cessation of all the routes) obtained through the frequency analysis by Flux Analyzer are also indicated. Table 3.5 details the calculated CEFs of human red blood cell metabolism, in cases of non-deficient enzyme metabolism and of the complete deficiencies of clinically essential enzymopathies. The *in silico* deficiency profile for each enzymopathy is also obtained for the whole range of 0 per cent -100 per cent deficiency.

3.2.4.1. Non-deficient Case. For the reactions of erythrocyte metabolism given in Table 3.3, the corresponding number of elementary modes is calculated to be 48 (Table 3.4). Each elementary mode represents a biochemical function (Schuster *et al*., 1998; Schuster *et al*., 2002c; Wiback and Palsson, 2002), as indicated in the table. There are 18 different overall stoichiometries. Half of the 48 EFMs are associated with lactate production, and the other half differs only in terms of the end product, which is pyruvate.

Table 3.4. Number of EFMs obtained for each enzymopathy

Enzymopathy	No.of EFMs	Disrupted EFMs	Indispensable Enzymes
None	48	-	HK, GAPDH, PGM, EN, PK
G6PD	16 (18)[b]	6-15, 21-30, 32, 33, 35, 36, 39-42, 45-48	HK, GAPDH, PGM, EN, PK, ALD, PFK, PGI, TPI
TPI	16 (34)[b]	1-5, 11-20, 26-31, 33, 34, 36-38, 41-44, 47, 48	HK, GAPDH, PGM, EN, PK, PPP enzymes
PGI	16	1-10, 16-25, 31, 32, 34, 35, 37-40, 43-46	HK, GAPDH, PGM, EN, PK, TPI, ALD, PFK, PPP Enzymes
DPGM	36	1, 6, 11, 16, 21, 26, 31-33, 34-36	HK, GAPDH, PGM, EN, PK
PGK	6	2-5, 7-10, 12-15, 17-20, 22-25, 27-48	HK, GAPDH, PGM, EN, PK, DPGM, DPGase
EN	0	All	All
GAPDH	0	All	All
HK	0	All	All
PGM	0	All	All
PK	0	All	All
LDH	24	1-15, 31-33, 37-42	HK, GAPDH, PGM, EN, PK
ALD[a]	16	1-5, 11-20, 26-31, 33, 34, 36-38, 41-44, 47, 48	HK, GAPDH, PGM, EN, PK, PGI, PPP Enzymes
PFK[a]	16	1-5, 11-20, 26-31, 33, 34, 36-38, 41-44, 47, 48	HK, GAPDH, PGM, EN, PK, PGI, PPP Enzymes

[a]ALD and PFK are enzyme subsets and hence have the same combination of hampered EFMs.
[b]Numbers in paranthesis indicate the number of EFMs when R5P or DHAP are considered as external in the corresponding enzymopathies.

In CEF analysis the "no deficiency" case points out that GSSGR reaction has the highest flux rate as shown in Table 3.5, which basically implies the importance of NADPH production in the human red blood cell metabolism. This finding is in parallel with the known fact that the production of glutathione is essential in this cell type since it reduces $NADP^+$ in order to prevent the cell from oxidative damage (Baynes and Dominiczak, 1999; Bronk, 1999).

Table 3.5. CEF analysis for complete deficiency. The first column gives the absolute values, whereas the others are the CEF ratios of deficiency case to the healthy one, indicating whether the CEF of corresponding reaction increased or decreased upon deficiency

	Enzyme	No Def.	G6PD Def.	TPI Def.	PGI Def.	DPGM Def.	PGK Def.
GLYCOLYTIC PATHWAY	HK	**3.00**	0.83	1.83	1.50	0.67	0.33
	PGI	**2.65**	0.94	2.34	0.00	0.71	0.40
	PFK	**1.97**	1.26	0.85	1.52	0.61	0.24
	ALD	**1.97**	1.26	0.85	1.52	0.61	0.24
	TPI	**1.97**	1.26	0.00	1.52	0.61	0.24
	GAPDH	**4.97**	1.00	1.11	1.51	0.64	0.30
	PGK	**3.75**	0.99	1.22	1.53	0.85	0.00
	PGM	**3.94**	0.97	1.19	1.57	0.81	0.38
	EN	**3.94**	0.97	1.19	1.57	0.81	0.38
	PK	**3.94**	0.97	1.19	1.57	0.81	0.38
	LDH	**1.62**	0.98	1.18	1.66	0.72	0.19
RL SHUNT	DPGM	**1.22**	1.01	0.76	1.46	0.00	1.22
	DPGase	**0.19**	0.61	0.58	2.44	0.00	7.86
PENTOSE PHOSPHATE PATHWAY	G6PD	**3.09**	0.00	3.71	1.46	0.78	0.50
	PGLase	**3.09**	0.00	3.71	1.46	0.78	0.50
	GL6PDH	**3.09**	0.00	3.71	1.46	0.78	0.50
	R5PI	**1.03**	0.02	3.71	1.46	0.78	0.50
	Xu5PE	**2.06**	0.01	3.71	1.46	0.78	0.50
	TKI	**1.03**	0.01	3.71	1.46	0.78	0.50
	TA	**1.03**	0.01	3.71	1.46	0.78	0.50
	TKII	**1.03**	0.01	3.71	1.46	0.78	0.50
NUCLEOTIDE PATHWAY	PRPPsyn	**0.24**	0.57	0.64	2.43	1.48	0.00
	PRM	**0.24**	0.57	0.64	2.43	1.48	0.00
	AdPRT	**0.16**	0.57	0.64	2.42	1.48	0.00
	PNPase	**0.24**	0.57	0.64	2.43	1.48	0.00
	HGPRT	**0.08**	0.58	0.62	2.43	1.48	0.00
	IMPase	**0.16**	0.58	0.63	2.43	1.48	0.00

Table 3.5. continued

	AMPase	**0.26**	0.59	0.61	2.43	1.48	0.00
NUCLEOTIDE	AMPDA	**0.08**	0.57	0.64	2.42	1.48	0.00
PATHWAY	ADA	**0.08**	0.57	0.64	2.42	1.48	0.00
	AK	**0.18**	0.60	0.60	2.44	1.48	0.00
	ApK	**0.24**	0.57	0.64	2.43	1.48	0.00
	GSSGR	**6.19**	0.00	3.71	1.46	0.78	0.50
CELLULAR	GSHox	**3.09**	0.00	3.71	1.46	0.78	0.50
OBJECTIVE	NaKATPase	**1.00**	1.08	0.83	1.35	1.07	0.00
	MemPhos	**1.06**	1.05	0.80	1.38	1.09	0.00
	MetHbRed	**3.35**	1.00	1.07	1.44	0.61	0.35
	D23PGdrain	**1.03**	1.08	0.79	1.27	0.00	0.00
	GSHpox	**3.09**	0.00	3.71	1.46	0.78	0.50

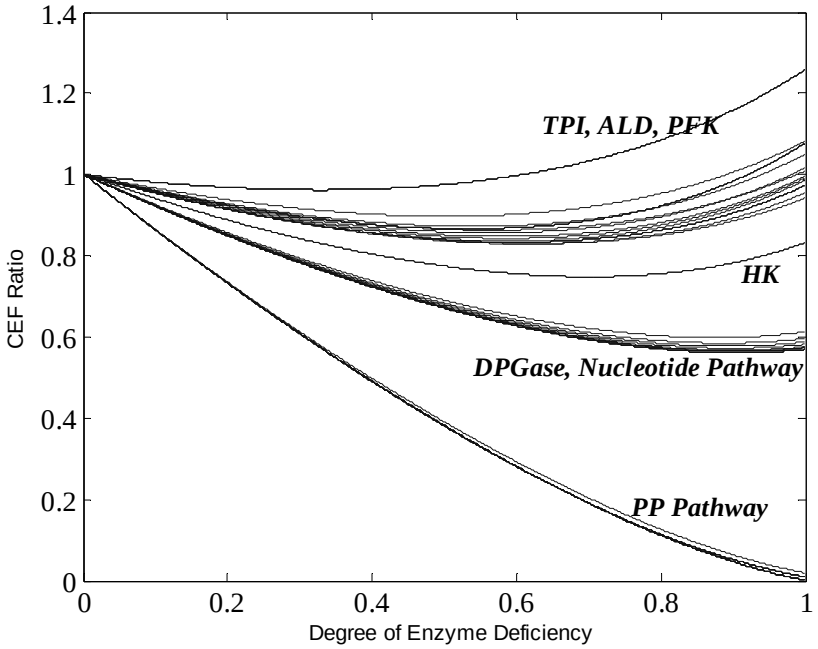
Nucleotide metabolism enzymes exhibit very low CEF values implying relative insignificancy of the pathway in the overall erythrocyte metabolism. Since the turnover of nucleotide pathway is reported to be very low (Ataullakhanov *et al*., 1996), this pathway is usually not considered in modeling analysis of erythrocytes (Mulquiney and Kuchel, 1999; Schuster and Holzhütter, 1995), in agreement with the low CEFs calculated here.

3.2.4.2. Glucose-6-phosphate dehydrogenase (G6PD) Deficiency. The experimental finding that in the absence of G6PD, there is still a pathway leading from glucose to ribose via TK and TA reactions (Pandolfi *et al*., 1995) was already verified using EFM analysis by blocking G6PD and considering ribose-5-phosphate (R5P) as external (Schuster *et al*., 1998). A substance is called external if it can be considered to be present in large excess so that its concentration is unaffected by the reactions under study (Schuster *et al*., 2002c). Based on this analysis, R5P is considered to be an external metabolite in the present deficiency analysis, too. The corresponding number of elementary flux modes is calculated to be 18. Common characteristics of these modes are that they do not involve the glutathione disulfide reductase and glutathione peroxidase reactions, which are necessary for preventing oxidative stress (Schuster *et al*., 1998; Mehta *et al*., 2000) and needed for cellular functions.

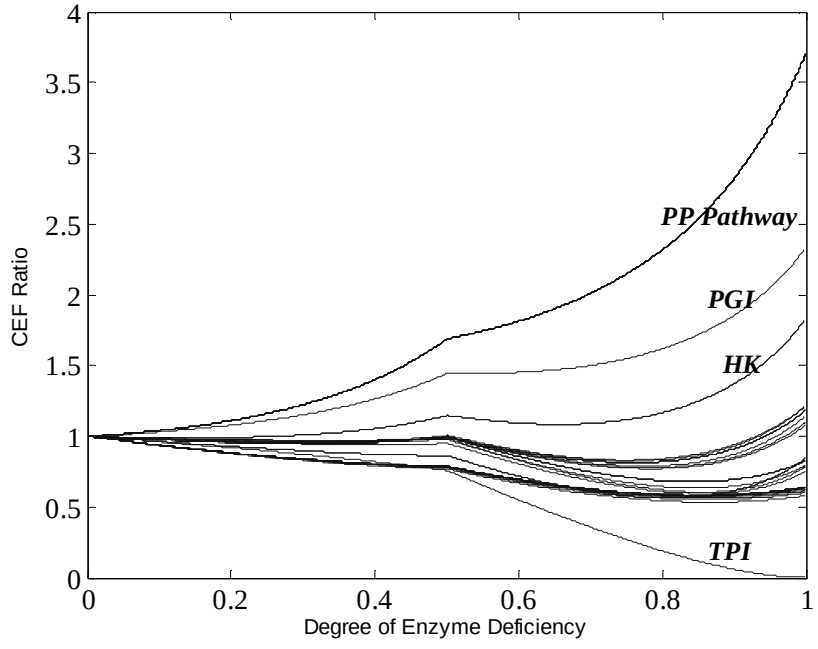
In normal erythrocytes, any oxidant that lowers available NADPH raises immediately the oxidative pentose phosphate pathway (r_{14}-r_{16}) rate several folds in order to supply the needed amount of this cofactor. On the other hand, red blood cells with G6PD deficiency cannot increase their OPPP rate sufficiently during the oxidative load (Bossi and Giardina, 1996) as predicted by CEF analysis results (Table 3.5). Zero fluxes were calculated for the oxidative PPP, and very low values were obtained for the nonoxidative part. The deficiency of G6PD blocks the PPP and renders those enzymes inefficient. On the other hand, the other objectives of the cell, i.e., ATP generation for the functioning of cation pump and for the membrane plasticity, D23PG production for the modulation of hemoglobin oxygenation, reduction of methemoglobin to functional form, were found to be unaffected.

Deficiency profile of this enzymopathy (Figure 3.6.a) indicates that the CEF values of metabolic reactions do not increase more than 1.3 fold, and nucleotide pathway is mildly affected. The only abrupt change occurs in PPP enzymes, and hence G6PD enzyme, which exhibit almost linear decrease upon enzyme deficiency. Deficiency profiles of the enzymes in the same enzyme subset were found to exhibit the same trends (eg. PPP pathway enzymes; glycolytic enzymes of PFK, ALD, TPI). Enzyme subsets are defined as the enzymes which always operate together in fixed flux proportions at steady state (Pfeiffer *et al*., 1999). The behaviour observed in CEF profiles supports the hypothesis that the enzymes belonging to the same subset share similar patterns of genetic regulations (Schuster *et al*., 2002b).

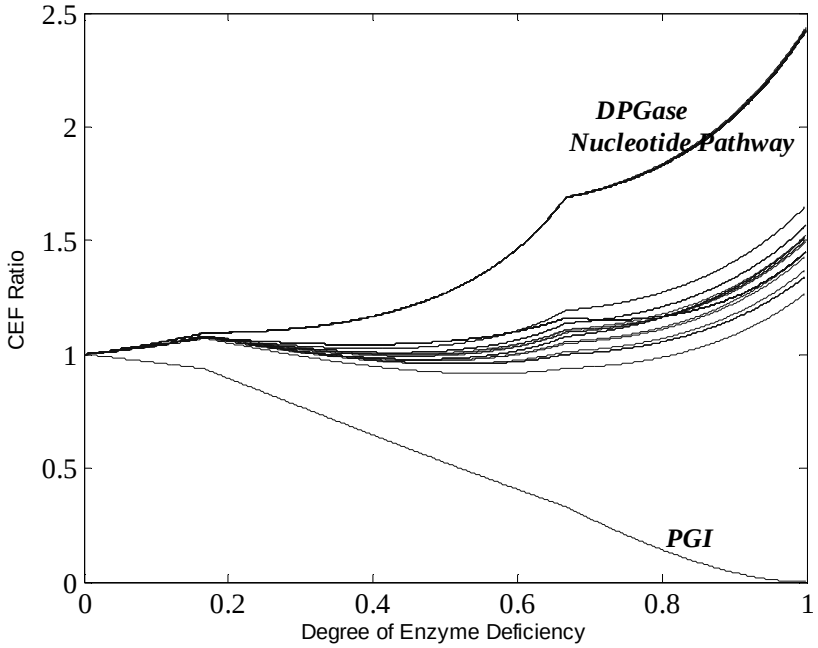
3.2.4.3. Triose Phosphate Isomerase (TPI) Deficiency. In a clinical study, TPI deficiency was investigated from the metabolic and genetic aspects and a high level of DHAP accumulation was reported (Hollan *et al*., 1997; Olah *et al*., 2002; Orosz *et al*., 1996; Repiso *et al*., 2002). On the basis of this literature information, EFM analysis was performed by taking DHAP as external metabolite. The number of functioning elementary flux modes was found as 34. In this case, similar glycolytic CEF values were calculated in TPI enzymopathy as in the healthy case. That is, the expression of the genes responsible for glycolytic enzymes does not differ much when TPI is deficient, in parallel with the clinically observed situation that glycolysis is unaffected in the patients (Repiso *et al*., 2002).



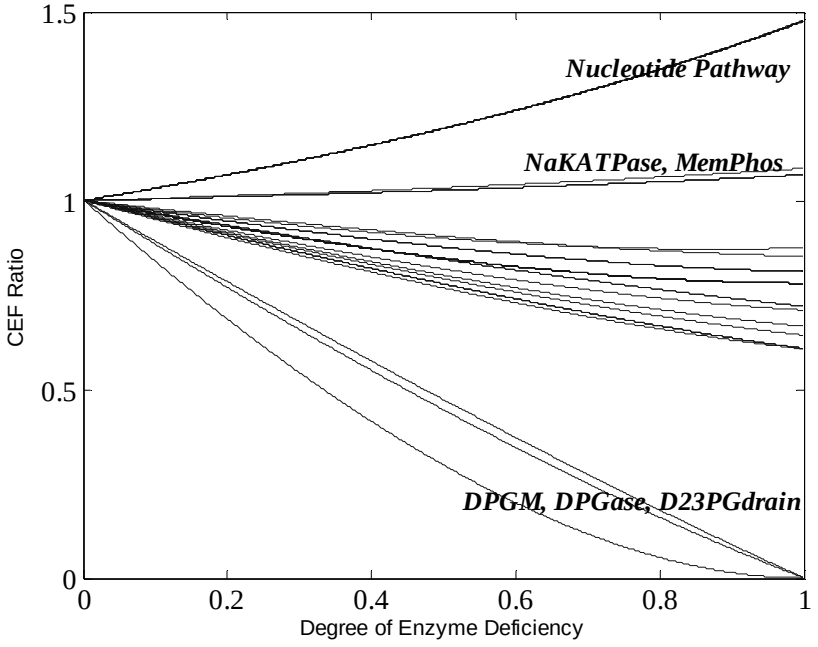
(a) G6PD enzymopathy



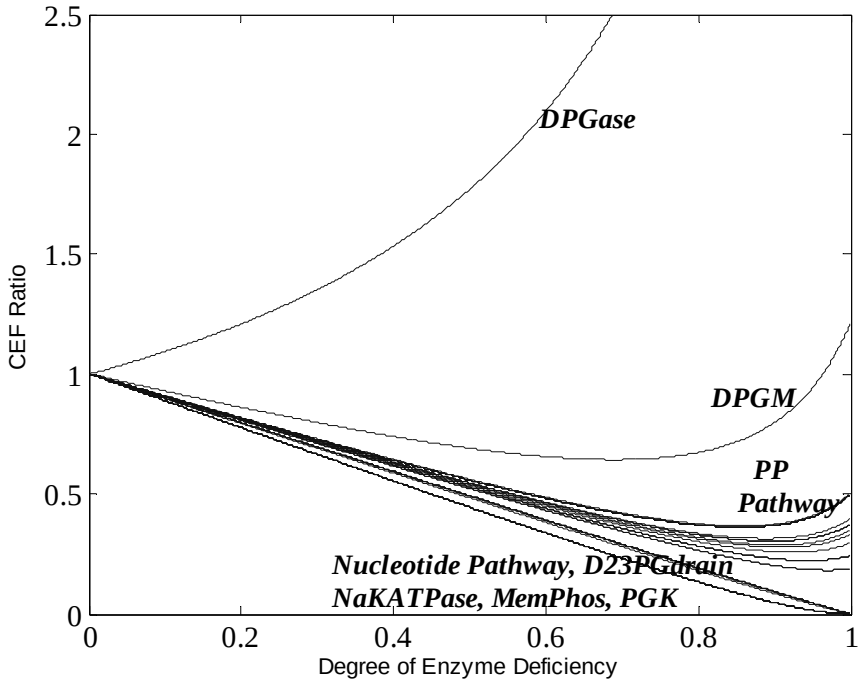
(b) TPI enzymopathy



(c) PGI enzymopathy



(d) DPGM enzymopathy



(e) PGK enzymopathy

Figure 3.6. Change of CEFs upon the activity change of enzymes of considered enzymopathies (a) G6PD (b) TPI (c) PGI (d) DPGM (e) PGK. y axis: CEF ratios with respect to the healthy case. In x axis, 0 fully functional enzyme, 1: complete deficiency. Names of the enzymes with outstanding behaviour are given in figures

Investigating the CEFs in complete TPI deficiency, the Rapoport-Luebering shunt (r_{12}, r_{13}) exhibited a decrease implying less D23PG production, and an increase was observed in pentose phosphate pathway enzymes (Table 3.5). This suggests that the cell is in high oxygenation state (Messana *et al.*, 1996) and fights with the deficiency by increasing its flux to OPPP, as a defense mechanism. The glucose utilization is not negatively affected in deficient cells, which is apparent by the high CEF values of HK and PGI reactions. The enzymes of ATP consumption reactions (NaKATPase and MemPhos) showed 20 per cent decrease, which is coincidental with the behavior of TPI deficiency that this enzymopathy leads to a generalized impairment of cellular energy supply (Jacobasch and Rapoport, 1996).

Deficiency profile (Figure 3.6.b) is associated with an increase in PP pathway and glutathione related enzymes. Efficiency of TPI reaction decreases only 20 per cent upon 50

per cent deficiency, after which a linear decrease is observed. The deficiency is also associated with about 3.5 fold increase in oxidative damage protecting reactions.

3.2.4.4. Phosphoglucose isomerase (PGI) Deficiency. CEF analysis of complete PGI deficiency resulted in increased values of glycolysis and nucleotide metabolism fluxes compared to "no deficiency" case (Table 3.5). The increase in PPP fluxes was not high compared to TPI deficiency, which implies that this deficiency does not have considerable influence on this pathway. Particularly, there are approximately 2.4 fold changes in the fluxes of reactions catalyzed by DPGase and nucleotide metabolism. Although, the decrease in glycolysis fluxes was the case reported in the literature (Jacobasch and Rapoport, 1996), theoretical CEF values indicate opposite behaviour.

The change of CEF value of PGI enzyme upon its deficiency (0 per cent -100 per cent) is slower until about 17 per cent deficiency, after which a linear decrease with higher slope is observed (Figure 3.6.c). The *in silico* profiles of most of the enzymes show similar behaviour associated with almost no change until 70 per cent deficiency, and then a slight increase; except nucleotide pathway enzymes and DPGase. This theoretical analysis does not find any other significantly affected enzyme for this enzymopathy.

3.2.4.5. Diphosphoglycerate mutase (DPGM) Deficiency. In DPGM deficiency case, the CEFs of essential reactions do not show abrupt changes compared to no deficiency case (Table 3.5). This implies that DPGM or DPGase deficiency is not critical, and has slight effects on red blood cells Jacobasch and Rapoport, 1996). A decrease in the rate of OPPP reactions, about 22 per cent, indicates that erythrocyte metabolism is not short of NADPH production, and is not under the possible attack of oxidative damage.

DPGM deficiency profile indicates that ATP consumption reactions, NaKATPase and MemPhos, are unaffected by this deficiency (Figure 3.6.d). This implies that this deficiency is milder compared to other enzymopathies; which is in agreement with other modeling attempts of this deficiency (Schuster and Holzhütter, 1995; Martinov *et al.*, 2000). These modeling approaches use kinetic information on enzymatic reactions in

contrary to this study, where the same conclusion on DPGM deficiency could be derived using only stoichiometric modeling.

3.2.4.6. Phosphoglycerate kinase (PGK) Deficiency. For this deficiency, only 6 EFMs were obtained. Complete PGK deficiency exhibits very low CEF for HK enzyme, indicating that glucose uptake is impaired with respect to healthy case (Table 3.5). This is in agreement with the reported decrease in glucose consumption in PGK deficient red blood cells (Jacobasch, 2000). Nucleotide pathway was also found to be non-functional in parallel with the emphasized lower yield of adenine nucleotides (Jacobasch, 2000). The only remarkable increase is associated with DPGase reaction (8 fold). This would also be expected due to the reported higher levels of D23PG in the patients. High level of this metabolite would supply a high amount of substrate for DPGase reaction, contributing to its enhanced activity.

The deficiency profile (Figure 3.6.e) shows that PGK activity changes almost linearly upon increasing deficiency. DPGM enzyme is negatively affected upon deficiency until the enzyme is 80 per cent deficient, after which it shows a slight increase in CEF value since it becomes an essential reaction in the case of complete enzyme deficiency.

3.2.5. Concluding Remarks

The enzymopathies of G6PD, TPI, PGI, DPGM and PGK in the human red blood cell were investigated by the help of EFM detection and CEF analysis. CEF analysis allowed the detection of the importance of each reaction in the EFMs as well as the relative change in the efficiencies of enzymatic reactions (*in silico* deficiency profiles). It revealed the importance of the glutathione mechanism in the human red blood cell, which prevents oxidation within the cell. The reactions catalyzed by the glutathione enzymes were found to have the highest CEFs in the erythrocyte metabolism. Application of new experimental measurement techniques and consequently obtaining detailed metabolic snapshot of clinically available enzymopathy cases would allow improved comparison and verification of here-reported theoretical results and eventually lead researchers to design drugs for patients suffering from these enzymopathies.

4. EFFECT OF CARBON SOURCE PERTURBATIONS ON TRANSCRIPTIONAL REGULATION OF METABOLIC FLUXES IN *S. CEREVISIAE*

Metabolic fluxes are functions of metabolite levels (metabolic regulation) and enzyme concentrations, and the latter are controlled at transcriptional, translational and/or post-translational levels (hierarchical regulation) (Nielsen, 2003; ter Kuile and Westerhoff, 2001). Functional genomics era has facilitated research on the type of flux regulation through the expression levels of metabolic genes (Lapujade *et al.*, 2004; Oh and Liao, 2001; Krömer *et al.*, 2004). A common approach in the literature is to compare flux levels calculated by flux balance analysis (FBA) or metabolic flux analysis (MFA) with mRNA levels (Famili *et al.*, 2003; Oh and Liao, 2000, Varela *et al.*, 2005). Since many of the reactions are not active under the optimum growth conditions determined by FBA, prediction is not even possible for a number of genes. This is also the case for the MFA approach (Lapujade *et al.*, 2004). Moreover, in FBA the occurrence of alternate optima cannot be excluded (Phalakornkule *et al.*, 2001; Mahadevan and Schilling, 2003; Urbanczik and Wagner, 2005). It was previously shown that these approaches do not account for the flexibility of the metabolic network and that the quality of the resultant prediction is greatly improved by the incorporation of flexibility (Stelling *et al.*, 2002). Elementary flux modes identified by the enumeration of the flux solution space using linear algebra (Schuster *et al.*, 2000) provide the missing flexibility information. Weighted sum of fluxes through these elementary modes, called control-effective fluxes (CEF), lead to the implicit incorporation of functionality and regulation into metabolic network structure (Stelling *et al.*, 2002; Cornish-Bowden and Cardenas, 2002; Çakır *et al.*, 2004). CEF changes were previously used for the prediction of transcriptome changes in carbon source shifts for *E. coli* (Stelling *et al.*, 2002) and *S. cerevisiae* (Chapter 3) metabolisms. Application to erythrocyte enzymopathies was also demonstrated (Chapter 3).

In this chapter, the stoichiometric metabolic model for *S. cerevisiae* used in Chapter 3 was improved and extended by the addition of reactions responsible for major amino acid pathways. The resulting system includes 77 metabolites and 83 reactions which are

governed by a total of 137 genes (Appendix D). Elementary flux modes of this reaction network were calculated for growth on different carbon substrates to determine CEFs which are the weighted sum of modes going through an enzyme. The fold changes of CEFs of reactions in the model in response to perturbations arising from carbon shifts were compared with that of expression levels of metabolic genes responsible for the enzymes of the reactions. The number of fluxes obeying an acceptable correlation was used to evaluate whether the metabolic fluxes are transcriptionally regulated for such perturbations. Hierarchical and/or metabolic regulation was assumed/concluded to be predominant in case of lack of correlation between the ratios of CEFs and mRNAs.

4.1. Computational Methods

4.1.1. Formulation

EFMs were calculated using FluxAnalyzer 5.3 (Klamt *et al.*, 2004). CEF calculations were performed under MATLAB 7.0 environment, and they are based on the efficiencies of calculated EFMs in terms of the chosen cellular objectives: production of biomass itself and ATP for maintenance. Efficiency of an EFM, and CEF of a reaction were calculated based on Equations 3.1 and 3.2, as explained in section 3.1.2. In Equation 3.1, EFMs which are equivalent in terms of cellular objectives are distinguished by assuming that the shorter pathways are more efficient as reflected in the denominator of the formulation (Stelling *et al.*, 2002; Schwarz *et al.*, 2005). This approach coincides with the recently suggested flux minimization objective (Holzhutter, 2004), which implies that the optimum flux distribution is the one which has minimum total flux. The ratio of CEFs of reactions at two different conditions was used to predict the expression ratios of metabolic genes responsible for the enzymes of the reactions in *S. cerevisiae*.

4.1.2. Methodology

Logarithms of the CEF and mRNA ratios for reactions/genes between two conditions were plotted against each other. A script was written in MATLAB to identify the points which cause the largest deviation from a preselected correlation, and these points were omitted from the plot one by one until a correlation (R^2) of 0.60 was reached. This

correlation coefficient value was selected as the threshold for an acceptable degree of correlation since it corresponds to a Pearson correlation coefficient around 0.80, which is considered to be the lower limit for a good correlation (Camacho *et al.*, 2005). Moreover, the correlation between logarithmic mRNA ratios of two different wild type strains (Williams *et al.*, 2002) in response to the same carbon shift was around 0.70 with slope being noticeably different from unity. This inherent variability in cell behaviour depending on its genotype cannot be reflected into metabolic stoichiometry since stoichiometric models are not strain-specific, which also justifies the selected threshold value.

The number of point omissions required to keep the regression coefficient, R^2, above 0.60 was assumed to be one of the criteria for identifying the type of regulation imposed on the fluxes for a particular carbon shift. If many points are to be omitted to reach the threshold, this means a) fluxes are not transcriptionally regulated, but regulated in post-transcriptional, translational or post-translational level (the other hierarchical control mechanisms are active) b) or there is predominant metabolic regulation corresponding to substantial changes in the metabolite levels, c) or both regulation types, metabolic and hierarchical, are active. In addition, a second qualitative criterion called 'correct prediction', which is based on the number of points in the first (up-regulation) and third (down-regulation) quadrants of the plotted coordinate axis, was also used, as employed by others (Famili *et al.*, 2003).

4.2. Results

The set of experimental gene expression data for carbon source perturbations used in this study are summarized in Table 4.1. If multiple genes correspond to a single reaction, their expression levels were summed up for each condition before the calculation of the fold change. The metabolic model consists of central carbon metabolism reactions as described in Chapter 3, and improved by the inclusion of the reactions involved in the synthesis of major amino acids (Appendix D). Other amino acids which contribute to the smaller portion of protein composition of *S. cerevisiae* (Förster *et al.*, 2003) were directly incorporated into the biomass reaction (r_{83a}) rather than including the individual reactions responsible for their formation into the set of stoichiometric reactions taken into consideration. The network could therefore be restricted to a manageable medium-scale

size. This processing was necessary to avoid the combinatorial explosion in the number of elementary flux modes with the increase in the number of considered reactions, which cannot be handled with the current algorithms (Klamt *et al.*, 2002, Schwarz *et al.*, 2005). The stoichiometric coefficients of the reaction leading to biomass formation were calculated on the basis of the biomass composition given by Förster *et al.*, 2003.

Table 4.1. Transcriptome datasets used in this study

Article	Source change	Fermentation type
DeRisi *et al.*, 1997	Carbon: Glucose- Ethanol	Batch
Lapujade *et al.*, 2004	Carbon: Glucose-Ethanol	Chemostat
Williams *et al.*, 2002	Carbon: Glucose-Acetate	Batch
Lapujade *et al.*, 2004	Carbon: Glucose-Acetate	Chemostat
Prokisch *et al.*, 2004	Carbon: Glucose-Lactate	Batch
Piper *et al.*, 2002	Oxygen: Aerobic-Anaerobic	Chemostat

Table 4.2. Number of EFMs for each studied carbon source for the biomass composition reported in Förster *et al.*, 2003. The numbers in paranthesis shows EFM numbers when the biomass composition of Gombert *et al.*, 2001 is employed for comparison

Substrate	EFMs - M83*	EFMs - M46
Glucose	136925 (184631)	13255
Ethanol	11427 (15099)	1225
Acetate	4240 (5452)	536
Lactate	25484 (34319)	2533

*In M83, the EFMs with simultaneous occurrence of GDH2 and GDH3 in reverse directions were not taken into account since this leads to transhdrogenase activity, which is known to be not available in *S. cerevisiae*.

The number of EFMs calculated for each carbon source is given in Table 4.2 for the model in Appendix D called M83 (based on the number of included reactions), and for a modified version of this model, M46, which only includes central carbon metabolism reactions as in Chapter 3. When the number of EFMs of the two models is compared, an approximately ten-fold increase is observed in the case of M83. Therefore, it may be concluded that the inclusion of amino acid reactions enables better and less-restricted representation of the microorganism flexibility. The coefficients of biomass constituents

were calculated also on the basis of another cellular macromolecular composition reported by Gombert *et al.*, 2001 for *S. cerevisiae* (r_{83b}), and this calculation led to noticeable differences in the resultant number of EFMs for the same carbon source (Table 4.2). However, variations between calculated CEFs for each EFM set were small and, therefore, the biomass composition in Förster *et al.*, 2003 was used in EFM and CEF calculations throughout this study.

The following strategy was pursued to distinguish active EFMs during growth on glucose in different fermentation types. For batch experiments (Table 4.1) operated in respiro-fermentative mode, EFMs producing any of the considered byproducts (ethanol, glycerol, acetate, succinate) were retained since this mode is mainly associated with simultaneous biomass and product formation; those producing only biomass were discarded, leading to 127872 EFMs instead of 136925. Biomass-only EFMs were considered in chemostat experiments since no by-product was detected in the medium at a dilution rate of 0.1 h^{-1} (Lapujade *et al.*, 2004), leading to 9600 EFMs instead of 136925. This approach was used to test the prediction capabilities of the previous models where all the EFMs had been used without such distinction for comparison with experimental data (Stelling *et al.*, 2002; Çakır *et al.*, 2004). The present strategy to include only active EFMs into the model was found to enable improved predictions of gene expression changes (results not shown).

Table 4.3 summarizes the simulation results for genes belonging to central carbon metabolism (45) for each case in Table 4.1, including the correlation coefficient, slope, correct qualitative prediction, and the omitted points to reach $R^2 = 0.60$. An acceptable correlation ($R^2 = 0.60$) with a slope close to the unity was possible by omitting at most 6 points for the studied carbon shifts (Table 4.3). Points which had to be omitted correspond to reactions whose CEF values do not show correlation with the change in expression levels of the genes encoding the enzymes catalyzing these reactions. It is highly probable that these fluxes are regulated at post-transcriptional and/or metabolic level. Correct prediction of qualitative up-regulation and down-regulation was above 76 per cent for all cases. These results (Table 4.3) indicate that fluxes corresponding to central carbon metabolism reactions are mainly transcriptionally regulated in carbon shift experiments. A clear lack of correlation was observed in the oxygen shift experiment, for which 20 points

had to be discarded to reach the threshold correlation (R^2 = 0.60). Plots of mRNA ratios versus CEF ratios for all cases studied are given in Figures 4.1 to 4.5. The present results also reveal the fact that central metabolic genes are predominantly upregulated in response to a shift from fermentative carbon source to a source with C-2 C-3 compounds as most of the points lie in the first quadrant in Figures 4.1 to 4.5.

Table 4.3. Results of simulations for genes belonging to central carbon metabolism

	Omissions for R^2:0.60	Correct Prediction*	Slope / R^2	Number of EFMs used in simulation	Omitted Genes	Not Applicable Genes#
Glucose/Ethanol, batch	3$	0.82 (36/44)	1.06/ 0.65	127872/ 11427	*pfk*$, *fbp1*$, *pyc*	*(gpp1-hor2)*
Glucose/Ethanol, chemostat	6	0.77 (33/43)	0.81/ 0.60	9600/ 7051	*pfk pyc ald4 pda fba tpi*	*(gpp1-hor2) bph1*
Glucose/Acetate, batch	3	0.76 (31/41)	1.11/ 0.63	127872/ 4238	*mae idp2 rpe*	*(gpp1-hor2) bph1 Pyc pfk*
Glucose/Acetate, chemostat	6	0.78 (32/41)	1.18/ 0.61	9600/ 4190	*adh1 rki lsc sol fba zwf*	*(gpp1-hor2) bph1 pfk pyc*
Glucose/Lactate, batch	4	0.84 (38/45)	0.89/ 0.60	127872/ 25482	*pfk idp2 pyk mae*	-
Aerobic/ Anaerobic, chemostat	20		1.25/ 0.62			

*The points with a fold change between 0.95-1.05 for either of model or experiment were considered to be correctly predicted.
$These genes were found to exhibit better agreement with CEF ratios in the analysis of Gasch *et al.*(2000).
#CEF ratio was either zero or infinity for these genes. Therefore, they could not be used in the correlation calculation.

The effect of the assumption regarding assignment of higher efficiency to the shorter pathways was tested for each of the perturbations analyzed here. Denominator of Equation 3.1 was not taken into account, thereby assigning equal efficiency to the EFMs with the same objective flux, regardless of the corresponding length. A lower correlation was

observed when the higher efficiency of shorter pathways was not considered. These results support the hypothesis that shortest modes contribute most to (cellular activity)/ correlation between CEFs and gene expression (Stelling *et al.*, 2002; Schwarz *et al.*, 2005).

4.3. Discussions

The correlation between mRNA ratios of the genes and corresponding CEF ratios were investigated for the genes belonging to central carbon metabolism and amino acid metabolism separately using M83 model.

4.3.1. Correlation between mRNA ratios for the genes of central carbon metabolism and corresponding CEF ratios

The present results indicate that the response of most of the central carbon metabolism genes to a perturbation in the carbon source is at transcriptional level and is transmitted hierarchically to flux level (Table 4.3). However, the same genes are found to be weakly correlated with CEFs in the case of oxygen shift, indicating that the response of the same genes to different perturbations is not shaped by a similar control mechanism.

For each carbon source perturbation, a small set of genes whose mRNA ratios were weakly correlated with the CEF ratios were omitted. For example, three genes, namely pfk1, fbp1 and pyc, displaying a weak correlation as a response to a diauxic shift in batch cultures (DeRisi *et al.*, 1997) were omitted (Table 4.3). Two of these genes (fbp1, pfk) are responsible for the expression of enzymes involved in conversion between fructose-6-phosphate and fructose-diphosphate in reverse directions. fbp1 is known to be active in ethanol growth whereas pfk is active during glucose growth. Although their up or down regulation matches with CEF predictions perfectly, the quantitative relation is absent. That is, relatively insensitive ratiowise response at gene expression level may indicate an amplified transmission of the signal to flux level. However, investigation of another dataset (Gasch *et al.*, 2000) for the same respiro-fermentative shift shows better correlation for these genes, and corresponding CEF values as shown in Figure 4.1 by square points. Thus, these genes may also be false-negatives resulting from the absence of replicates.

For the glucose/ethanol shift in the chemostat culture, glycolysis pathway genes (pfk, pyc, ald4, pda, fba, tpi) are found to be the mainly deviating ones, undergoing other kinds of regulation rather than transcriptional (Table 4.3). This is supported by a recent study, which shows that glycolytic genes are regulated at the level of proteome in response to the same perturbation (Kolkman *et al.*, 2005). The present analysis indicates a good correlation between the magnitude of change in CEFs and transcript levels of genes, with the exception of these six points. Here, this analysis gives a better correlation than the comparison made using MFA based fluxes by Lapujade *et al.*, 2004, where 19 out of 43 genes could not be included in correlation analysis since the corresponding MFA-based fold change was either zero or infinity, and the fold changes of 21 of the remaining points showed a correlation above the threshold ($R^2 = 0.60$), with a slope several folds higher than unity (3.5). This indicates that the use of general metabolic capabilities of the microorganism under a carbon source as reflected in calculated EFMs, rather than focusing on a single flux distribution, results in better representation of the hierarchical behaviour of the control in gene expression. This is also valid for the diauxic shift in batch cultures. CEF approach with 82 per cent qualitative correct prediction (Table 4.3) is superior to the FBA approach with 61 per cent qualitative correct prediction (Famili *et al.*, 2003) which is based on the number of up-regulated and down-regulated points that are in agreement between experimental and simulation results.

For the glucose/acetate shift in chemostat cultures, three of the six omitted genes belong to pentose phosphate pathway (rki, sol, zwf). The other three are from different pathways, lsc from TCA cycle, fba and adh1 from glycolytic pathway. The corresponding fluxes are similarly assumed to be subjected to regulation types other than transcriptional (Table 4.3). Unlike chemostat cultures, lack of transcriptional regulation through fluxes of reactions governed by two different genes, namely *idp2* and *mae*, is implied in case of batch cultures for glucose-acetate shift (Table 4.3).

For genes of central carbon metabolism predictions by M83 was better than that of M46, meaning that further incorporation of the possible paths spanning amino acid pathways reflects the flexibility of the organism better (results not shown).

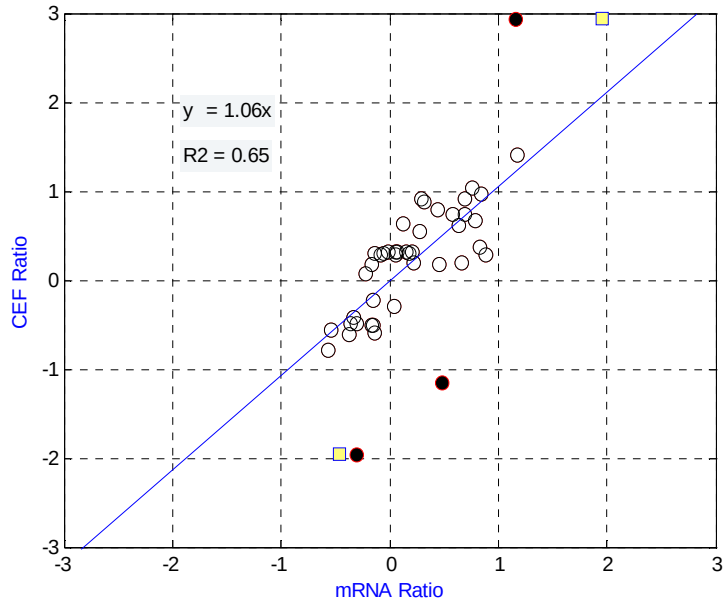


Figure 4.1. Comparison of model-based and data-based ratios for carbon shift from glucose to ethanol in batch cultures. Filled circles are the omitted points to reach $R^2 = 0.60$. The squares are obtained by using the data from Gasch *et al.*, 2000

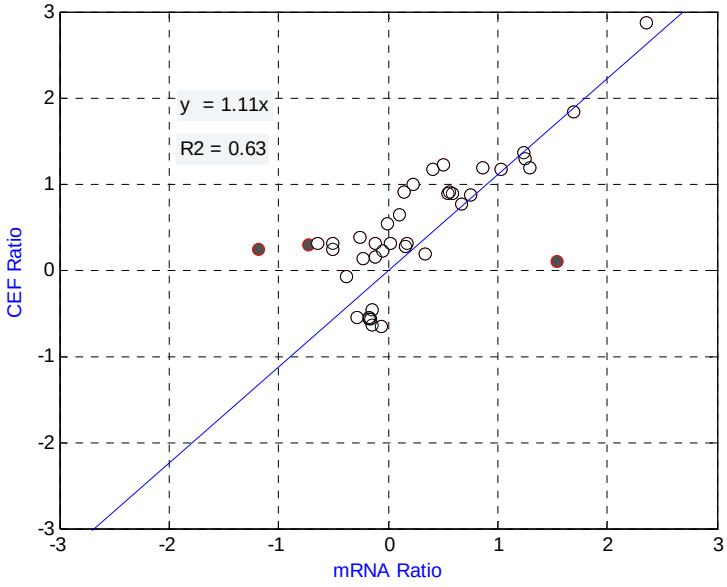


Figure 4.2. Comparison of model-based and data-based ratios for carbon shift from glucose to acetate in batch cultures. Filled circles are the omitted points to reach the selected cut-off value of $R^2 = 0.60$

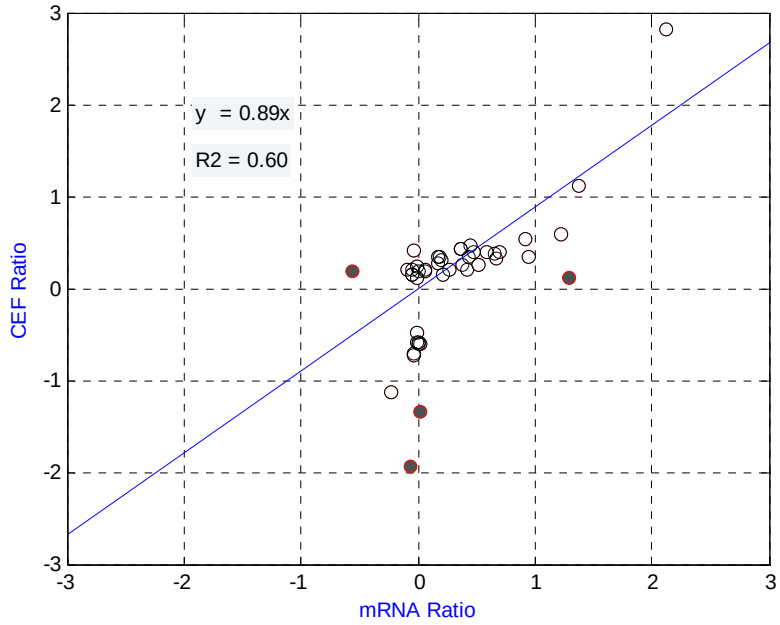


Figure 4.3. Comparison of the model-based and data-based ratios for carbon shift from glucose to lactate in batch cultures

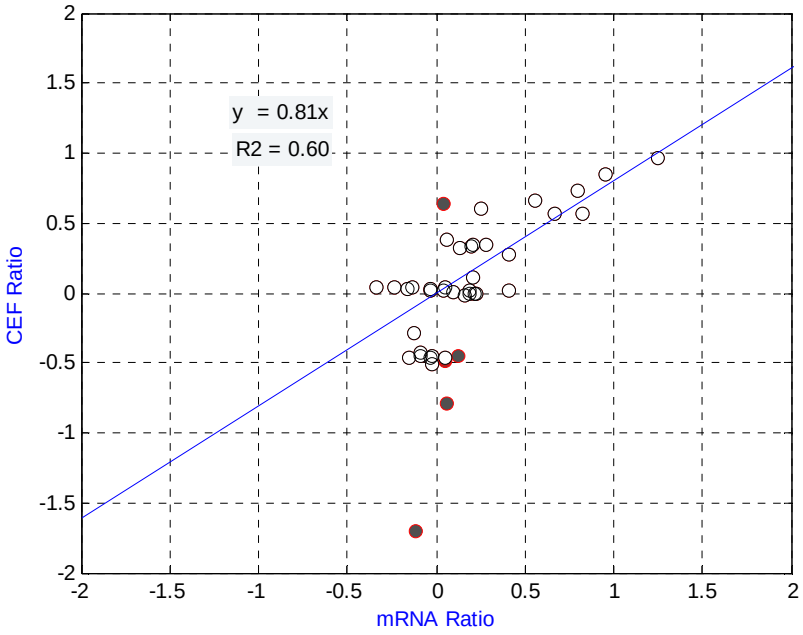


Figure 4.4. Comparison of the model-based and data-based ratios for carbon shift from glucose to ethanol in chemostat cultures

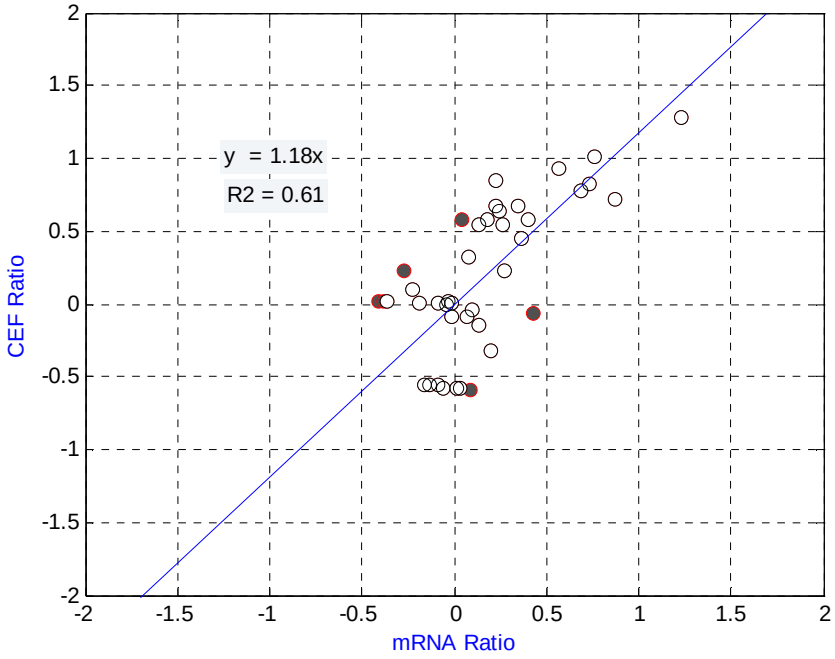


Figure 4.5. Comparison of the model-based and data-based ratios for carbon shift from glucose to acetate in chemostat cultures

4.3.2. Correlation between mRNA ratios for the genes of amino acid pathways and corresponding CEF ratios

For respiratory chemostat datasets, it is difficult to establish a correlation between ratios of expression levels of amino acid genes and corresponding CEF ratios since numerical values of both experimental mRNA and model CEF ratios for these genes are very close to unity. Therefore, these genes do not have pronounced effect on the resultant correlation and slope.

For respiro-fermentative batch datasets, on average ten more points had to be removed from the graph to get the predetermined correlation of $R^2 = 0.60$. That is, there was a lack of correlation between the ratios of expression levels of amino acid pathway genes and corresponding CEF ratios. The observed lack of correlation for amino acid genes was also obvious for *E. coli* model (Stelling *et al.*, 2002). There was no positive correlation between five genes available in the model belonging to amino acid metabolism, consistent with the behaviour observed for *S. cerevisiae* as summarized above. However, one should be cautious to judge this lack of correlation as the weakness of transcriptional regulation

for genes involved in amino acid synthesis pathways. The experiments analyzed include a direct perturbation in the carbon source. The resultant effect on central carbon metabolism could be captured by the model more easily as the perturbation was directly related to this metabolism. Since there was no perturbation for amino acid pathways such as the change in nitrogen source, the results may not reflect the real situation. Additionally, the lack of correlation in batch cultures may be explained by the use of rich media in the fermentations. Due to their availability in the media, the amino acids might not have been resynthesized within the cell, which may be another cause of poor correlation. Therefore the correct analysis of amino acid pathways requires specially designed experiments. Moreover, it has been reported (Grotkjær, 2005) that the change in the growth rate has a direct influence on the mRNA levels. The lack of correlation for the relatively unperturbed amino acid pathways can be attributed to the possible difference in microbial growth rates on the compared carbon sources in the batch experimental datasets. Furthermore, the mRNA ratios may have been biased by the normalization methods employed for the analysis of transcriptome data.

On the other hand, as will be presented in Chapter 5, comparison of the significance of statistical change in transcriptome and metabolome profiles of *S. cerevisiae* under different conditions (Çakır *et al.*, 2006) led us to the conclusion that almost all of the genes governing amino acid metabolism were metabolically regulated with or without transcriptional regulation. Although this finding cannot be stated as a strong support due to the reasonings listed above, this may be another explanation for the lack of correlation observed here. Then, this poor correlation implies that central carbon metabolism genes are much more transcriptionally regulated than those governing amino acid metabolism. Therefore, in this study the focus was on the analysis of central carbon metabolism since this was the part mostly affected by the here-studied carbon-source perturbations as explained above.

4.3.3. Effect of media and strains on the transcriptional regulation of fluxes

The fluxes of central carbon metabolism are found to be mainly transcriptionally regulated in response to carbon source perturbations (Table 4.3). In order to investigate the effect of strain type on the regulation of fluxes, the experimental dataset for glucose-

acetate shift in batch cultures (Williams *et al.*, 2002) was used. This study includes the identification of changes in transcriptome as a response to the same type of perturbation for two different wild type yeast strains (W303, SK1). The result presented in Table 4.3 is for W303 strain and indicates that fluxes of central metabolism of this strain is subject to transcriptional regulation, with only three disobeying fluxes. The analysis of the other strain (SK1) revealed a requirement of omission of seven more points to reach the predetermined correlation ($R^2 = 0.60$) in addition to three omissions for W303. This result suggests that the regulation behaviour can strongly depend on the genotype of strain itself as suggested elsewhere (Ferea *et al.*, 1999; Brem *et al.*, 2002; Townsend *et al.*, 2003; Jansen *et al.*, 2005; Çakır *et al.*, 2006).

W303 strain is suggested to exhibit more fermentative behaviour than SK1 strain in glucose containing medium (Williams *et al.*, 2002). The expression levels of genes involved in respiratory metabolism were higher for SK1 strain than for W303 strain. This information was used to test the introduced approach of distinguishing active EFMs operating in respiratory and respiro-fermentative growth. CEF analysis and comparison of CEF and mRNA ratios for SK1 were performed considering all EFMs for glucose growth instead of taking only those co-producing biomass with any of by-products. The underlying assumption is that those producing only biomass must also be active in this strain displaying a more respiratory behaviour. Resultant number of omissions was reduced to 8 for SK1 strain. On the other hand, use of all EFMs for W303 strain caused an increase in the number of omissions to five. It should be noted that the incorporation of an information on the phenotypic/fermentative behaviour of the strain into the analysis may improve the prediction of the fluxes that are transcriptionally regulated.

In order to investigate the effect of media on the regulation of fluxes, the experimental dataset for glucose-lactate shift in batch cultures, where the changes in transcriptome as a response to this perturbation for YPD and synthetic complete media were reported (Prokisch *et al.*, 2004), was used. The same type of analysis was also carried out for synthetic complete media, and the results were compared with that of YPD in Table 4.3. The number of the omitted fluxes increased to eight, indicating the effect of medium components on the regulation type of particular fluxes.

4.4. Alternative Approaches to CEF Calculation

Two additional approaches can be followed to generate the set of flux distributions as an input to CEF calculation. The glucose-ethanol diauxic shift case in batch cultures (DeRisi *et al.*, 1997) is used here to demonstrate the prediction power of these approaches for metabolic transcriptome changes, as compared to EFM-derived approach for CEF ratios.

The first approach is the uniform sampling of the solution space. There are two different alternatives for such a sampling, i.e. Monte-Carlo and hit-and-run. Of the two methods that can be used to uniformly sample the flux solution space constrained by the stoichiometric matrix and the reaction irreversibility information, Monte-Carlo approach is not applicable for large-scale models. The other method, hit-and-run sampling was employed here to get a set of flux distributions sampled randomly and uniformly. Detailed formulation of the approach is given in Appendix E. The sampling was performed until about 4,000 flux distributions are obtained. Hit-and-Run algorithm was executed to get 8 million sample points; and only every 2000^{th} point was saved in order to prevent the interdependency of the consecutively generated samples. Collection of higher number of flux distributions did not result in noticeable difference in calculated CEFs. Therefore, these sampling parameters were used in simulating yeast growth on both glucose and ethanol. Then, the collected flux distributions were used in the calculation of CEFs. CEF calculation was repeated for three different sampling runs, to detect if the flux distributions generated by sampling shows variability. No significant variation was observed among the results of different executions. Figure 4.7 presents the correlation between sampling-based CEF ratios and transcript ratios. The corresponding figure based on EFM-based CEF ratios (Figure 4.1) is regenerated as Figure 4.6 without omitting any points for comparison.

As another alternative to CEF calculation, maximization/minimization of a randomly selected linear objective function was used in linear programming-based determination of flux distributions. After collecting enough number of flux distributions, CEF calculations were repeated, and the resultant correlations were calculated. Generation of about 1000 flux distributions by randomized FBA was considered to be enough since execution of this approach in different times led to flux distributions with almost the same CEF values.

Figure 4.8 presents the correlation between the fold change of mRNA levels (DeRisi *et al.*, 1997) and that of multiFBA-based CEFs.

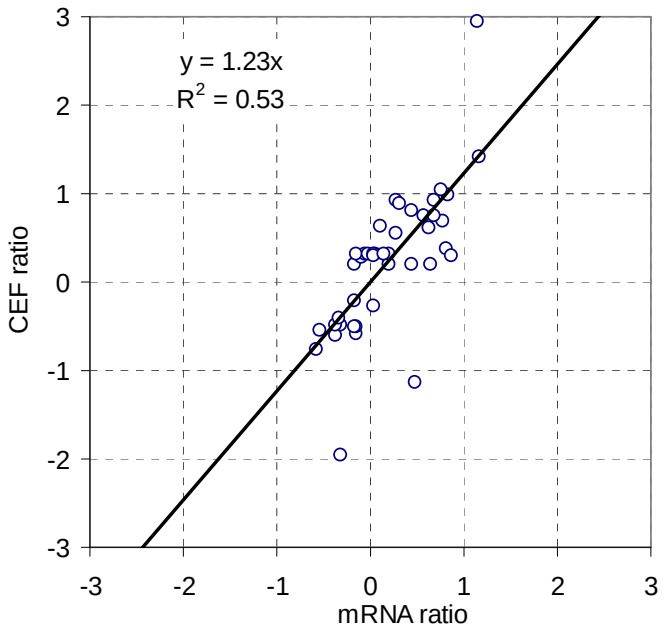


Figure 4.6. Correlation between experimental mRNA ratios and EFM-based CEF ratios for carbon shift from glucose to ethanol in batch cultures

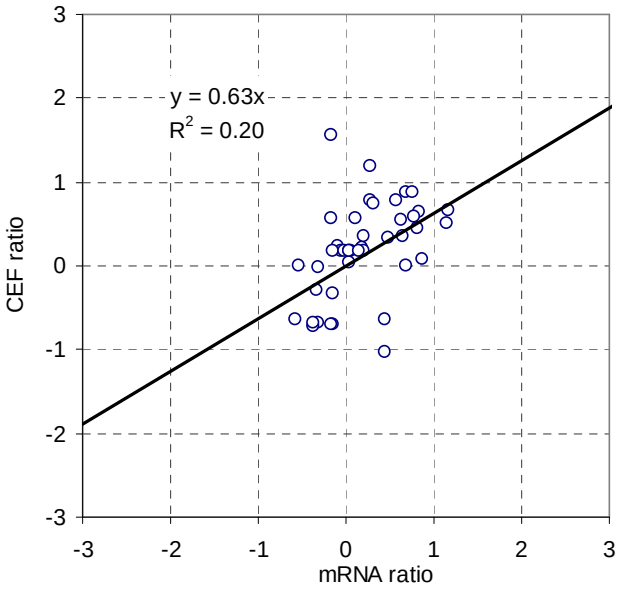


Figure 4.7. Correlation between experimental mRNA ratios and hit-and-run-based CEF ratios for carbon shift from glucose to ethanol in batch cultures

As it can be seen from both of the figures (4.7 and 4.8), the two alternative approaches lead to noticeably worse correlation with experimental data, compared to the correlation when EFM-based CEF values are used.

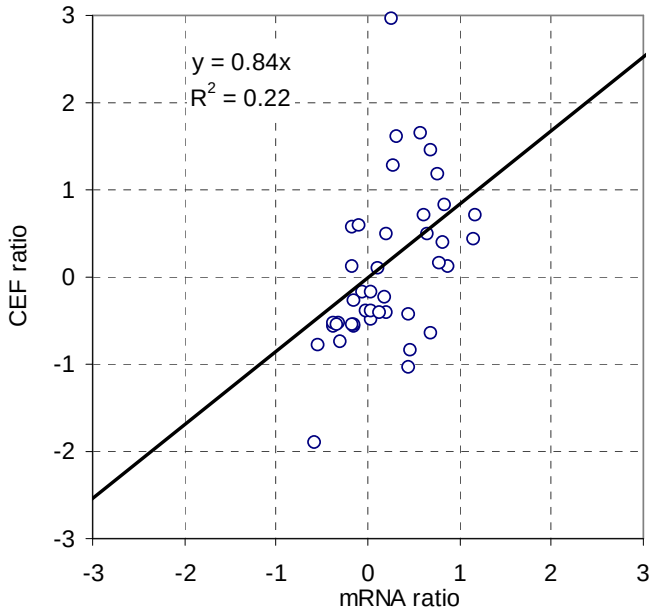


Figure 4.8. Correlation between experimental mRNA ratios and multi-LP-based CEF ratios for carbon shift from glucose to ethanol in batch cultures

4.5. Concluding Remarks

The hierarchical transmission of transcriptome changes to flux level was investigated using control effective fluxes rather than the fluxes derived from flux balance analysis. The degree of high correlation between transcriptome and fluxome obtained by CEF approach points out that the major reason for the lack of correlation reported so far between gene expressions and fluxes was due to neglecting the flexibility information of the network in operation. The detailed analysis using CEFs has shown that fluxes of central carbon metabolism are predominantly regulated at the transcriptional level in response to changes in carbon source. Regulation of amino acid metabolism seems to be mainly at the metabolic level; however, a definite conclusion can not be drawn since the analyzed perturbations were not directly related to amino acid metabolism. Therefore, analysis of nitrogen shift experiments will provide important information on the regulatory mechanism

of amino acid pathways. These results lead to the hypothesis that if the applied perturbation has a direct effect on a metabolic pathway, then the genetic response of that pathway at mRNA stage is propagated into the flux stage, as demonstrated for central carbon metabolism in this study.

5. INTEGRATION OF METABOLOME DATA WITH METABOLIC NETWORKS REVEALS REPORTER REACTIONS

One of the goals of systems biology is to obtain overall quantitative description of cellular systems. This is currently not achievable since the number of components and interactions involved in these systems is quite large resulting in a very large parameter space. Thus, methods are required to reduce the dimensionality and particularly identify key regulatory points in the many different cellular processes. Metabolism is a good starting point to develop such analysis methods as it is studied in great detail and well annotated. Furthermore, genome-scale metabolic models have been developed for many different cellular systems (Edwards & Palsson, 2000; Förster *et al*., 2003; Sheikh *et al*., 2005), and besides their use for simulation of cellular function (Edwards *et al*., 2001; Famili *et al*., 2003; Price *et al*., 2004b) these models can serve as scaffolds for analysis of genome-scale biological data (Covert *et al*., 2004; Borodina & Nielsen, 2005). This has been demonstrated recently for analysis of transcriptome data, where the use of genome-scale metabolic models enabled identification of co-regulated sub-networks and reporter metabolites (Patil & Nielsen, 2005). Although transcriptome data provides an overview of the global regulation in the metabolism, understanding of cellular physiology is incomplete without knowledge of metabolome owing to the high connectivity in metabolic networks and inherent inter-dependency between enzymatic regulation, metabolite levels and fluxes (Nielsen, 2003). Metabolites, acting as intermediates of biochemical reactions, play a crucial role within a living cell by connecting many different operating pathways. Metabolite levels are determined by the concentrations and the properties of the surrounding enzymes, making their levels a complex function of many cellular regulatory processes in different dimensions. Thus, the metabolome represents a snapshot of the functioning metabolism of the cell and hence provides valuable information about regulation of several different cellular processes (Villas-Bôas *et al*., 2005c). Consequently, in recent years there has been increased focus on analysis of the metabolome (Sumner *et al*. 2003; Bino *et al*., 2004; Villas-Bôas *et al*., 2005c). Even though traditional data analysis methods like principal component analysis, clustering analysis and chemometrics have shown to be efficient for analysis of this kind of data (Raamsdonk *et al*., 2001; Allen *et al*., 2003), there are some limitations with these methods for uncovering the underlying

biological principles (Weckwerth *et al.*, 2004). Furthermore, there are still only few example studies on the use of metabolome data to understand regulatory principles in metabolism.

Functional analysis of cellular metabolism and, in particular, integration of metabolome data with other omics-data demands (semi-)quantitative measurements of key metabolites. However, a problem with metabolomics is the scarcity of targeted quantitative data, and often metabolome analysis is (at best) semi-quantitative even though there is a trend towards more quantitative analysis (Nielsen & Oliver, 2005). Although it is currently not yet possible to quantify all the metabolites in a cellular system (Goodacre *et al.*, 2004, Fernie *et al.*, 2004), a high-throughput GC-MS method that allows semi-quantitative identification of several metabolites in *S. cerevisiae* was recently developed (Villas-Bôas *et al*., 2005a; Devantier *et al*., 2005a). In the latter studies, the levels of 52 unique metabolites (out of 584 reported unique metabolites in the genome-scale yeast model, Förster *et al*., 2003) were determined in genetically different yeast strains under different environmental conditions. Specifically, metabolites playing important roles in the central carbon metabolism and amino acid biosynthesis were able to be identified.

In order to understand the regulatory principles underlying the changes in metabolite levels an algorithm was developed that enables integration of such quantitative metabolome data with genome-scale models by using a graph theoretical representation of the metabolism. The application of this algorithm is demonstrated here for the metabolome data reported by Villas-Bôas *et al*., (2005a) and Devantier *et al*. (2005a). The algorithm includes preprocessing of a genome-scale yeast model such that the fraction of measured metabolites within the model is enhanced, and hereby it is possible to map significant alterations associated with a perturbation even though a small fraction of the complete metabolome is measured. The significance of changes in the metabolite levels is used to identify reporter reactions around which the most significant coordinated metabolite changes are observed. Reporter reaction analysis is an attempt to infer the differential reaction significance based on metabolite measurements, and hence provides a basis for understanding the underlying cellular processes responding to the perturbations. It is further demonstrated that through combination with transcriptome data, reporter reactions may provide clues on whether regulatory control at a given reaction node is at the

metabolic level or at the hierarchical level. Hereby, the reported approach represents an attempt to map different layers of regulation within metabolic networks through combination of metabolome and transcriptome data.

5.1. Methods

5.1.1. Graph Representation

In the present study, the metabolic network ENZSUB-3, the pre-processed model (see section 5.2), was represented as a bipartite undirected graph in order to identify reporter reactions. Reactions and metabolites were both taken as nodes, and the edges denoted the interactions between them (Patil & Nielsen, 2005). Hence, the resulting graph consisted of 317 nodes.

Different genetic and environmental perturbations associated with the two datasets (Devantier *et al.*, 2005a; Villas-Bôas *et al.*, 2005a) were analyzed. The graph representation was used to identify 'reporter reactions' for these perturbations. The algorithm used in the simulations is a modification of the algorithm recently developed by Patil & Nielsen, 2005, which was based on the analysis of transcriptoma data to identify so-called reporter metabolites, the spots in the metabolism with substantial transcriptional regulation. The modified algorithm herein has the capability of identifying reporter reactions, the putative key points in the metabolism in terms of metabolic regulation (Figure 5.1).

5.1.2. Significance Test

The significance of change for the experimental metabolite levels between any two conditions were determined by comparing the levels with the aid of a statistical-test, thereby quantifying the effect of the associated perturbation. For each of the perturbations, the statistical test was applied to the experimental data following the normalization process described by Villas-Bôas *et al.* (2005a). Briefly, the normalization process is such that the within-group variances among replicates are reduced and between-group variances are maximized. The Mann-Whitney rank-sum u-test is a nonparametric statistical test which

has no *a priori* assumption about the distribution type of the data. It was preferred over the standard t-test since the distribution of levels of some of the metabolites among the replicates, especially NAD+ and NADPH, was found to be skewed rather than normal distributed. The Student t-test assumes normal distribution of the data and compares the mean values whereas the u-test compares medians rather than means. Furthermore, median is a better measure for skewed distributions since it is less sensitive to the extreme scores that can be encountered in the replicates.

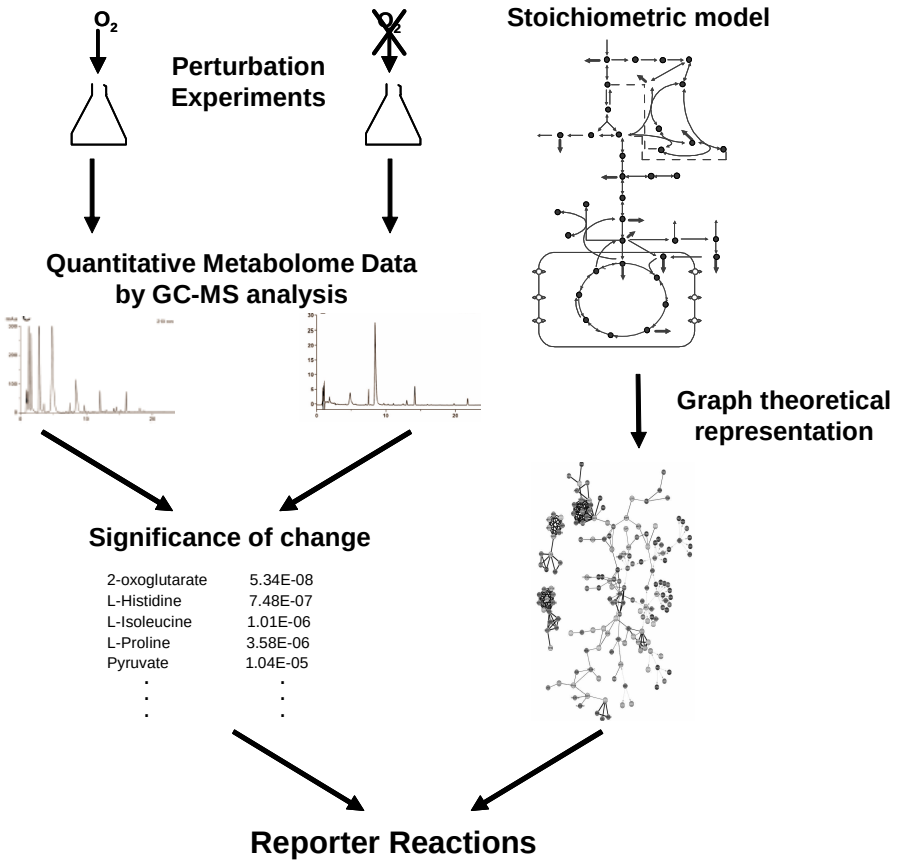


Figure 5.1. Reporter reaction algorithm to identify differential reaction significance by integrating metabolome data with metabolic networks

5.1.3. Strategy for the Lack of Data

Since the utilized reporter reaction algorithm depends on the scoring of reactions based on the p-values of involved metabolites, the lack of p-values for the 94 metabolites that remain unmeasured in the final ENZSUB-3 model must be handled. Random

assignment from GC-MS peaks was used to overcome the problem of the unavailable data. GC-MS spectra contain a large number of unknown peaks due to unmeasured metabolites. All the peaks in GC-MS spectra were deconvoluted for each replicate. The output was normalized by using a Python code which minimizes the sample variability within the classes (Villas-Bôas *et al*., 2005a). Afterwards, the peaks in the spectra within a selected time interval (0.15 minutes) were binned to account for the fluctuations in the retention times using a MATLAB algorithm. This has resulted in the overall detection of 236 unknown peaks for the first dataset (Villas-Bôas *et al*., 2005a), with 116, 178 and 201 non-zero peak comparisons for genetic perturbations under aerobic and anaerobic conditions and environmental perturbations respectively, and 240 unknown peaks for the second dataset (Devantier *et al*., 2005a) with 129 and 174 non-zero peak comparisons for the environmental perturbation of laboratory and industrial strains respectively. The significance of change for these unknown peaks was quantified for each perturbation by means of p-values using the u-test. These p-values were randomly assigned to the unmeasured metabolites.

5.1.4. Reporter Reaction Analysis

Resultant p-values were converted to z-scores using an inverse normal cumulative distribution function for further analysis. Each reaction in the constructed graph was scored by calculating the score of the subnetwork formed by its k neighboring metabolites, and z-values of the metabolites were used in the scoring.

$$Z_{reaction} = \frac{1}{\sqrt{k}} \sum Z_{metabolite,k} \tag{5.1}$$

$Z_{reaction}$ score was then corrected for background distribution using the mean (μ_k) and standard deviation (σ_k) of z-scores of metabolite groups of the same size, obtained by random sampling from the same metabolic network.

$$Z_{corrected-reaction} = \frac{Z_{reaction} - \mu_k}{\sigma_k} \tag{5.2}$$

In order to minimize the sensitivity of reporter reactions to the randomly selected p-values for the non-measured metabolites as mentioned above, the reporter-reaction algorithm was executed 1000 times by repeating the random assignment in each case. This repetition eliminated the effect of the p-values of the assigned peaks on results. For each reaction, the z-scores in each repetition were averaged to get a final z-score. Those reactions with the highest z-scores (typically $z > 1.28$, corresponding to $p < 0.10$) can be defined as reporter reactions for a system with complete metabolome data. Since available experimental data were not complete, the calculated z-scores were used for deducing the relative significance of the reactions in the analyzed perturbations. Namely, the main focus is comparative analysis of reactions among the studied perturbations as revealed by Figure 5.3, rather than comparing a reaction to another based on its Z-score. The underlying reason is to avoid potentially incorrect conclusions due to the unmeasured metabolites which have randomly assigned p-values. Additionally, the analyzed reactions have a high percentage of measured metabolite content as indicated in Tables 5.1 and 5.3. In the case of low coverage of measured metabolite content, this method should be followed with caution as the resultant Z-scores of reactions will become insignificant, and such reactions will not be picked up as reporters. However, in future when analytical methods have been further improved it is likely that more metabolites can be measured, and one will overcome this shortcoming and our approach may then be used to infer more solidly about the level of regulation at different parts of large metabolic networks.

5.1.5. Computational Tools and RepRxn MX Software Package

METATOOL 4.3 (Pfeiffer *et al.*, 1999) was used for the identification of enzyme subsets in the UNCOMP model. The codes written in MATLAB 7.0 (MathWorks Inc.) were utilized for the model pre-processing summarized above and to call the algorithm written in C++ for reporter reaction identification. Flux Balance Analysis was performed using in-house software BioOpt employing LINDO API for linear optimization. A software package which works under MATLAB 7.0 and includes 13 scripts was developed to automate the preprocessing steps (Figure 5.2), and reporter reaction algorithm. This package, named as RepRxn MX, is detailed in Appendix G. Deconvolution of peaks in GC-MS spectra for the identification of metabolites based on a metabolite library and for

the random peak assignment was achieved using AMDIS software (Stein *et al.*, 1999), and the peak normalization software was kindly provided by J. F. Moxley.

5.2. Model Preprocessing

Due to the large chemical diversity of the metabolome there is currently no single analytical method that enables analysis of the complete metabolome. Even the best analytical methods reported to date for metabolome analysis therefore only cover a small fraction of the metabolites present in genome-scale metabolic models. The unavailability of data for a large number of metabolites is one of the major problems associated with mapping (and hence integration) of metabolome data on to genome-scale metabolic networks. In order to overcome this fundamental problem, the genome-scale model of Forster *et al.* (2003) is pre-processed so as to obtain a reduced model where the fraction of experimentally measured metabolites was enriched. This processing was done by systematically eliminating unmeasured metabolites from the metabolic network. It should be noted that the model pre-processing is dependent on the metabolome data that are available, and the pre-processing will have to be done for each case. However, following the flow-chart depicted in Figure 5.2 this pre-processing is relatively straight forward and can easily be done also for other metabolic networks.

The yeast genome-scale model includes three compartments (mitochondria, cytosol and external space) with 844 metabolites (559 cytosolic, 164 mitochondrial, 121 external) and 1175 reactions (Förster *et al.*, 2003). Within the context of this model, metabolites present in more than one compartment are treated as if they are different entities in each compartment. However, the experimental data used in this analysis (and most of the datasets available to date) can only differentiate between extracellular and intracellular space. Since metabolite levels in different cellular compartments are not available, the cytosolic/mitochondrial compartmentation of the model was removed and corresponding metabolites were represented as one, with their corresponding reactions conserved. Also, there are a number of duplicate reactions due to the presence of isoenzymes in the model, and these reactions were lumped into single reactions since metabolome data alone does not provide information that enables distinction between the operations of different isoenzymes. As a result, the ‘processed’ model (Uncompartmented model, UNCOMP)

consists of 677 metabolites (559 internal, 118 external) with 725 reactions, including transport reactions. With this model the experimental data used here amount to about 12 per cent of these 677 metabolites (52 internal, 32 external).

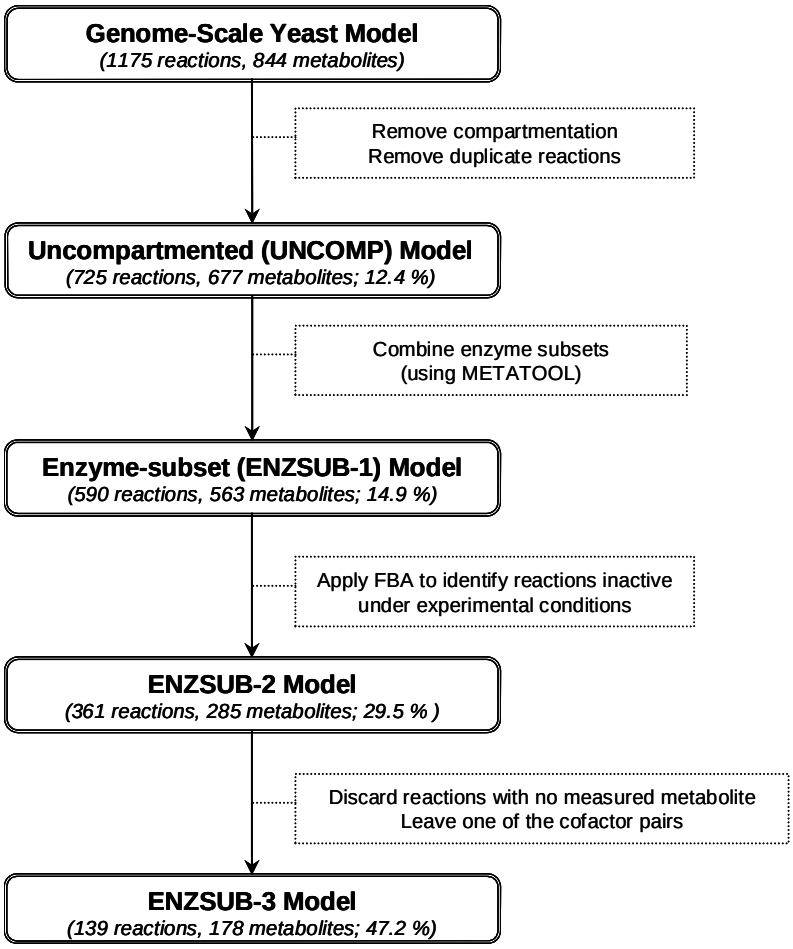


Figure 5.2. The preprocessing of the model to reduce the fraction of unmeasured metabolites and to focus on reactions involving measured metabolites. Percentages indicate the fraction of measured metabolites in each model

Enzyme subsets are enzymes that always operate together in fixed flux proportions at steady state (Pfeiffer *et al*., 1999; Schuster *et al*., 2002b), often representing enzymes in linear pathways. Accordingly, the intermediate metabolites in enzyme subsets can be assumed to be similarly affected by the perturbations. The uncompartmented model (UNCOMP) was further reduced in size by using METATOOL 4.3 (Pfeiffer *et al*., 1999; Dandekar *et al*., 2003) and thus representing each enzyme subset as a single reaction. The

resulting model (Enzyme-subset model, ENZSUB-1) consists of 563 metabolites and 590 reactions and it has about 15 per cent of the metabolites measured within the data used. Since the removal of the metabolites in linear pathways also led to the omission of six measured metabolites, the reactions containing these metabolites were restored back into the ENZSUB-1 model. To further increase the fraction of the measured metabolites, potentially inactive (or potentially low flux) reactions were removed. This was done by using Flux Balance Analysis (FBA) (Varma & Palsson, 1994; Kauffmann *et al*., 2002) for simulation of fluxes at specific environmental conditions used in the experiments (aerobic and anaerobic batch cultivation in glucose-limited minimal media). ENZSUB-1 model was used to simulate the fluxes with the objective of optimum growth. Then, the maximum and the minimum flux for each reaction in the model were obtained by constraining the specific growth rate between its optimum value and 50 per cent of the optimum. Reactions that had zero flux in the FBA analysis (at both optimum values) were considered as potentially invariant between the studied perturbations and thus omitted from the ENZSUB-1 model. The resulting model had 349 reactions involving 267 metabolites. The here-used FBA-based approach for model reduction does not necessarily imply that the eliminated reactions are inactive and that the metabolites involved in these reactions not present in the cell. However, it is assumed that as these reactions are likely to carry very low fluxes under the studied conditions, the associated metabolite pools are likely to be weakly affected due to changes in the fluxes through these reactions. Although this approach is useful, the assumption is not fool-proof as certain measured metabolites were intermediates in pathways with zero fluxes (Pimelic Acid, PIMExt, Myristic Acid, C140xt, trans-4-hydroxy-L-proline, Itaconate, Nicotinate, 4-Aminobenzoate, THMxt). The first six of these metabolites were detected as ‘invariant’ by the FBA approach due to the fact that that these metabolites are not connected to the overall network (Förster *et al*., 2003). However, here reactions involving these measured metabolites were restored back, and the resulting model comprised a total of 285 metabolites participating in 361 reactions (Figure 5.2). Even though certain reactions may be removed from the analysis by using this approach, the algorithm will still correctly identify reporter reactions, given the metabolome dataset. The resulting metabolic network, ENZSUB-2 model, was substantially enriched in terms of the content of measured metabolites (now accounting for about 30 per cent).

In order to further focus the analysis only on reactions involving measured metabolites, ENZSUB-3 was constructed by keeping only reactions that involved at least one measured metabolite. Additionally, only one member of the NADH/NAD^+, NADPH/$NADP^+$, $FADH_2$/FAD^+ cofactor pairs, when available, was retained in the remaining reactions since the levels of members of each pair were assumed to be interdependent. The resulting metabolic network, ENZSUB-3, included a total of 178 metabolites participating in 139 reactions, which corresponds to more than 47 per centt of the available quantitative metabolome data (Figure 5.2). The 139 reactions included in the model are given in the Appendix F.1.

The significance of change in the levels of metabolites between any two conditions was calculated by applying a statistical test (see methods section). However, it is difficult to deduce which reactions in the cell are affected most by only judging the significance of the change in metabolite levels, since the number of the metabolic reactions in the cell is high and one metabolite usually appears in more than one reaction. Thus, a normalized z-score for each reaction was calculated based on the z-values of its neighboring metabolites (p-values of individual metabolites were converted to z-scores by using inverse normal cumulative distribution function, see methods section). Here it is assumed that the calculated reaction z-scores can be regarded as an indicator of the significance of how the reactions respond to the studied perturbation at metabolic level. This assumption is based on the fact that metabolite levels are governed by changes in fluxes and enzyme activities (Nielsen, 2003). Reactions exhibiting significant changes (typically $z > 1.28$, corresponding to $p < 0.10$) for the perturbations analyzed were identified by using the graph representation of the derived metabolic model, ENZSUB-3, and listed in Table 5.1 and Table 5.3. A loose cut-off was deliberately chosen. The rason was not to be too-biased in the light of the fact that measurements were not available for all of the metabolites in the model, and thus the resultant p-values are in fact, in general, shifted to high values due to randomly selected p-values for those unmeasured metabolites.

5.3. Effect of an altered redox metabolism and oxygen availability

As a first demonstration of the developed approach, data from metabolome analysis of two different *S. cerevisiae* strains was considered. The strains were a wild type

laboratory strain (CEN.PK.113-7D) and a redox engineered strain, which was carried out in batch cultures under two different environmental conditions (aerobic and anaerobic) in standard mineral media with glucose as the sole carbon source (Villas-Bôas *et al*., 2005a). The redox engineered strain carrying a deletion of the NADPH-dependent glutamate dehydrogenase encoded by *GDH1* and an over-expression of the NADH-dependent glutamate dehydrogenase encoded by *GDH2* was constructed by dos Santos *et al*. (2003). Three different perturbations were analyzed here: genetic change under both aerobic and anaerobic conditions (wild type versus redox engineered strain), and environmental change for the wild type strain (aerobic versus anaerobic). Since it was reported that sample-to-sample variability exceeds flask-to-flask variability, replicate samples from different shake flasks were treated equivalently (Villas-Bôas *et al*., 2005a). Accordingly, the metabolome dataset includes around 15 intracellular and nine extracellular replicates for each experimental condition. The dataset used in this study is available in the supplemental material as normalized abundances of GC-MS peaks.

Comparison of the wild type and mutant strains revealed that the genetic changes do not alter the basic growth characteristics in aerobic (dos Santos *et al*., 2003) and anaerobic (Nissen *et al*., 2000) batch cultivations. The here-developed approach, however, captures the associated changes in different cellular pathways by identifying a number of significantly affected reactions due to these perturbations. The detected reactions (Table 5.1) belong to many different amino acid pathways, indicating a widespread effect of the mutation on the cellular metabolism. The present integrated approach also differentiates between the genetic perturbation under aerobic and anaerobic conditions as there are reactions that are specific to each condition.

Genetic perturbations (wild type versus redox engineered) used in the present study are directly related to a changed redox metabolism. Environmental perturbation (aerobic versus anaerobic) is, however, also associated with a changed redox metabolism due to the direct effect of oxygen availability on the operation of the TCA cycle and the pentose phosphate pathway, and hence on the redox state of the cell. This is also reflected in the identified reporter reactions since a number of common significantly changed reactions are observed for the two different types of perturbation (Table 5.1, Table 5.2).

Table 5.1. Reactions with significant z-scores (z > 1.28) in response to genetic perturbations by altered redox metabolism and environmental perturbation[a,b,c]. The number of measured metabolites and the total number of metabolites for each reaction are also given in parentheses

Genetic Perturbation (aerobic)			*Genetic Perturbation (anaerobic)*			*Environmental Perturbation (wild type strain)*		
VALsyn	(4/4)	2.90	AGX1	(4/4)	2.67	UGA^{ES}	(5/5)	2.41
ALT	(4/4)	2.83	ALT	(4/4)	2.35	ALT	(4/4)	2.34
LEUsyn^{ES}	(5/6)	2.66	**PROsc**	(2/3)	2.08	AGX1	(4/4)	2.34
TYRsyn	(3/4)	2.54	LEUsyn^{ES}	(5/6)	1.80	CAR2	(3/4)	1.95
CAR2	(3/4)	2.50	ASP3-1	(2/3)	1.78	LEUsyn^{ES}	(5/6)	1.95
PHEsyn^{ES}	(3/5)	2.25	**U46_**	(3/4)	1.64	TYRsyn	(3/4)	1.92
AGX1	(4/4)	2.01	**CHA1p**	(2/3)	1.58	VALsyn	(4/4)	1.87
AAT	(4/4)	1.86	PHEsyn^{ES}	(3/5)	1.57	PHEsyn^{ES}	(3/5)	1.74
ILEsyn^{ES}	(6/7)	1.77	**PUT1**	(2/3)	1.55	SERsyn^{ES}	(4/6)	1.67
SUCsc	(2/3)	1.66	VALsyn	(4/4)	1.54	**GAD1**	(2/3)	1.47
SDH	(2/3)	1.63	**GLY1**	(2/3)	1.50	**GDH13**	(3/4))	1.44
HISsyn^{ES}	(4/10)	1.58	SERsyn^{ES}	(4/6)	1.41	ASP3-1	(2/3)	1.39
ASP3-1	(2/3)	1.57				GDH2	(3/4)	1.38
GDH2	(3/4)	1.55				**MYRsc**	(2/2)	1.36
DLD	(2/4)	1.51				ILEsyn^{ES}	(6/7)	1.36
UGA^{ES}	(5/5)	1.48				HISsyn^{ES}	(4/10)	1.34
SERsyn^{ES}	(4/6)	1.46				**GLYsyn**	(2/4)	1.30
LEU4	(2/4)	1.36				**U155_**	(4/4)	1.29
FUM	(2/2)	1.28						

[a]Reactions specific to each perturbation are given in bold letters.
[b]ES means that the corresponding reaction is an enzyme subset consisting of combination of more than one reaction.
[c]sc in some of the reaction names stands for 'secretion', indicating that they are secretion reactions.

The glutamate decarboxylase reaction (GAD1) appears as a significantly changed reaction specific to the environmental perturbation of the wild type cells, which implies a major role of this reaction during respiratory growth (Table 5.1). Indeed, it was reported (McCammon *et al.*, 2003) that the defects in any of the 15 TCA cycle genes, associated with the slowing down of the respiratory metabolism, result in a substantial decrease in the mRNA levels of *GAD1*, which is in agreement with the findings reported here. GAD1

constitutes the first step of the glutamate catabolic pathway towards succinate (Coleman *et al*., 2001). The downstream steps of the pathway are catalyzed by Uga1p and Uga2p (UGA^{ES}), which are affected most by the environmental perturbation (Table 5.1). Detection of all reactions of this pathway (GAD1, UGA^{ES}) as responsive to the oxygen availability (Figure 5.3.a) indicates that they have a key role in succinate production via glutamate under anaerobic conditions where the yeast is secreting succinate. In fact, this pathway was found to be activated during oxidative (Coleman *et al*., 2001) or osmotic (sugar) (Erasmus *et al*., 2003) stress to control the redox balance of the cell.

Table 5.2. Grouping of the reactions with significant z-scores (z > 1.28) given in Table 5.1. Reactions common to all perturbations and perturbation-specific ones are grouped separately. Complete reactions can be followed from Appendix F.1

Reactions			*Genetic. Perturb. (aerobic)*	*Genetic Perturb. (anaerobic)*	*Environm. Perturb. (wild type)*
Reactions common to all perturbations					
VALsyn	oival ↔ val, *TA*	(4/4)	**2.90**	**1.54**	**1.87**
ALT	pyr ↔ ala, *TA*	(4/4)	**2.83**	**2.35**	**2.34**
$LEUsyn^{ES}$	glu + ippmal → leu, *TA*	(5/6)	**2.66**	**1.80**	**1.95**
$PHEsyn^{ES}$	prph → phe, *TA*	(3/5)	**2.25**	**1.57**	**1.74**
AGX1	gox + ala ↔ gly + pyr	(4/4)	**2.01**	**2.67**	**2.34**
$SERsyn^{ES}$	p3g → ser, *TA*	(4/6)	**1.46**	**1.41**	**1.67**
ASP3-1	asp → asn	(2/3)	**1.57**	**1.78**	**1.39**
Reactions common to two perturbations					
TYRsyn	4hpp → tyr, *TA*	(3/4)	**2.54**	0.95	**1.92**
CAR2	orn → glusal, *TA*	(3/4)	**2.50**	0.90	**1.95**
$ILEsyn^{ES}$	obut + pyr → ile, TA	(6/7)	**1.77**	1.07	**1.36**
$HISsyn^{ES}$	…. → his, *TA*	(4/10)	**1.58**	0.68	**1.34**
GDH2	glu → akg	(3/4)	**1.55**	1.13	**1.38**
UGA^{ES}	Gaba → succ, *TA*	(5/5)	**1.48**	0.70	**2.41**
Perturbation-specific reactions					
AAT	oac ↔ asp, *TA*	(4/4)	**1.86**	0.73	1.12
SUCsc	→ succE	(2/3)	**1.66**	0.27	1.26

Table 5.2. continued

SDH	succ ↔ fum	(2/3)	**1.63**	0.31	1.00
DLD	lac → pyr	(2/4)	**1.51**	1.18	1.11
LEU4	oival → ippmal	(2/4)	**1.36**	0.39	0.37
FUM	fum ↔ mal	(2/2)	**1.28**	0.27	0.57
PROsc	→ proE	(2/3)	1.24	**2.08**	1.26
U46_	thr → ac	(3/4)	0.92	**1.64**	0.98
CHA1p	ser → pyr	(2/3)	1.13	**1.58**	1.25
PUT1	pro → pr5carb	(2/3)	0.97	**1.55**	0.86
GLY1	acal + gly → thr	(2/3)	1.07	**1.50**	1.03
GAD1	glu → gaba	(2/3)	0.13	0.70	**1.47**
GDH13	akg → glu	(3/4)	1.12	0.53	**1.44**
MYRsc	myrist → myristE	(2/2)	0.20	0.51	**1.36**
GLYsyn	ser ↔ gly	(2/4)	0.55	1.22	**1.30**
U155_	akg+ malE ↔ mal +akgE	(4/4)	0.53	0.94	**1.29**

*TA: transaminase activity (conversion of glutamate to alpha-ketoglutarate)

Although the glyoxylate cycle is generally believed to be repressed during growth on glucose, Villas-Bôas *et al*. (2005b) found that an alternative pathway for glyoxylate biosynthesis is active in *S. cerevisiae*. Examination of the z-scores of reactions involving glyoxylate for all the analyzed perturbations revealed that AGX1 (reaction of enzyme encoded by *YFL030w*), which enables synthesis of glyoxylate from glycine, has much higher scores for all the perturbations compared to the reactions of the glyoxylate pathway (ICL and MLS) (Figure 5.3.b). Thus, the analysis made here supports the presence of an alternative pathway catalyzed by AGX1 leading to the biosynthesis of glyoxylate from glycine.

Reporter reaction analysis also identifies that the genetic perturbation results in metabolic changes around the genes that are perturbed (Figure 5.3.c). Thus, the reaction responsible for the over-expressed gene in the redox-engineered strain, GDH2, has a significant z-score for the genetic perturbation under aerobic condition. It should be mentioned that a genetic perturbation of a gene should not necessarily result in that the corresponding reaction comes out as a reporter reaction, as certain genetic perturbations may lead to only small changes in metabolite levels. However, in this case there are two

genetic modifications around α-ketoglutarate and glutamate (deletion of *GDH1* and over-expression of *GDH2*) which leads to identification of GDH2 as reporter. For the genetic change under anaerobic conditions, the detected significance of GDH2 is comparably lower. However, an indirect effect of the genetic modification in the glutamate biosynthesis can be observed from the presence of transaminase activity associated with some of the identified reporter reactions for this perturbation (conversion of glutamate to α-ketoglutarate by ALT, LEUsyn^{ES}, PHEsyn^{ES}, VALsyn, SERsyn^{ES}, Table 5.1, Table 5.2). On the other hand, the aerobic-anaerobic shift for the wild-type gives rise to nearly the same z-score for GDH2 reaction as the genetic perturbation under aerobic conditions. One explanation for this similarity in behavior would be that oxygen availability may have a direct effect on glutamate dehydrogenase genes; that is, cessation of oxygen uptake or manipulation of redox metabolism may result in similar effects on this node in the metabolism. In fact, in chemostat cultures, *GDH2* is associated with a significant transcription change when subjected to the same environmental perturbation (Piper *et al*., 2002). On the other hand, it is not possible to make a definite interpretation about the effect of the mutation on the deleted gene, *GDH1*, by looking at the z-score of GDH13 reaction since the reaction catalyzed by Gdh1p is identical with that catalyzed by Gdh3p. Consequently, what is reflected by this z-score is the 'combined' response of these two enzymes. The reason that the GDH13 reaction is not identified as a reporter reaction whereas the GDH2 reaction is identified can only be explained by either a different response in the co-factor level as a consequence of the perturbations, i.e. the NADPH/NADP$^+$ levels do not change as much as the NADH/NAD$^+$ levels, or due to measurement errors of these co-factors (these co-factors are inherently difficult to measure).

Since TCA cycle activity is known to be low under anaerobic conditions, associated effect of genetic mutation under this condition is expected to be weaker than the other two perturbations analyzed. The z-scores for the SDH and FUM reactions (both being part of the TCA cycle) are clearly in agreement with this expectation (Figure 5.3.d). These two reactions are also members of the electron transport system, and this further explains why the metabolites surrounding these reactions exhibit remarkably weaker coordinated change in the genetic perturbation under anaerobic condition than in the other perturbations.

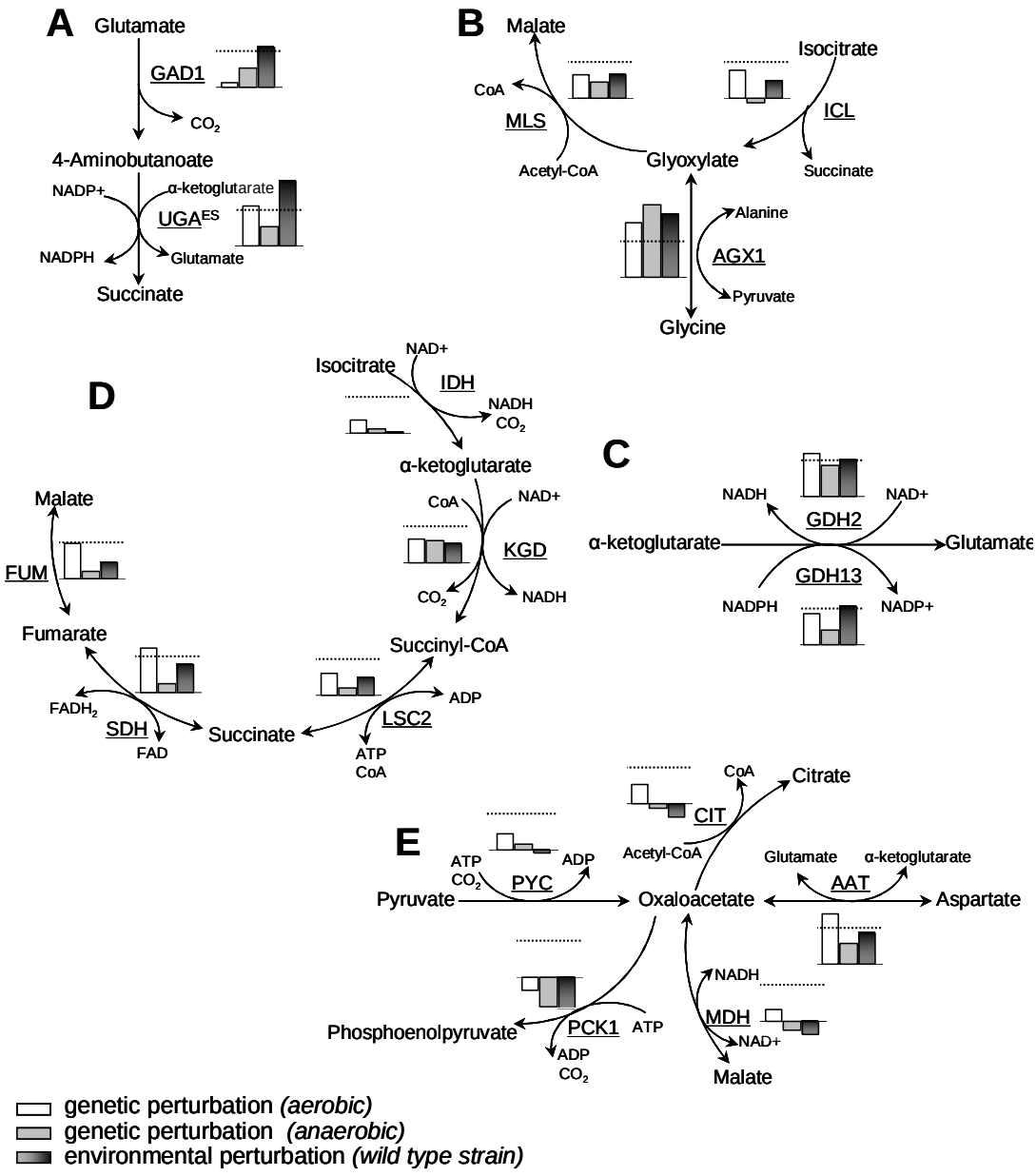


Figure 5.3. Example pathway structures based on z-scores of reactions, which demonstrate the metabolomic response of the selected reactions for the effect of an altered redox metabolism and aerobic/anaerobic growth. The dashed lines correspond to the cut-off of 1.28 (p = 0.10)

Similarly, the z-scores of key reactions involving oxaloacetate suggest that these reactions are mainly affected in the redox engineered strain under aerobic conditions (Figure 5.3.e), and AAT, a transamination reaction leading to the conversion of

oxaloacetate to aspartate, appears to be the key reaction where oxaloacetate is involved. There is no literature data available about the effect of the genetic perturbation on this metabolic reaction but as the genetic perturbation results in a changed ratio of glutamate to 2-oxoglutarate (Villas-Bôas *et al.*, 2005a) that may have effected this important transamination reaction.

5.4. Effect of very-high-gravity fermentation

As a second demonstration of the developed approach, metabolome data from two different *S. cerevisiae* strains, a laboratory strain (CEN.PK.113-7D) and an industrial strain used for fuel ethanol production (hereafter termed as "Red Star"), was used. For both strains the data were obtained from anaerobic batch cultures under two different cultivation conditions; exponential growth in a glucose containing standard mineral media and the stationary phase in a maltodextrin containing very-high-gravity (VHG) mineral media (Devantier *et al.*, 2005a). Environmental perturbations obtained through variation in the media were analyzed here for each strain. The intracellular metabolome dataset includes four replicates for the standard medium and eight replicates for the VHG medium. The extracellular metabolome dataset has six replicates for each condition. The complete dataset is available in the supplemental material.

As for the first case study discussed above, the two media perturbations analyzed revealed the same trend for the glyoxylate reactions, pointing to substantial regulation of the AGX1 reaction node in both perturbations (data not shown). In case of the glutamate metabolism, all the reactions have noticeably higher z-scores, except GDH2, implying that this pathway is highly affected by VHG associated media changes. All of the TCA cycle reactions shown in Figure 5.3.d have very low z-scores, in accordance with the fact that the cycle is barely operational under any of the experimental conditions studied (anaerobic fermentations). For reactions involving oxaloacetate, AAT again appears to play the major role as observed in the first data set, in parallel with the graph shown in Figure 5.3.e.

The reaction governed by Gad1p, which catalyzes decarboxylation of glutamate – a reaction that is generally considered to be associated with stress, is found to be significantly changed in both strains when the media was changed (Table 5.3). A

noticeably lower score was obtained for comparison of the two strains grown on the standard medium (results not shown), which shows that the standard medium imposes less stress compared with the VHG medium where sugar and ethanol stresses are predominant. The appearance of all reactions (GAD1, UGA^{ES}) involved in the glutamate catabolic pathway as reporter reactions when the media is perturbed (Table 5.3) points to the fact that this perturbation has a major effect on the amino acid metabolism, and probably also on the redox balance in the cell. The results of transcriptome analysis for the same strains in standard and VHG media (Devantier *et al.*, 2005b) indicate that the strains have differences in their redox balancing, confirming the finding reported here.

A large number of transport reactions were found to have significant z-scores (Table 5.3). GC-MS analysis of extracellular metabolites in the VHG medium revealed many more metabolites compared to what is found in the standard medium, explaining the appearance of transport reactions as significant. The here-reported algorithm allowed us to identify and quantify the secretion reactions which are mostly affected from the media change, by integrating both intracellular and extracellular measurements to the reaction network. Secretion of a number of amino acids (glutamate, aspartate, proline, alanine and glycine), and succinate, pyruvate and lactate are commonly and significantly regulated in response to media perturbation for both the laboratory and the red star strain. On the other hand, detection of strain-specific secretion patterns (valine, citrate and alpha-ketoglutarate, Table 5.3) points to differences in operation of the metabolic network in the two strains, possibly arising from the difference in the redox metabolism of the two strains.

Since the change in the fermentation medium led to ethanol and osmotic stress for both strains (Devantier *et al.*, 2005a), it is not surprising that many of the reactions are shared in the identified lists for the two strains in the media comparison (Table 5.3). Transcriptome analysis of this dataset revealed that a substantial part of the significantly changed genes were involved in protein synthesis and amino acid metabolism (Devantier *et al.*, 2005b). Thus, amino acid pathway reactions detected by reporter reaction analysis (Table 5.3) are in accordance with the transcriptome data. Absence of amino acid synthesis in VHG media due to the cessation of growth in the stationary phase can be a possible cause of the observed differences.

Table 5.3. Effect of media change (standard vs. VHG medium) on each strain. Reactions with significant z-scores (z > 1.28) are shown[a,b]. z-scores of gene expression changes are also given. z_{RE}: z-scores of reactions calculated by the developed approach, z_{GE}: z-scores of genes/gene groups calculated from asscociated p-values from transcriptome data.

Media Change for laboratory strain (CEN.PK113-7D)				**Media Change for industrial strain (Red Star)**			
		z_{RE}	**z_{GE}**			**z_{RE}**	**z_{GE}**
ALT	(4/4)	2.50	2.48	ALT	(4/4)	2.48	1.80
AGX1	(4/4)	2.45	0.86	AGX1	(4/4)	2.31	2.21
UGA [ES]	(5/5)	2.18	1.69	UGA [ES]	(5/5)	2.23	0.38
ECM40	(3/4)	1.85	2.39	**U155_**	(4/4)	2.01	-
GLUsc	(2/3)	1.85	1.17	ASN	(4/7)	1.85	0.54
ASN	(4/7)	1.84	2.30	**TYRsyn**	(3/4)	1.84	3.41
CAR2	(3/4)	1.74	0.57	GLUsc	(2/3)	1.81	0.79
LYSsyn [ES]	(7/8)	1.67	2.46	PHEsyn [ES]	(3/5)	1.65	0.98
TRP23	(3/5)	1.67	1.31	TRP23	(3/5)	1.65	0.78
ASP3-1	(2/3)	1.47	1.45	PROsc	(2/3)	1.45	0.90
CHA1p	(2/3)	1.47	0.93	ALAsc	(2/3)	1.45	0.75
U42_ -43_	(2/3)	1.47	-	GLYsc	(2/3)	1.45	0.80
ASPsc	(2/3)	1.43	2.09	LACsc	(2/3)	1.45	0.86
PROsc	(2/3)	1.43	1.40	PYRsc	(2/3)	1.45	0.86
ALAsc	(2/3)	1.43	1.90	SUCsc	(2/3)	1.45	-
GLYsc	(2/3)	1.43	1.66	**CITsc**	(2/3)	1.45	-
LACsc	(2/3)	1.43	0.81	**AKGsc**	(2/3)	1.45	-
PYRsc	(2/3)	1.43	0.81	**U88_**	(2/3)	1.43	-
SUCsc	(2/3)	1.43	-	GAD1	(2/3)	1.43	1.21
GLY1	(2/3)	1.41	0.41	**ILEsyn [ES]**	(6/7)	1.42	1.90
VALsc	(2/3)	1.38	1.55	ASP3-1	(2/3)	1.41	1.34
PHEsyn [ES]	(3/5)	1.36	1.66	U42_-43_	(2/3)	1.41	-
GAD1	(2/3)	1.29	1.46	**LEUsyn [ES]**	(5/6)	1.38	0.91
				ASPsc	(2/3)	1.29	0.86

[a] Reactions specific to each perturbation are given in bold letters.
[b]Number of measured metabolites and total number of metabolites for each reaction are also given in parentheses

5.5. Integration of metabolome data with transcriptome data for understanding regulation

For the latter case study, where the effect of a VHG medium was analyzed on the metabolome of laboratory and industrial strains, there was also performed genome wide expression analysis (Devantier *et al*., 2005b). This basically enables further evaluation of mode of regulation for the different reactions in the reduced metabolic network. ter Kuile & Westerhoff (2001) introduced the concept of metabolic regulation and hierarchical regulation, where the first indicates that regulation of flux is at the level of enzyme kinetics, i.e. through changes of the metabolite levels, and the second indicates that regulation of flux is at the level of enzyme production /activity (transcription/translation/post-translational modification). As both metabolite data and transcription data are available for this case study, it was investigated whether it was possible to identify the type of regulation at the individual reaction level. A major obstacle for this kind of analysis is, however, that information about changes in fluxes for the analyzed conditions is not available, and such data would also be difficult to obtain. Although there are efficient methods for obtaining data on the metabolic fluxes in the central carbon metabolism (Nielsen, 2003), it is difficult to get good estimates for the fluxes in all pathways of the metabolic network analyzed here, and even though the fluxes can be calculated by using flux balance analysis, this method is not well suited to give precise estimates for the actual fluxes in networks where there are redundant pathways. In order to proceed with analysis, it was therefore assumed that whenever there was a coordinated significant change in metabolite levels around a reaction, then it is very likely that the flux through this reaction is also changing. However, there is no guarantee that the flux through this reaction is also changed as there could also be a change in the enzyme concentration, or there could even be altered allosteric regulation of the enzyme, thus keeping the flux unchanged. Thus, the assumption made may result in identification of some false positives, but still the analysis would clearly lead to identification of reactions around which there is at least one level of regulation (and possibly several levels of regulation), and these reactions will therefore be referred as being metabolically regulated. For all the reactions that are not identified as reporter reactions one can not infer anything about whether the flux has changed, but still it can be deduced from the transcription data whether there has occurred regulation at the hierarchical level, and even though this does

not necessarily mean hierarchical regulation of the flux these reactions will be referred as being hierarchically regulated. This deduction can still be informative as indicator of the logic of transcriptional regulatory machinery governing gene expression. For cases where there was a significant change at the transcriptional level for an identified reporter reaction this was considered to be a situation where there was mixed regulation.

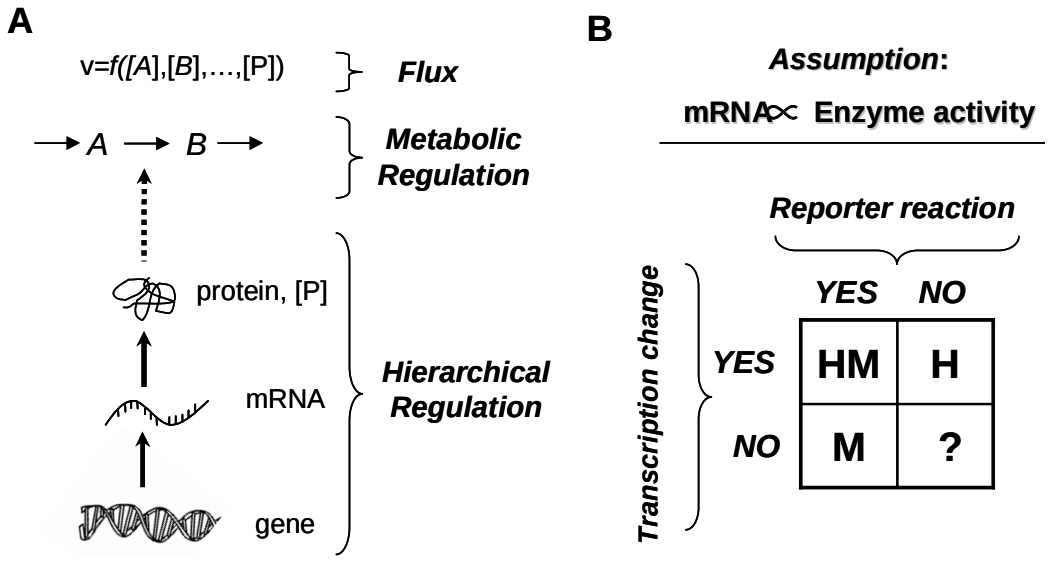


Figure 5.4. a) Major components of flux regulation b) Classification of (reporter) reactions with respect to regulation type. mRNA levels were assumed to reflect enzyme activities

The metabolic network includes several enzymes (hence reactions) governed by multiple genes. Thus, in order to infer about the significance of change in expression levels for the reactions the transcript levels for all genes coding for the same reaction were summed up before applying the statistical test. The p-values of transcripts were then calculated by using a t-test with unequal variance, and further converted into z-scores to enable a comparison with the z-scores of reactions based on metabolome data.

Using this approach all the reactions of the metabolic network was grouped into whether they were metabolically or hierarchically regulated (or a combination or not regulated at all) for the VHG dataset. To score the magnitude of the regulation at the hierarchical and metabolic levels, the corresponding z-scores were used. Hereby the qualitative evaluation of z-scores emerging from the transcriptome and the metabolome data enabled us to get an indication of regulation within the metabolic network (see Figure

5.4, Table F.1). The cases where only the transcript z-score is significantly changed can be scored as points with possible hierarchical regulation, whereas the opposite case where only the metabolite based z-score has significantly changed implies metabolic regulation of the corresponding reaction (Rossel *et al*., 2005). When both z-scores are significant there is regulation shared at both levels, and when none of the z-scores are significant, it is not possible to infer about at which level there is regulation.

Figure 5.5 and Table F.1 presents magnitude of the regulation for the reactions of the metabolic network, ENZSUB3, at the hierarchical and metabolic levels for the effect of very high-gravity (VHG) fermentation media on laboratory (CEN.PK113-7D) and industrial (RS) strains. z-scores calculated based on gene expression changes (zGE) and based on changes in the surrounding metabolites (zRE) are shown. Red means a positive z-score, and green means a negative z-score indicating that the regulation is insignificant. Reactions were color-coded with respect to their z-scores using $z = 1.28$ ($p = 0.10$) as the cut-off value to decide on the corresponding regulation type. *yellow:* hierarchically regulation. *black:* metabolically regulation. *violet:* mixed regulation. *white:* statistically insignificant score for both type.

Of the 121 reactions in the model having corresponding genes associated with them, the number of reactions predicted to be regulated hierarchically, metabolically, and at both levels were 56, seven, and 14 respectively for the media perturbation with the laboratory strain, and 31, 14, and five for the same perturbation with the industrial strain (Figure 5.5, Table F.1). For the laboratory strain, 44 reactions were found to be relatively irresponsive to the perturbation. On the other hand, the number of potentially unregulated reactions was much higher (71) for the industrial strain. One explanation for the observed predominance of transcriptional regulation could be the fact that the strains protect themselves against the applied perturbation by mainly changing their gene expression to minimize the changes in the metabolome; an observation also encountered in plants (Hirai *et al*., 2004). Figure 5.5 and Table F.1 suggest that metabolic regulation is mainly predominant for secretion reactions and amino acid pathways with or without simultaneous hierarchical regulation, the sole exceptions being proline and methionine/cysteine pathways. It is logical to identify the latter as subjected to different regulation since they are involved in pathways with

sulfur assimilation and there were no direct perturbation on sulfur utilization in the experimental study.

The type of regulation for a number of reactions differs between the two strains, which supports the finding that gene expression pattern can vary within different *S. cerevisiae* strains (Ferea *et al*., 1999; Brem *et al*., 2002; Townsend *et al*., 2003; Jansen *et al*., 2005). Ferea *et al*. (1999) have reported altered expression levels of genes involved in metabolite transport for strains obtained by adaptive evolution in glucose limited cultures. This observation presents an interesting analogy to our analysis, as the industrial strain is also likely to be a result of adaptive evolution. Similarly, different wild type strains were found to have widespread variations in expression of genes involved in amino acid metabolism (Townsend *et al.*, 2003). In order to further validate that the metabolism is different in the industrial and laboratory strains, we performed principal component analysis of the metabolome data for the VHG medium dataset (Figure 5.6). This shows a clear distinction of the strains indicating that the strains behave remarkably different at the level of metabolome. Our analysis systematically combines the transcriptome and metabolome and deduces the underlying regulation causing these differences in metabolism. Notably, following a change to a high-gravity fermentation medium, transcriptional regulation of metabolism is much more pre-dominant in the laboratory strain as compared to the industrial strain; whereas the number of reporter reactions between two strains is around the same with a 70 per cent overlap (Table 5.3). This strongly suggests that although the industrial strain has a better adaptation of its transcriptional program for high-gravity media, there is still similar metabolic regulation pattern to the laboratory strain. The difference in strains in terms of their response to the same perturbation is, again, very visible in the secretion reactions where laboratory strain attempts to regulate them also at transcriptional level, whereas industrial strain relies predominantly on metabolic control (Figure 5.5, Table F.1). The lesser degree of transcriptional regulation in the industrial strain could benefit the cells by reducing the investment of resources in transcriptional regulatory machinery.

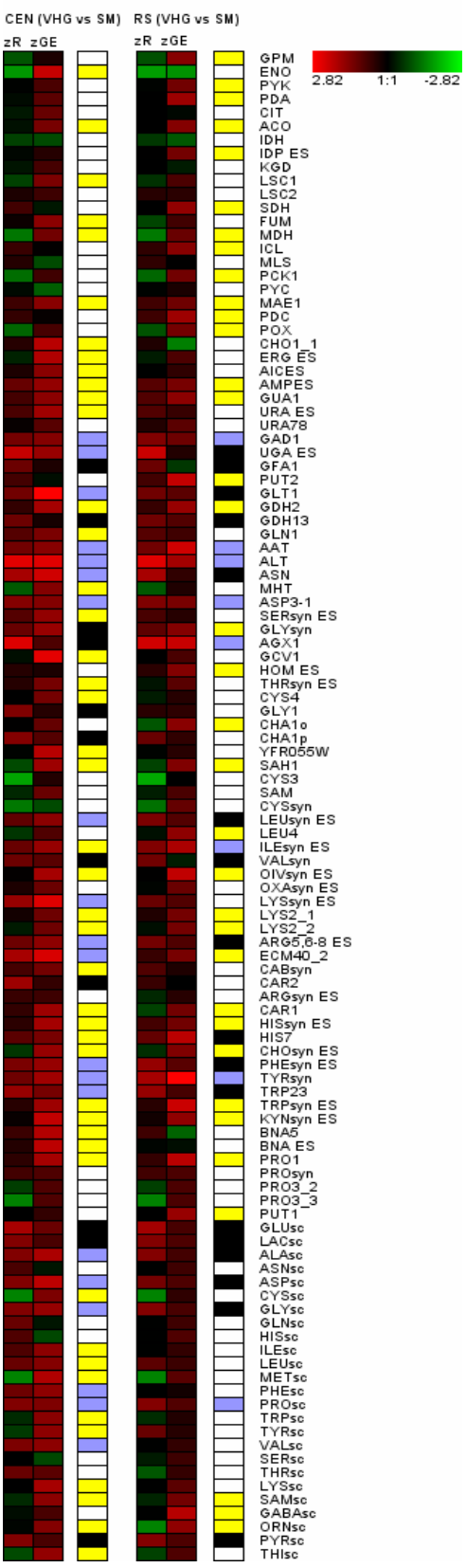


Figure 5.5. Magnitude of the regulation for reactions of the metabolic network, ENZSUB3

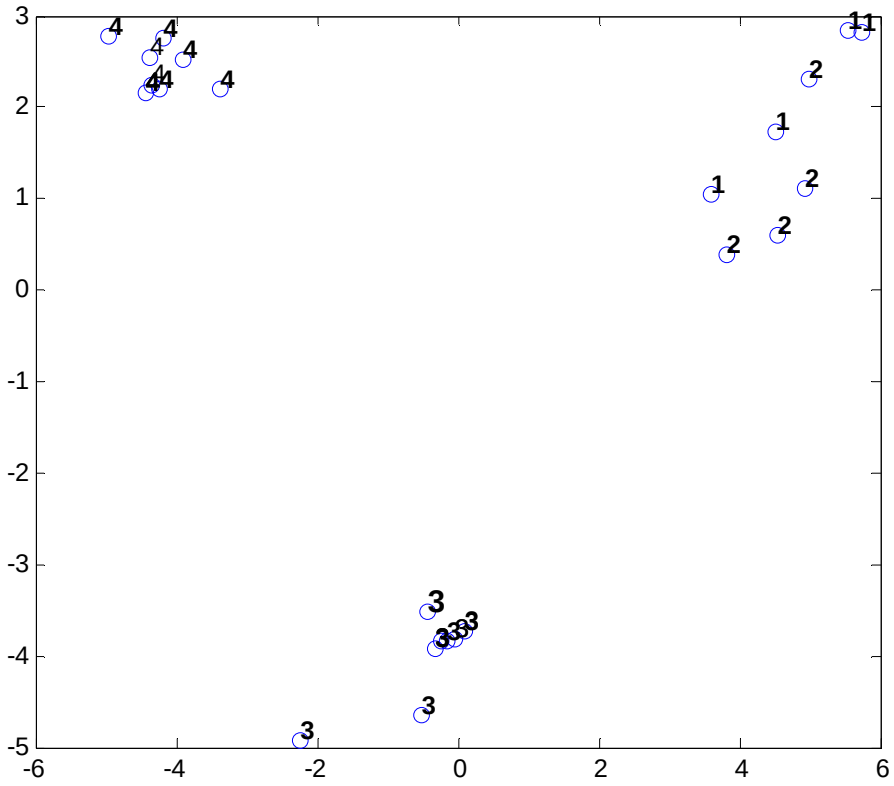


Figure 5.6. Principal component analysis of the dataset which involves the effect of very-high-gravity fermentation. 1: CEN.PK strain in standard media; 2: Industrial strain in standard media; 3: CEN.PK strain in VHG media; 4: Industrial strain in VHG media)

5.6. Low Coverage of Measured Metabolite Content by the Preprocessed Model

Due to the large chemical diversity of the metabolome there is currently no single analytical method that enables analysis of the complete metabolome. Even the best analytical methods reported to date for metabolome analysis therefore only cover a small fraction of the metabolites present in genome-scale metabolic models. To overcome this fundamental problem the genome-scale model was pre-processed as discussed in section 5.2. However, the final model still includes a number of unmeasured metabolites. Here, the effects and limits for the percentage of unmeasured (and hence modeled) metabolites on the applicability of the developed algorithm are discussed.

It is difficult to give a unique favorable ratio between measured and unmeasured metabolites as a threshold for the application of the algorithm. This is due to the fact that it

is not only the ratio that matters, but also how the measured metabolites are distributed throughout the metabolic network, i.e. if a lot of metabolites are measured but they all cluster within a certain part of the metabolic network, it is not possible to deduce anything about what is happening in other parts of the network. Even if the ratio is high (e.g. higher than 0.5) it may be problematic if the measured metabolites are distributed across different parts of the metabolic network, and there will therefore hardly be any metabolites that coexist in the same reactions, making the resultant z-scores more difficult to estimate reliably.

Additionally, an increase in the ratio between measured and modeled metabolites will result in the inclusion of some of the reactions which had been removed in the model reduction steps since none of whose participating metabolites were measured, and some of those reactions may happen to have significant z-scores and appear in the reporter list. On the other hand, the decrease in the ratio, namely in the number of measured metabolites, may lead to elimination of some of the reactions in the reporter list due to the low fraction of measured metabolites participating in a reaction.

To summarize, the proposed method is less likely to pick false-positives, but a change in the measured metabolite coverage will lead to a change in the number of false-negatives. That is, the decrease in the ratio will only lead to an increase in the number of false-negatives, but still we will have correctly detected reactions. And since the discussion of the results of the developed algorithm is centered on the reactions detected as significant, the change in the ratio of measured to unmeasured metabolites is not likely to affect the obtained results substantially, at least in the sense that the detected reactions will never include false positives.

To illustrate some of the above-discussed points, 10 measured metabolites out of 84 was randomly labeled as unmeasured and the reporter reaction algorithm was executed for the case where aerobic and anaerobic conditions for the wild type strain are compared. The original analysis had resulted in 18 reactions associated with significant z-scores. After randomly removing 10 measured metabolites from the analysis, it was observed that 2 reactions were removed from the the list of originally detected reactions, as some of their corresponding metabolites were unmeasured. In another random removal of 10 measured

metabolites, nine (out of the 18) of the significant reactions from the original list were deemed insignificant after removal of 10 metabolites.

Based on the features of the developed algorithm, certain guidelines can be provided for the metabolome measurements in order to effectively exploit the reporter approach; (i) Measurement of metabolites that participate in many reactions (hubs in the metabolic network) will certainly increase the coverage of the algorithm. (ii) Measurement of metabolites that participate in certain closely related pathways (metabolites that are closely placed in the network) will increase the confidence in the obtained z-scores for reactions in those pathways.

5.7. Concluding Remarks

An integrative algorithm based on metabolome data was introduced for the identification of reporter reactions, defined as the reactions that are responding to a genetic or environmental perturbation through a coordinated variation in the levels of surrounding metabolites. It is demonstrated that the algorithm functions even with a small number of measured metabolites (84), which is a typical situation for several currently used technologies. Moreover, the method developed is suitable for mapping the entire alterations associated with a specific perturbation, depending on the advances in analytical detection techniques enabling the measurement of a larger number of metabolites.

Furthermore, when integrated with transcriptome data the developed approach can be used to infer information about whether a reaction is metabolically regulated or whether it is hierarchically regulated. The approach can therefore be regarded as a genome-scale approach towards the integration of different types of omics data by using metabolic networks as a scaffold in order to understand the architecture of metabolic regulatory circuits.

6. METABOLIC PATHWAY ROBUSTNESS OF *S. CEREVISIAE*

Metabolic robustness is defined as the ability of the metabolic network to adjust its fluxes in response to environmental or genetic changes without changes to the phenotype (Edwards and Palsson, 1998). In a way, it is the resistance of cell metabolism against perturbations. Structural robustness of an organism can be determined to see whether a cell can tolerate the elimination of some enzymes by mutations (Wilhelm *et al.*, 2004). Several methods were suggested as *in silico* measures of metabolic robustness of cells, based on the lethality of the cells in response to the deletion of each metabolic gene (enzyme). These methods requires calculation of EFMs, the possible paths from substrates to products. If, after the deletion of an enzyme, there are still remaining EFMs with active biomass growth reaction, this mutant is predicted to be viable. Three of these measures are proposed by Wilhelm et al., 2004, in which calculated EFMs are used as the basis. The forth measure is based on a recently developed concept; minimal cut sets (MCS) (Klamt and Gilles, 2004). MCS calculation relies on pre-calculated EFMs, and can be defined as the minimal set of reactions whose inactivation will lead to guaranteed failure in certain cellular network functions. If the cellular network function to be tested is biomass growth, the failure means lethality.

6.1. In silico Metabolic Robustness Measures

Number of EFMs is an indicator of flexibility as well as redundancy. However, it is reasoned that redundancy is not directly identical with robustness since systems with the same number of EFMs were shown to have different robustness scores (Wilhelm *et al.*, 2004). Therefore, a measure must be defined which compares the entire system with the mutated system.

6.1.1. Overall Global Robustness

The structural robustness of a metabolic network to the knockout (deficiency) of one enzyme, E_i, can be calculated based on the fraction of remaining EFMs after its knock-out.

If EFMs in the original system is denoted as z, and that of the perturbed network is designated by z_i, the ratio z_i/z will be important. Since the aim is to calculate overall robustness of the network rather than the robustness to a specific enzyme knockout, this ratio must be calculated for all the enzymes E_i available in the considered reaction system. The arithmetic mean of all these ratios were suggested as a quantification of the global robustness of the entire metabolic network (Wilhelm *et al.*, 2004);

$$R_1 = \frac{\sum_{i=1}^{r} z^{(i)}}{r \times z} \tag{6.1}$$

where *r* is the total number of reactions in the system.

6.1.2. Product-based Minimal Robustness

This measure is based on the essential products of the reaction system for the organism of interest. Namely, if any of these essential products cannot be produced, the organism is non-viable. Therefore, product-specific robustness should be taken into account. This can be reflected into the robustness measure as follows; to calculate the robustness concerning product P1, only the EFMs which produce this product must be chosen, discarding the others. Then, the same formula introduced in Equation 6.1 can be applied to this set of EFMs. Product-based minimal robustness hypothesizes that the robustness of the metabolic network of interest is equal to the minimum of robustness values calculated for every available essential product (Wilhelm *et al.*, 2004).

$$R_2 = \min\{R_1^{P1},\ R_1^{P2},\ R_1^{P3}, \ldots.\ R_1^{Pn}\} \tag{6.2}$$

6.1.3. Product-based Global Robustness

It may happen that product-based robustness of one product may be quite low, but most of the random mutations would affect the EFMs producing the other products. This third masure takes this fact into account, by using the arithmetic mean of product-based robustness values rather than the minimum of them (Wilhelm *et al.*, 2004).

$$R_3 = \frac{\sum_{i=1}^{n} R_1^{P(i)}}{n} \tag{6.3}$$

where *n* denotes the number of products of the metabolic network.

6.1.4. Minimal Cut Set- based Robustness

Minimal cut sets, the minimal set of enzymes whose complete inhibition prevents the functioning of a pre-defined target reaction (eg. biomass formation), are calculated based on the EFMs of the system. The algorithm (Klamt and Gilles, 2004), also assigns a fragility coefficient to every enzyme in the model. Fragility of a reaction is defined as the reciprocal of the average size of all MCSs in which that reaction is involved. It is a number between zero and one. The closer the number to zero, it is less fragile, meaning more robust. Klamt and Gilles also define a network fragility coefficient, F, to enable an overall quantification of the structural fragility. F is defined as the average of the fragility coefficients over all reactions. Corresponding robustness score can be defined as (1 - F). Theoretically, this measure is more sound than the previous measures since it also takes multiple deletions (double, triple, quadrople,…) into account rather than only single deletions.

6.2. Robustness of Yeast Central Metabolism compared to *E. coli*

6.2.1. EFM and MCS Calculation

Minimal cut sets, were determined for a total of 16 fermentable and nonfermentable substrates such as glucose, acetate, glycerol, ethanol, in *Saccharomyces cerevisiae* and *Escherischia coli*, for the cellular objective of biomass production. The biochemical reaction set used covers central carbon metabolism of the yeast (54 reactions and 52 metabolites; Chapter 3; Çakır *et al.*, 2004). EFM calculation was done by METATOOL and FluxAnalyzer 5.0 softwares. The first three measures were calculated in MATLAB 7.0, whereas the calculation of the fourth measure was performed using FluxAnalyzer 5.0 with the following target reaction; biomass formation. To deduce relative robustness of yeast metabolism, the same approach was also applied to *E. coli* metabolism which consists of

52 reactions and 49 metabolites, adapted from Stelling *et al*., 2002 and Carlson and Srienc, 2004. The total number of reactions differs in two organisms since there are a few organism-specific metabolic reactions. The other major difference between the two microorganisms reflected into the reactions is that *S. cerevisiae* has a compartmentalized metabolism. Therefore, metabolites which cannot pass through the mitochondrial membrane are distinguished by treating them as different metabolites in cytosol and mitochondria (NADH, NADPH, Acetyl-CoA). Table 6.1 gives the substrates with corresponding number of calculated elementary flux modes and minimal cut sets for both microorganisms.

Table 6.1. Elementary flux modes and Minimal cut sets for *S. cerevisiae* and *E. coli* grown on 16 different substrates

	S. cerevisiae			*E. coli*		
	per cent			per cent		
Substrate	EFM	EFM-BIOM*	MCS	EFM	EFM-BIOM*	MCS
Glucose	15996	97.7	616	4532	51.4	541
Ethanol	1890	94.0	78	225	39.1	43
Acetate	932	89.5	90	443	46.0	98
Glycerol	4736	97.5	228	306	52.0	117
Succinate	5000	95.9	219	1204	61.4	248
Pyruvate	5879	95.0	269	970	54.0	222
Lactate	4816	96.3	198	666	55.3	214
2-ketoglutarate	2699	93.4	103	831	67.4	236
Malate	3481	95.3	173	829	53.6	173
Ribose	6783	93.4	461	907	25.1	170
Xylose	7105	93.7	492	1847	21.2	262
Erythrose	7731	94.3	707	946	27.9	264
Sedoheptulose	5689	92.2	480	1385	26.3	248
Fumarate	3481	95.3	165	829	53.6	156
Citrate	3730	94.4	154	1648	58.7	312
Oxaloacetate	3714	93.3	249	1391	52.8	245

* Percentage of biomass-producing EFMs. These EFMs were used in the calculation of MCSs.

Among many other substrates, glucose is the natural substrate for microorganisms. Therefore, microorganisms should have higher adaptation for growth in media containing

glucose. That is, they should exhibit more flexible growth on glucose than other carbon substrates. In accordance with this expectation, both microorganisms have highest number of EFMs when glucose is the substrate (Table 6.1). Additionally, theoretical results presented in the table reveals that yeast has noticably more flexible metabolism than *E. coli* in general because of having higher number of EFMs for the same substrates.

Figures 6.1 and 6.2 present the general trend of EFMs and MCSs for the studied substrates for both microorganisms. Calculated Pearson correlation between the numbers of EFMs and MCSs for yeast is 0.79 whereas there is a higher correlation for *E. coli* (0.92). Therefore, it can be concluded that although they are not directly proportional in a precise way, number of EFMs and MCSs are correlated. That is, the minimal sets of reactions which can perform a function (EFMs) are interrelated to the minimal set of reactions whose removal impedes a certain function (MCSs).

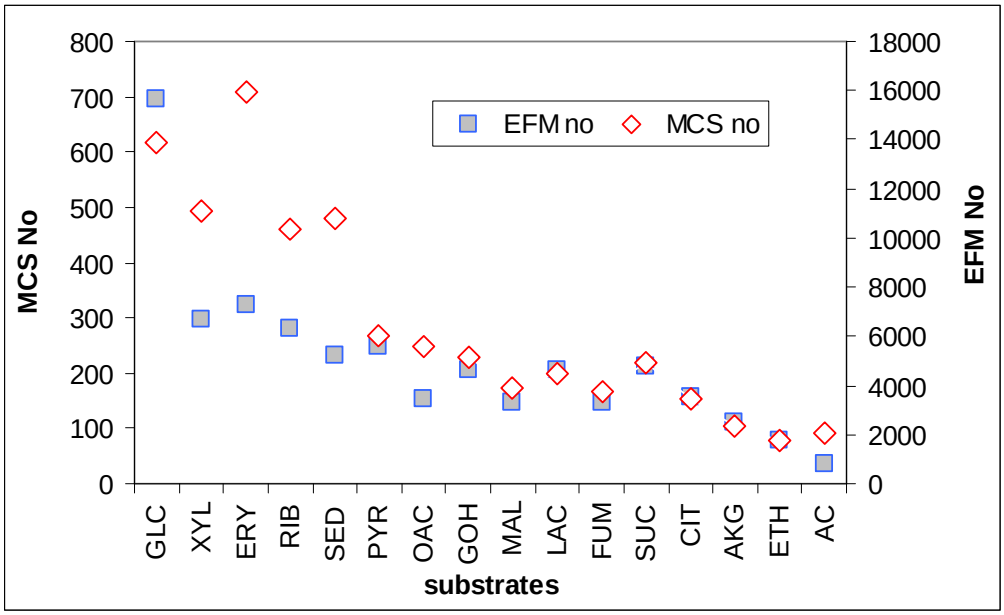


Figure 6.1. Comparison of calculated EFMs and MCSs for 16 different substrates of *S.cerevisiae*. The substrates in x-axis are ranked with respect to the corresponding robustness score in decreasing order

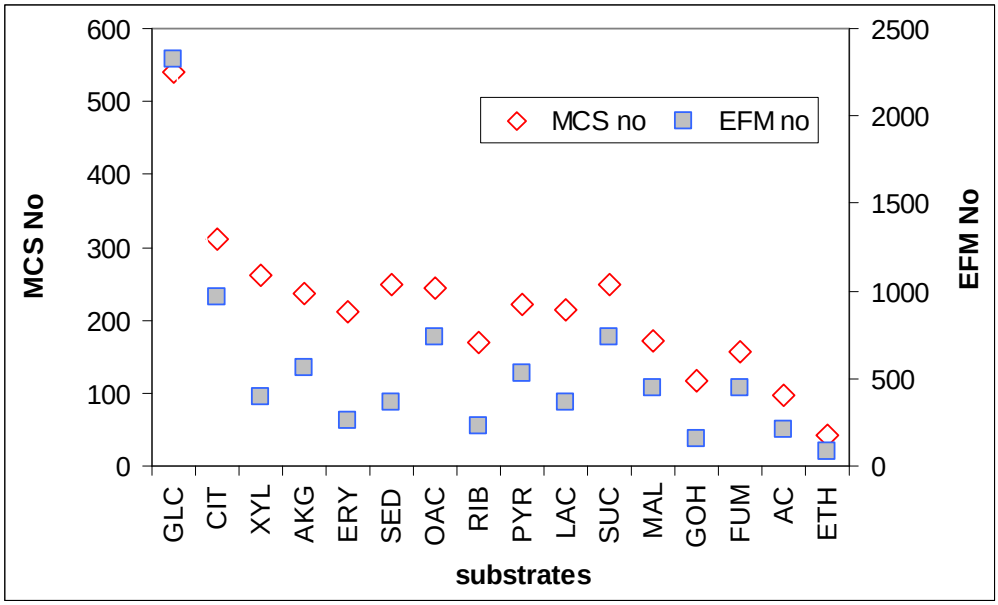


Figure 6.2. Comparison of calculated EFMs and MCSs for 16 different substrates of *E. coli*. The substrates in x-axis are ranked with respect to the corresponding robustness score in decreasing order

6.2.2. Results of Robustness Measures

Calculation of overall global robustness score (R_1) led to very close robustness values for each substrates in *S. cerevisiae*, with minimum score belonging to growth on ethanol (0.305), and maximum belonging to oxaloacetate and malate (0.331). Overall, these results suggest that yeast metabolism is prone to fragility, regardless of the substrate type (Table 6.2). For *E. coli*, on the other hand, the range of robustness values are wider. Ethanol is found to be the most robust substrate (0.512), with glucose being the most fragile one (0.412). The results are surprising since they are opposed to the biological expectation. Glucose is the most common substrate for *E. coli*, and the microorganism must be quite robust to genetic perturbations when it grows on this natural substrate. These findings put doubts on the credibility of this robustness measure. Additionally, the scores imply more robust behaviour of *E. coli* compared to S. *cerevisiae*.

Essential products of microorganisms can be considered as biomass production and ATP production for maintenance. Second measure (R_2) suggests the calculation of robustness for both objectives, and assigning the minimum of them as the robustness score.

Therefore, biomass-based robustness was calculated using only biomass-producing EFMs. Similarly, use of only reactions which utilizes maintenance reaction leads to maintenance-based robustness. Calculated maintenance-based robustness gives higher values for all substrates, whereas product-based ones are very similar to those obtained as R_1. For *S. cerevisiae*, the minimum of the two values gives scores in almost the same magnitude of R_1, ranging from 0.282 (ethanol growth) to 0.321 (erythrose and ribose). This measure again suggests a fragile yeast metabolism (Table 6.2). Whereas the previous measure (R_1) indicated a clearly more robust *E. coli* metabolism; product-based minimal robustness leads to noticeably closer scores to *S. cerevisiae* scores, ranging from 0.341 (glycerol) to 0.395 (malate). However, the scores are again higher than that belonging to yeast.

The third score is calculated by averaging maintenance- and biomass- based robustness scores. Interestingly, growth of yeast on acetate has the highest score in this case, whereas it was one of the lowest according to R_2. The most fragile (least robust) growth was found to be on glycerol (0.302). For *E. coli*, malate is again the most robust substrate. Results are given in Table 6.2.

In general, these three measures do not result in consistent results for yeast. The Pearson correlation between R_1 and R_2 is 0.74, between R_1 and R_3 is 0.24, and between R_2 and R_3 is -0.31. This means that, scores obtained by R_1 and R_2 measures are moderately correlated, whereas R_3 results in values not parallel with R_1 and R_2. This tendency is not the same for *E. coli*, where R_1 and R_2 are correlated with a Pearson value of 0.25; and correlation between R_2 and R_3, and R_1 and R_3 are 0.56 and 0.63 respectively.

One possible reason for the incapability of the first three measures to reflect differences between growth on different substrates is that they test only the effect of single deletions. MCS-based fourth score, on the other hand, takes all possible deletion mutants into account. Fourth measure depends on network fragility coefficient, which is automatically calculated by FluxAnalyzer. Compared to other scores, this score is more widely distributed between zero and one (Table 6.2), allowing the comparison of relative robustness of different substrates. Additionally, this measure predicts glucose as the most robust substrate for both microorganisms, in parallel with the biological expectation. The most fragile substrates, on the other hand, are ethanol for *E. coli* (0.230) and acetate for

yeast (0.297). In general, substrate robustness is found to be higher when the microorganism is yeast, according to the fourth measure. The three exceptions are for citrate, 2-ketoglutarate and acetate. As it is known, acetate is the natural by-product for *E. coli*, which becomes the first substrate in diauxic shift following glucose. For yeast, however, ethanol is the preferred substrate after diauxic shift. Acetate can only be produced in very low amounts by yeast. Therefore, yeast must be more fragile in acetate containing media. Theoretical finding is in accordance with this expectation. Figure 6.3 gives comparative robustness of substrates for yeast and *E. coli* depending on MCS-based robustness score.

Table 6.2. Robustness scores for both microorganisms calculated with four different measures

	R1		R2		R3		R4	
	S. cer.	*E. coli*	*S. cer.*	*E. coli*	*S. cer.*	*E. coli*	*S. cer.*	*E. coli*
Glucose	0.311	0.416	0.307	0.356	0.310	0.412	0.562	0.519
Xylose	0.326	0.457	0.319	0.388	0.329	0.433	0.545	0.486
Erythrose	0.327	0.451	0.321	0.375	0.326	0.426	0.541	0.472
Ribose	0.328	0.448	0.321	0.374	0.331	0.424	0.531	0.448
Sedoheptulose	0.329	0.426	0.320	0.362	0.332	0.405	0.530	0.459
Pyruvate	0.317	0.463	0.302	0.363	0.324	0.501	0.511	0.447
Oxaloacetate	0.331	0.482	0.311	0.381	0.359	0.530	0.500	0.455
Glycerol	0.308	0.466	0.301	0.341	0.302	0.423	0.477	0.411
Malate	0.331	0.504	0.317	0.395	0.329	0.542	0.477	0.417
Lactate	0.319	0.463	0.307	0.351	0.313	0.464	0.476	0.440
Fumarate	0.328	0.496	0.313	0.386	0.327	0.534	0.466	0.402
Succinate	0.319	0.463	0.307	0.376	0.320	0.517	0.462	0.430
Citrate	0.316	0.462	0.299	0.369	0.314	0.454	0.435	0.492
2-ketoglutarate	0.322	0.474	0.302	0.393	0.312	0.521	0.393	0.477
Ethanol	0.305	0.512	0.282	0.350	0.306	0.436	0.324	0.230
Acetate	0.320	0.466	0.284	0.344	0.473	0.471	0.297	0.323

One should also note that there is no one-to-one relationship between the robustness scores and the number of elementary modes. In some cases, the number of EFMs (flexibility) decreases whereas robustness score increases. For example, *E. coli* can

produce biomass in 204 different ways during growth on acetate, with MCS-robustness score being 0.320. On the other hand, it has noticeably less alternative pathways for biomass production when the substrate is glycerol (159). However, glycerol as substrate leads to more robust metabolism, with a MCS-robustness score of 0.410. In some other cases, flexibility and robustness score change in the same direction (Figure 6.4, Figure 6.5).

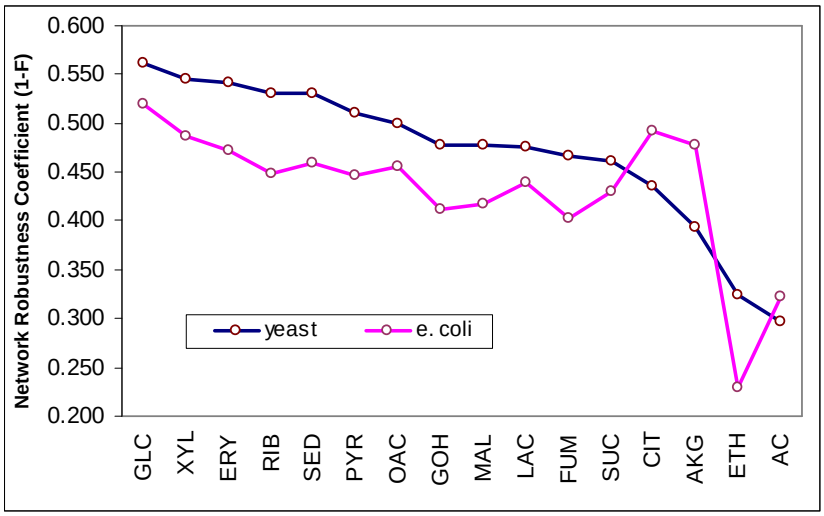


Figure 6.3. Robustness of *E. coli* and *S. cerevisiae* metabolisms on different substrates based on the fourth measure which depends on minimal cut set calculation

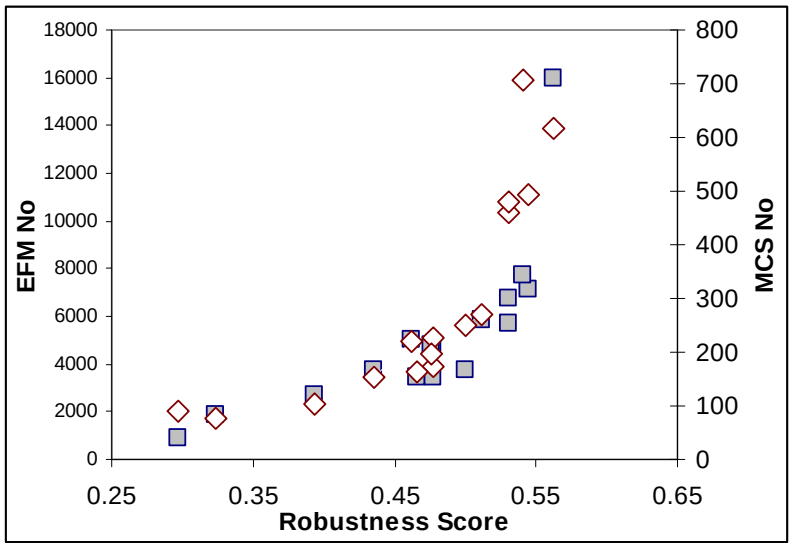


Figure 6.4. Comparison of robustness scores for growth on different substrates with the number of EFMs for *S. cerevisiae* metabolism.

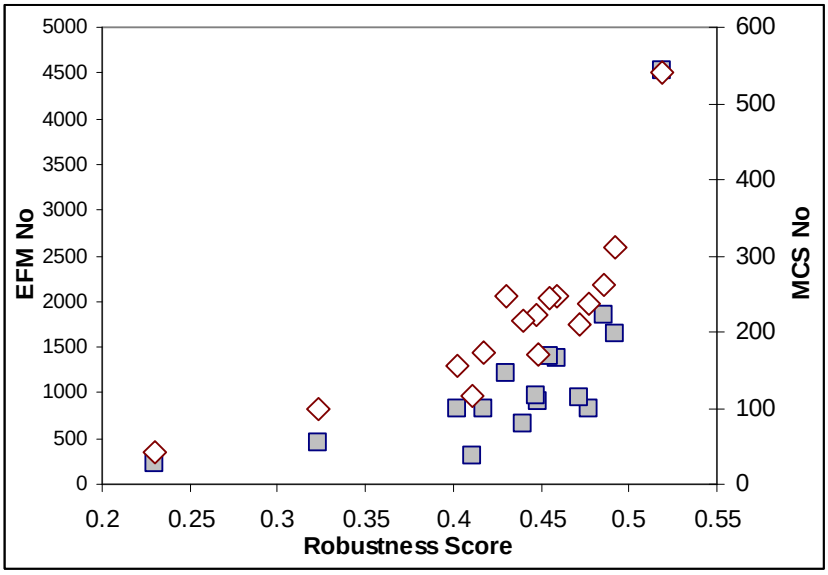


Figure 6.5. Comparison of robustness scores for growth on different substrates with the number of EFMs for *E. coli* metabolism.

6.3. Concluding Remarks

The *in silico* survival analysis of the *S. cerevisiae* cells in response to single or multiple gene deletions was made by using different robustness measures. Minimal-cut-set-based measure was shown to be the one mostly agreeing with the known behaviour of yeast. Among many different carbon substrates, the natural substrate, glucose, was shown to be the substrate on which the microorganism exhibits the most robust behaviour among others. Comparison with *E. coli* metabolism indicated a more robust metabolic structure of *S. cerevisiae*.

7. CONCLUSIONS AND RECOMMENDATIONS

7.1. Conclusions

A system-level understanding of the metabolism was pursued in this research by stoichiometric modeling approaches. The following conclusions can be drawn on the basis of the results presented.

- Control effective fluxes were shown to better in explaining the changes that occur in metabolic transcriptome in response to a perturbation.
- Applicability of CEF approach to *S. cerevisiae* and human metabolism was successfully shown, reinforcing the universality of the approach.
- Metabolic fluxes of central carbon metabolism were shown to be transcriptionally regulated for carbon source perturbations. Oxygen source perturbation was found to shift the flux regulation into post-transcriptional or metabolic level.
- Metabolic fluxes of amino acid metabolism were shown to be post-transcriptionally or metabolically regulated. Specially designed experiments are required to test this finding.
- Integration of high-throughput metabolome data with metabolic networks was achieved to identify reporter reactions, even with a small number of measured metabolites.
- Further integration of reporter reaction approach with transcriptome data allowed inferring information about the regulation type (metabolic and/or hierarchical) of the reactions.
- Using minimal-cut-set based robustness measure, glucose was shown to be the substrate on which *S. cerevisiae* exhibits most robust metabolism. Comparison with bacterium *E. coli* indicates a more robust metabolic structure of the yeast. The *in silico* analysis is helpful to guide research on phenomics.

Thereby, stoichiometric metabolic modeling approaches were shown to be successful in yeast systems biology research.

7.2. Recommendations

In chapter 4, based on transcriptome data, predictions were made about the regulation type of fluxes in response to carbon source perturbations. Metabolome analysis of carbon shift experiments and subsequent analysis of this data to calculate metabolically regulated 'reporter' reactions will allow a quick verification of the predictions made in chapter 4 about omitted points- not transcriptionally regulated fluxes. The idealized case would be to obtain also proteome data, thereby the level at which the fluxes are regulated can be identified in a high-throughput manner.

An experimental fluxome analysis for the studied perturbations can be performed to identify whether the metabolic/hierarchical regulation prediction made in chapter 5 about the reactions are in agreement at flux level. Alternatively, a model-based fluxome analysis based on EFMs can also be performed as introduced in chapter 3, thereby making a bridge between the two methods presented in chapters 3-4 and in chapter 5.

Furthermore, the model driven analysis presented in chapter 5 for the integration of metabolomic and transcriptomic data is flexible and may allow integration of other types of omics data, such as proteomics, and this will refine the method presented herein to account for the genome-scale alterations in response to genetic as well as environmental perturbations, and hence allow genome-scale identification of all levels of regulation in the metabolism.

Robustness measures utilized in chapter 6 identify lethal multiple deletions. The advent of phenomics technologies to identify high-throughput lethality deletions will allow verification and refinement of the results, and subsequently improve the employed stoichiometric model. Additionally, experiments on minimal media with each studied carbon source can be conducted to identify corresponding growth rates. Thereby, it can be checked whether the growth rates are in quantitative correlation with corresponding number of EFMs or MCSs.

APPENDIX A: FLUX BALANCE ANALYSIS FORMULATION

A.1. Formulation of Balance Equations

For a biochemical system with *n* reactions and *m* metabolites, stoichiometric matrix, S, has dimensions of m×n, providing that,

$$\mathbf{S} \times \mathbf{v} = \mathbf{b} \tag{A.1}$$

Here, **v** shows the reaction rate vector, and **b** represents the net accumulation rate of the metabolites in the medium.

After a stoichiometric matrix is formulated, a negative identity matrix is added next to this stoichiometric matrix with dimensions m×m. The new matrix is called **A** and is a $m \times (n+m)$ matrix.

$$\mathbf{A} = [\mathbf{S} \quad -\mathbf{I}] \tag{A.2}$$

The physical meaning for this addition is the incorporation of net accumulation rates of metabolites into calculations. Since there are *m* metabolites, identity matrix is in dimensions of $m \times m$.

With this incorporation, balance equations for this *m* compound- *n* reaction system are written as,

$$\mathbf{A} \times \mathbf{R} = \mathbf{0} \tag{A.3}$$

where **R** is a rate vector containing both the reaction rates and the net accumulation rates of the metabolites, i.e,. it is a $((n+m) \times 1)$ matrix.

Then, the metabolites in the system are grouped according to their accumulations in the medium. Those that have negligible pools in the medium are called intracellular. The

substrates taken from medium and the products secreted to the medium are known as extracellular metabolites. If number of extracellular substances in the system is denoted as *p*, then (*m*-*p*) metabolites are intracellular.

Since the accumulation rates of the intracellular metabolites are zero, the columns of identity matrix part of **A** corresponding to these metabolites are omitted. Hence, (*m*-*p*) columns are deleted from **A** matrix. This new matrix is called **A'**. The dimension of **A'** is $m \times (n+p)$. **R** vector is also changed to **R'**, which is a $((n+p) \times 1)$ vector.

$$\mathbf{A}' \times \mathbf{R}' = \mathbf{0} \tag{A.4}$$

There are $(n+p)$ unknowns in the system, i.e. n reaction rates and p accumulation rates for extracellular substances are unknown. Rank**A'** gives number of independent equations in the system. The difference is the degrees of freedom of the system.

$$dof = (n+p) - rank\,\mathbf{A}' \tag{A.5}$$

If *dof* is greater than zero, then the system is underdetermined. All real metabolic systems are underdetermined by their nature. To make the system determined, a number of rates equal to *dof* must be specified.

A.2. Solution of Undetermined Systems

If there is no external measurement of metabolites in the system, or if the number of measurements are less than *dof* of the system, the system remains underdetermined. Thus, the number of possible solutions is infinite. For the solution of such a system, linear programming is applied and this method is named as Flux Balance Analysis (FBA).

First, the matrix **A'** defined in Equation A.4 must be partitioned according to the measured accumulation rates as follows,

$$\mathbf{A}^{calc} \times \mathbf{R}^{calc} + \mathbf{A}^{meas} \times \mathbf{R}^{meas} = \mathbf{0} \tag{A.6}$$

Here, the corresponding columns of **A'** for measured metabolites are moved to a new matrix called $\mathbf{A}^{meas}$. The remaining part of **A'** is called $\mathbf{A}^{calc}$. The corresponding rate vector **R'** is also partitioned accordingly. If k of the p external metabolites are measured, $\mathbf{A}^{meas}$ and $\mathbf{A}^{calc}$ will have dimensions of $m \times k$ and $m \times (n+(p-k))$ respectively.

The second term of Equation A.6 is denoted with vector **b** since it is calculable.

$$-\mathbf{A}^{meas} \times \mathbf{R}^{meas} = \mathbf{b} \tag{A.7}$$

$\mathbf{R}^{meas}$ is a (k$\times$ 1) matrix. Since $\mathbf{A}^{meas}$ and $\mathbf{R}^{meas}$ are known, the resultant **b** vector can be calculated numerically. As a result, the following final formula is obtained,

$$\mathbf{A}^{calc} \times \mathbf{R}^{calc} = \mathbf{b} \tag{A.8}$$

Equation A.8 is the basis for the solution of underdetermined biochemical reaction systems. This equation describes a set of equations that are used as equality constraints in the linear programming solution.

Additional to Equation A.8, the lower and upper bounds of each of the reactions stored in $\mathbf{R}^{calc}$ constitute inequality constraints to the problem. These constraints can easily be derived from reaction reversibility information.

$$\mathbf{R}^{lb} \leq \mathbf{R}^{calc} \leq \mathbf{R}^{ub} \tag{A.9}$$

An objective function Z is needed for the solution. The objective function is selected as maximization/ minimization of the accumulation (production/consumption) of a particular metabolite. This is mathematically expressed by a row vector with dimensions equal to the dimensions of $\mathbf{R}^{calc}$. In this row vector, f, the entry corresponding to the net accumulation rate of the metabolite to be optimized is entered as 1 (for minimization) or -1 (for maximization) with other entries being zero;

$$Z = f \times \mathbf{R}^{calc} \tag{A.10}$$

Here, *f* shows the mentioned row vector. All elements of *f* is zero except the one for the net accumulation rate of the metabolite to be optimized. Hence, *Z*, the objective function, is a 1x1 vector. *Z* is not directly used in the solution, but *f* is utilized as an indication of objective function.

Table A.1 summarizes all these theoretical definitions. The stoichiometric network can be solved using optimization packages such as TOMLAB or MOSEK under MATLAB.

Table A.1. The summary of formulation of balance equation and the adaptation of these formulations to the considered reaction system

	General Formulation
Number of reactions	n
Number of metabolites	m
Number of extracellular metabolites	p
Number of intracellular metabolites	m-p
S	m×n
v	n×1
b	m×1
I	m×m
A	$m\times (n+m)$
R	$(n+m) \times 1$
A’	m× (n+p)
R’	$(n+p) \times 1$
Number of unknown rates	$(n+p)$
Number of independent equations	rank **A’**
dof	$(n+p) - rank\,\mathbf{A}'$
Number of measured accumulation rates	k
A^{calc}	$m\times (n+(p-k))$
A^{meas}	(m× k)
R^{meas}	(k× 1)
R^{calc}	$(n+(p-k)) \times 1$
f	$1\times (n+(p-k))$

APPENDIX B: REACTION SET FOR CENTRAL METABOLISM OF YEAST

B.1. Reactions

Reversible reactions in the glycolysis/gluconeogenesis pathway were written in the direction of their occurrence. Superscripts E and G in gene names show that the corresponding reaction is particular to growth in ethanol or glucose. External metabolites were written in bold font in their uptake and secretion reactions.

1. Substrate Uptake

(1)	1 **GLUC** + 1 ATP $\rightarrow$ 1 GLUC6P + ADP	*GLK1,HXK*^{G}
(2)	1 **GAL** + 1 ATP $\rightarrow$ 1 GLUC6P + ADP	*GAL1,5,7*
(3)	1 **ETOH** + 1 NAD_{cyt} $\rightarrow$ 1 ACAL + 1 $NADH_{cyt}$	*ADH2*^{E}

2. Glycolysis &Gluconeogenesis

(4)	1 GLUC6P $\leftrightarrow$ 1 FRUC6P	*PGI1*
(5)	1 FRUC6P + 1 ATP $\rightarrow$ 1 FRUCDP + ADP	*PFK1,2*^{G}
(6)	1 FRUCDP $\rightarrow$ 1 FRUC6P	*FBP1*^{E}
(7)	1 FRUCDP $\leftrightarrow$ 1 GA3P + 1 DHAP	*FBA1*
(8)	1 DHAP $\leftrightarrow$ 1 GA3P	*TPI1*
(9)	1 GA3P + 1 NAD_{cyt} $\leftrightarrow$ 1 P13G + 1 $NADH_{cyt}$	*TDH1,2,3*
(10)	1 P13G + 1 ADP $\leftrightarrow$ 1 P3G + 1 ATP	*PGK1*
(11)	1 P3G $\leftrightarrow$ 1 P2G	*GPM1,2,3*
(12)	1 P2G $\leftrightarrow$ 1 PEP	*ENO1,2*
(13)	1 PEP + 1 ADP $\rightarrow$ 1 PYR + 1 ATP	*PYK1,2*^{G}
(14)	1 DHAP + 1 $NADH_{cyt}$ $\rightarrow$ 1 GOH3P + 1 NAD_{cyt}	*GPD1,2*
(15)	1 GOH3P $\rightarrow$ 1 **GOH**	*GPP*
(16)	1 PYR $\rightarrow$ 1 ACAL + 1 CO_2	*PDC1,2,5*
(17)	1 ACAL + 1 $NADH_{cyt}$ $\rightarrow$ 1 **ETOH** + 1 NAD_{cyt}	*ADH1,4*^{G}
(18)	1 ACAL + 1 $NADP_{cyt}$ $\rightarrow$ 1 AC + 1 $NADPH_{cyt}$	*ALD6*

(19)	1 ACAL + 1 NAD_{mit} → 1 **AC** + 1 $NADH_{mit}$	*ALD4*
(20)	1 AC + 2 ATP → 1 $ACCOA_{cyt}$ + 2 ADP	*ACS1,2*
(21)	1 PYR + 1 NAD_{mit} → 1 $ACCOA_{mit}$ + 1 $NADH_{mit}$ + 1 CO_2	*PDA1,2, PDB*
(22)	1 PYR + 1 ATP + 1 CO_2 → 1 OAC + 1 ADP	*PYC1,2*
(23)	1 OAC+ 1 ATP → 1 PEP + 1 ADP + 1 CO_2	*PCK1*^{E}
	3. Pentose Phosphate Pathway	
(24)	1 GLUC6P + 1 $NADP_{cyt}$ → 1 G15L + 1 $NADPH_{cyt}$	*ZWF1*
(25)	1 G15L → 1 P6G	*SOL1,2,3,4*
(26)	1 P6G + 1 $NADP_{cyt}$ → 1 RIBL5P + 1 $NADPH_{cyt}$ + 1 CO_2	*GND1,2*
(27)	1 RIBL5P ↔ 1 RIB5P	*RKI1*
(28)	1 RIBL5P ↔ 1 XYL5P	*RPE1*
(29)	1 RIB5P + 1 XYL5P ↔ 1 SED7P + 1 GA3P	*TKL, TKI*
(30)	1 SED7P + 1 GA3P ↔ 1 FRUC6P + 1 E4P	*TAL1*
(31)	1 XYL5P + 1 E4P ↔ 1 FRUC6P + 1 GA3P	*TKI, TKL*
	4. Citric Acid Cycle	
(32)	1 OAC + 1 $ACCOA_{mit}$ → 1 CIT	*CIT1,3*
(33)	1 CIT ↔ 1 ISOCIT	*ACO1*
(34)	1 ISOCIT + 1 NAD_{mit} → 1 AKG + 1 $NADH_{mit}$ + 1 CO_2	*IDH1,2*
(35)	1 ISOCIT + 1 $NADP_{mit}$ → 1 AKG + 1 $NADPH_{mit}$ + 1 CO_2	*IDP1*
(36)	1 ISOCIT + 1 $NADP_{cyt}$ → 1 AKG + 1 $NADPH_{cyt}$ + 1 CO_2	*IDP2*
(37)	1 AKG + 1 NAD_{mit} → 1 SUCCOA + 1 $NADH_{mit}$ + 1 CO_2	*KGD1,2*
(38)	1 SUCCOA + 1 ADP ↔ 1 SUC + 1 ATP	*LSC1,2*
(39)	1 SUC + 1 FAD → 1 FUM + 1 $FADH_2$	*SDH1,2,3*
(40)	1 FUM + 1 $FADH_2$ → 1 SUC + 1 FAD	*OSM1*
(41)	1 FUM ↔ 1 MAL	*FUM1*
(42)	1 MAL + 1 NAD_{mit} ↔ 1 OAC + 1 $NADH_{mit}$	*MDH1*
(43)	1 MAL + 1 $NADP_{mit}$ → 1 PYR + 1 CO_2 + 1 $NADPH_{mit}$	*MAE1*
(44)	1 $ACCOA_{cyt}$ → 1 $ACCOA_{mit}$	*CAT2*
(45)	1 $NADH_{cyt}$ + 1 NAD_{mit} → 1 NAD_{cyt} + 1 $NADH_{mit}$	*ShuttleX*
	5. Glyoxylate Shunt	
(46)	1 OAC + 1 $ACCOA_{cyt}$ → 1 CIT	*CIT2*^{E}

(47) 1 ISOCIT $\rightarrow$ 1 GLYO + 1 SUC *ICL1* ^{E}

(48) 1 GLYO + 1 $ACCOA_{cyt}$ $\rightarrow$ 1 MAL *MLS1* ^{E}

(49) 1 MAL + 1 NAD_{cyt} $\leftrightarrow$ 1 OAC + 1 $NADH_{cyt}$ *MDH2* ^{E}

6. Oxidative Phosphorylation

(50) 24 ADP + 20 $NADH_{mit}$ + 10 O_2 $\rightarrow$ 24 ATP + 20 NAD_{mit} *NADHX*

(51) 24 ADP + 20 $FADH_2$ + 10 O_2 $\rightarrow$ 24 ATP + 20 FAD *FADHX*

(52) 1 ATP $\rightarrow$ 1 ADP *MAINT*

7. Biomass Formation

(53) 3 $ACCOA_{mit}$ + 24 $ACCOA_{cyt}$ + 11 AKG + 3 E4P + 6 P3G + 1 GOH3P + 6 PEP + 18 PYR + 3 RIB5P + 25 GLUC6P + 10 OAC + 16 NAD_{cyt} + 6 NAD_{mit} + 90 $NADPH_{cyt}$ + 22 $NADPH_{mit}$ + 254 ATP $\rightarrow$ 10000 **BIOM** + 16 $NADH_{cyt}$ + 6 $NADH_{mit}$ + 90 $NADP_{cyt}$ + 22 $NADP_{mit}$ + 254 ADP *BIOMX*

B.2. Reaction Abbreviations

ACO	Aconitate synthetase
ACS	Acetyl-coenzyme A synthetase
ADH	Alcohol dehydrogenase
ALD6	Aldehyde dehydrogenase ($NADP_{cyt}$ dependent)
ALD4	Aldehyde dehydrogenase (NAD_{mit} dependent)
BIOMX	Biomass formation
CAT	Carnitine O-acetyltransferase
CIT1,3	Citrate synthase (cytosolic)
CIT2	Citrate synthase (mitochondrial)
ENO	Enolase
FADHX	Electronic chain: reoxidation of FADH
FBA	Aldolase
FBP	Fructose 1,6-biphosphatase
FUM	Fumarate hydratase

GAL	Galactokinase
GLK,HXK	Glucokinase, Hexokinase
GND	Phosphogluconate dehydrogenase
GPD	Glycerol-3-phosphate dehydrogenase
GPM	Phosphoglycerate mutase
GPP	Glycerol phosphatase
ICL	Isocitrate lyase
IDH	Isocitrate dehydrogenase (NAD_{mit} dependent)
IDP1	Isocitrate dehydrogenase ($NADP_{mit}$ dependent)
IDP2	Isocitrate dehydrogenase ($NADP_{cyt}$ dependent)
KGD	Alpha-ketogluterate dehydrogenase
LSC	Succinate-CoA ligase
MAE	Malic enzyme
MDH1	Malate dehydrogenase (NAD_{mit} dependent)
MDH	Malate dehydrogenase (NAD_{cyt} dependent)
MLS	Malate synthase
NADHX	Electronic chain: reoxidation of NADH
OSM	Fumarate reductase
PCK	Phosphoenolpyruvate carbboxykinase
PDA, PDB	Pyruvate dehydrogenase
PDC	Pyruvate decarboxylase
PFK	Phosphofructokinase
PGI	Glucose-6-phosphate isomerase
PGK	3-phosphoglycerate kinase
PYC	Pyruvate carboxylase
PYK	Pyruvate kinase
RKI	Ribose 5-phosphate isomerase
RPE	Ribulose-phosphate 3-epimerase
SDH	Succinate dehydrogenase
ShuttleX	Reoxidation of cytosolic NADH to mitochondrial NADH
SOL	6-phosphoglucono-lactonase
TAL	Transaldolase
TDH	Glyceraldehyde-3-phosphate dehydrogenase

TKL	Transketolase
TPI	Triosephosphate isomerase
ZWF	Glucose-6-phosphate dehydrogenase

APPENDIX C: METABOLITE AND REACTION ABBREVIATIONS FOR RED BLOOD CELL METABOLIC NETWORK

C.1. Metabolite Abbreviations

ADE	adenine
ADO	adenosine
D13PG	1,3-diphosphoglycerate
D23PG	2,3-diphosphoglycerate
DHAP	dihydroxyacetone phosphate
ERY4P	erythrose 4-phosphate
F6P	fructose-6-phosphate
FDP	frucotose diphosphate
G6P	glucose-6-phosphate
GA3P	glyceraldehydes-3-phosphate
GL6P	6-phosphogluco lactone
GLC	glucose
GO6P	6-phosphogluconate
GSH	reduced glutathione
GSSG	oxidized glutathione
H_2O_2	hydrogen peroxide
Hb	hemoglobin
HYPX	hypoxanthine
IMP	inosine monophosphate
INO	inosine
LAC	lactate
MetHb	methemoglobin
P2G	2-phosphoglycerate
P3G	3-phospho glycerate
PEP	phosphenolpyruvate
PRPP	5-phosphoribosyl pyrophosphate

PYR	pyruvate
R1P	ribose-1-phosphate
R5P	ribose-5-phosphate
RL5P	ribulose-5-phosphate
SED7P	sedoheptulose 7-phosphate
XYL5P	xylose-5-phosphate

C.2. Reaction Abbreviations

ADA	adenosine deaminase (EC 3.5.4.4)
ADPRT	adenine phosphoribosyltransferase (EC 2.4.2.7)
AdylK	adenylate kinase (EC 2.7.4.3)
AK	adenosine kinase (EC 2.7.1.20)
ALD	aldolase (EC 4.1.2.13)
AMPase	adenosine monophosphate phosphohydrolase (EC 3.1.3.5)
AMPDA	adenosine monophosphate deaminase (EC 3.5.4.6)
D23PGdrain	2, 3 diphosphoglyerate drain
DPGase	diphosphoglycerate phosphatase (EC 3.1.3.13)
DPGM	diphosphoglycerate mutase (EC 5.4.2.4)
EN	enolase (EC 4.2.1.11)
G6PDH	glucose-6-phosphate dehydrogenase (EC 1.1.1.49)
GAPDH	glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12)
GL6PDH	phosphogluconate dehydrogenase (EC 1.1.1.44)
GSHox	glutathione oxidase (EC 1.8.3.3)
GSHpox	glutathione peroxidase (EC 1.11.1.9)
GSSG-R	glutathione-disulfide reductase (EC 1.8.1.7)
HGPRT	hypoxanthine phosphoribosyltransferase (EC 2.4.2.8)
HK	hexokinase (EC 2.7.1.1)
IMPase	inosine monophosphate phosphohdrolase (EC 3.1.3.5)
LDH	lactate dehydrogenase (EC 1.1.1.27)
MemPhos	membrane phosphorylation
MetHbRed	methemoglobin reductase

NaKATPase	sodium-potassium cation pump
PFK	phosphofructokinase (EC 2.7.1.11)
PGI	glucose-6-phosphate isomerase (EC 5.3.1.9)
PGK	phosphoglycerate kinase (EC 2.7.2.3)
PGLase	Phosphogluconolactonase (EC 3.1.1.31)
PGM	phosphoglycerate mutase (EC 5.4.2.1)
PK	pyruvate kinase (EC 2.7.1.40)
PNPase	purine-nucleoside phosphorylase (EC 2.4.2.1)
PRM	phosphoribomutase (EC 5.4.2.7)
PRPPsyn	phosphoribosylpyrophosphate synthetase (EC 2.7.6.1)
RPI	ribose-5-phosphate isomerase (EC 5.3.1.6)
TA	transaldolase (EC 2.2.1.2)
TK	transketolase (EC 2.2.1.1)
TPI	triosephosphate isomerase (EC 5.3.1.1)
XPI	ribulose phosphate epimerase (EC 5.1.3.1)

APPENDIX D: REACTION SET FOR CENTRAL AND AMINO ACID METABOLISM OF YEAST

The reaction network contains 55 central metabolic reactions including the secretion & uptake mechanisms (governed by 91 genes). The remaining 28 reactions belong to the amino acid metabolism (governed by 46 genes).

Reversible reactions in the glycolysis/gluconeogenesis pathway were written in the direction of their occurrence. Superscripts E and G in gene names show that the corresponding reaction is particular to growth in ethanol or glucose. External metabolites were written in bold font in their uptake and secretion reactions.

	1. Substrate Uptake	
(1)	1 **GLUC** + 1 ATP $\rightarrow$ 1 GLUC6P + ADP	*GLK1, HXK1,2*^{G}
(2)	1 **ETOH** + 1 NAD_{cyt} $\rightarrow$ 1 ACAL + 1 $NADH_{cyt}$	*ADH2*^{E}
(3)	1 **NH_3** $\rightarrow$	*MEP1,2,3*
	2. Glycolysis &Gluconeogenesis[a]	
(4)	1 GLUC6P $\leftrightarrow$ 1 FRUC6P	*PGI1*
(5)	1 FRUC6P + 1 ATP $\rightarrow$ 1 FRUCDP + ADP	*PFK1,2*^{G}
(6)	1 FRUCDP $\rightarrow$ 1 FRUC6P	*FBP1*^{E}
(7)	1 FRUCDP $\leftrightarrow$ 1 GA3P + 1 DHAP	*FBA1*
(8)	1 DHAP $\leftrightarrow$ 1 GA3P	*TPI1*
(9)	1 GA3P + 1 NAD_{cyt} $\leftrightarrow$ 1 P13G + 1 $NADH_{cyt}$	*TDH1,2,3*
(10)	1 P13G + 1 ADP $\leftrightarrow$ 1 P3G + 1 ATP	*PGK1*
(11)	1 P3G $\leftrightarrow$ 1 P2G	*GPM1,2,3*
(12)	1 P2G $\leftrightarrow$ 1 PEP	*ENO1,2, ERR1,2,3*
(13)	1 PEP + 1 ADP $\rightarrow$ 1 PYR + 1 ATP	*PYK1,2*^{G}
(14)	1 DHAP + 1 $NADH_{cyt}$ $\rightarrow$ 1 GOH3P + 1 NAD_{cyt}	*GPD1,2*
(15)	1 GOH3P $\rightarrow$ 1 GOH	*GPP1, HOR2*

(16)	1 PYR $\rightarrow$ 1 ACAL + 1 CO_2	*PDC1,5,6*
(17)	1 ACAL + 1 $NADH_{cyt}$ $\leftrightarrow$ 1 ETOH + 1 NAD_{cyt}	*ADH1,4,5*^{G}
(18)	1 ACAL + 1 $NADP_{cyt}$ $\rightarrow$ 1 AC + 1 $NADPH_{cyt}$	*ALD6*
(19)	1 ACAL + 1 NAD_{mit} $\rightarrow$ 1 AC + 1 $NADH_{mit}$	*ALD4,5*
(20)	1 AC + 2 ATP $\rightarrow$ 1 $ACCOA_{cyt}$ + 2 ADP	*ACS1,2*
(21)	1 PYR + 1 NAD_{mit} $\rightarrow$ 1 $ACCOA_{mit}$ + 1 $NADH_{mit}$ + 1 CO_2	*PDA1,2, PDB1, PDX1, LPD1*
(22)	1 PYR + 1 ATP + 1 CO_2 $\rightarrow$ 1 OAC + 1 ADP	*PYC1,2*
(23)	1 OAC+ 1 ATP $\rightarrow$ 1 PEP + 1 ADP + 1 CO_2	*PCK1*^{E}
	3. Pentose Phosphate Pathway	
(24)	1 GLUC6P + 1 $NADP_{cyt}$ $\rightarrow$ 1 G15L + 1 $NADPH_{cyt}$	*ZWF1*
(25)	1 G15L $\rightarrow$ 1 P6G	*SOL3,4*
(26)	1 P6G + 1 $NADP_{cyt}$ $\rightarrow$ 1 RIBL5P + 1 $NADPH_{cyt}$ + 1 CO_2	*GND1,2*
(27)	1 RIBL5P $\leftrightarrow$ 1 RIB5P	*RKI1*
(28)	1 RIBL5P $\leftrightarrow$ 1 XYL5P	*RPE1*
(29)	1 RIB5P + 1 XYL5P $\leftrightarrow$ 1 SED7P + 1 GA3P	*TKL1,2*
(30)	1 SED7P + 1 GA3P $\leftrightarrow$ 1 FRUC6P + 1 E4P	*TAL1, YGR043C*
(31)	1 XYL5P + 1 E4P $\leftrightarrow$ 1 FRUC6P + 1 GA3P	*TKL1,2*
	4. Citric Acid Cycle	
(32)	1 OAC + 1 $ACCOA_{mit}$ $\rightarrow$ 1 CIT	*CIT1,3*
(33)	1 CIT $\leftrightarrow$ 1 ISOCIT	*ACO1,2*
(34)	1 ISOCIT + 1 NAD_{mit} $\rightarrow$ 1 AKG + 1 $NADH_{mit}$ + 1 CO_2	*IDH1,2*
(35)	1 ISOCIT + 1 $NADP_{mit}$ $\rightarrow$ 1 AKG + 1 $NADPH_{mit}$ + 1 CO_2	*IDP1*
(36)	1 ISOCIT + 1 $NADP_{cyt}$ $\rightarrow$ 1 AKG + 1 $NADPH_{cyt}$ + 1 CO_2	*IDP2,3*
(37)	1 AKG + 1 NAD_{mit} $\rightarrow$ 1 SUCCOA + 1 $NADH_{mit}$ + 1 CO_2	*KGD1,2, LPD1*
(38)	1 SUCCOA + 1 ADP $\leftrightarrow$ 1 SUC + 1 ATP	*LSC1,2*
(39)	1 SUC + 1 FAD $\rightarrow$ 1 FUM + 1 $FADH_2$	*SDH1,2,3,4*
(40)	1 FUM + 1 $FADH_2$ $\rightarrow$ 1 SUC + 1 FAD	*OSM1*
(41)	1 FUM $\leftrightarrow$ 1 MAL	*FUM1*
(42)	1 MAL + 1 NAD_{mit} $\leftrightarrow$ 1 OAC + 1 $NADH_{mit}$	*MDH1*
(43)	1 MAL + 1 $NADP_{mit}$ $\rightarrow$ 1 PYR + 1 CO_2 + 1 $NADPH_{mit}$	*MAE1*

(44)	1 $ACCOA_{cyt}$ → 1 $ACCOA_{mit}$	*CAT2, YAT1*
(45)	1 ACAL + 1 $NADH_{mit}$ ↔ 1 ETOH + 1 NAD_{mit}	*ADH3*
	5. Glyoxylate Shunt	
(46)	1 OAC + 1 $ACCOA_{cyt}$ → 1 CIT	$CIT2^{E}$
(47)	1 ISOCIT → 1 GLYO + 1 SUC	$ICL1,2^{E}$
(48)	1 GLYO + 1 $ACCOA_{cyt}$ → 1 MAL	$MLS1,2^{E}$
(49)	1 MAL + 1 NAD_{mit} ↔ 1 OAC + 1 $NADH_{cyt}$	$MDH2,3^{E}$
	6. Oxidative Phosphorylation	
(50)	24 ADP + 20 $NADH_{mit}$ + 10 O_2 → 24 ATP + 20 NAD_{mit}	*NADHX*
(51)	24 ADP + 20 $FADH_2$ + 10 O_2 → 24 ATP + 20 FAD	*FADHX*
(52)	1 ATP → 1 ADP	*MAINT*
	7. Product Secretion	
(53)	→ 1 **GOH**	*FPS1*
(54)	→ 1 **ETOH**	
(55)	→ 1 **AC**	*BPH1*
(56)	→ 1 **SUC**	
	8. Glutamate & Glutamine Metabolism	
(57)	1 AKG + 1 $NADPH_{cyt}$ + 1 NH_3 → 1 GLT + 1 $NADP_{cyt}$	*GDH1,3*
(58)	1 GLT + 1 NAD_{cyt} → 1 AKG + 1 $NADH_{cyt}$ + 1 NH_3	*GDH2*
(59)	1 GLT + 1 ATP + 1 NH_3 → 1 GLN + 1 ADP	*GLN1*
(60)	1 AKG + 1 GLN + 1 $NADH_{cyt}$ → 2 GLT + 1 NAD_{cyt}	*GLT1*
(61)	1 GLT → 1 CO_2 + 1 GABA	*GAD1*
(62)	1 AKG + 1 GABA → 1 GLT + 1 SUCSAL	*UGA1*
(63)	1 SUCSAL + 1 $NADP_{cyt}$ → 1 SUC + 1 $NADPH_{cyt}$	*UGA2*
	9. Aspartate & Asparagine & Alanine Metabolism	
(64)	1 OAC + 1 GLT ↔ 1 AKG + 1 ASP	*AAT1,2*
(65)	2 ATP + 1 GLN + 1 ASP ↔ 2 ADP + 1 GLT + 1 ASN	*ASN1,2*
(66)	1 ASN → 1 ASP + 1 NH_3	*ASP3-4,1*
(67)	1 PYR + 1 GLT ↔ 1 AKG + 1 ALA	*ALT1,2*
	10. Leucine & Valine Metabolism	
(68)	2 PYR → 1 CO2 + 1 ACLAC	*ILV2,6*

(69)	1 ACLAC + 1 $NADPH_{mit}$ → 1 $NADP_{mit}$ + 1 DHVAL	*ILV5*
(70)	1 DHVAL → 1 OIVAL	*ILV3*
(71)	$ACCOA_{mit}$ + 1 OIVAL → 1 IPPMAL	*LEU4*
(72)	1 IPPMAL + 1 NAD_{cyt} → 1 $NADH_{cyt}$ + 1 OICAP + 1 CO_2	*LEU2*
(73)	1 OICAP + 1 GLT ↔ 1 AKG + 1 LEU	*BAT1,2*
(74)	1 OIVAL + 1 GLT ↔ 1 AKG + 1 VAL	*BAT2*
	11. Serine & Glycine Metabolism	
(75)	1 P3G + NAD_{cyt} → 1 PHP + 1 $NADH_{cyt}$	*SER3,33*
(76)	1 PHP + 1 GLT → 1 AKG + 1 P3SER	*SER1*
(77)	1 P3SER → 1 SER	*SER2*
(78)	1 SER → 1 PYR + 1 NH_3	*CHA1, SDL1*
(79)	1 SER ↔ 1 GLY + 1 C1	*SHM1,2*
(80)	1 ALA + 1 GLYO ↔ 1 PYR + 1 GLY	*AGX1*
(81)	1 GLY + NAD_{mit} → 1 C1 + $NADH_{mit}$ + CO_2 + NH_3	*GCV1*
	12. AICAR synthesis	
(82a)	1 R5P + 1 P3G + 8 ATP + 1 NAD_{mit} + 1 NAD_{cyt} + 2 $NADPH_{cyt}$ → 1 $NADH_{mit}$ + 1 $NADH_{cyt}$ + 2 $NADP_{cyt}$ + 8 ADP + 1 AICAR	*ADE4,5,7,8,6,5, 2,1,13- CCM*
(82b)	1 R5P + 6 ATP + 2 GLN + 1 GLY + 1 C1 + 1 CO_2 + ASP → 2 GLU + 1 FUM + 6 ADP + 1 AICAR	*ADE4,5,7,8,6,5, 2,1,13 -CCMAA*
	13. Biomass Formation	
(83a)	0.05877 PYR + 0.03293 OAC + 0.03430 AKG + 0.01767 P3G + 0.04302 $ACCOA_{cyt}$ + 0.00996 $ACCOA_{mit}$ + 0.00802 RB5P + 0.00889 E4P + 0.01777 PEP + 0.04400 G6P + 0.00187 GOH3P + 0.00201 AICAR + 0.28658 $NADPH_{cyt}$ + 0.03080 $NADPH_{mit}$ + 0.05056 NAD_{cyt} + 0.01716 NAD_{mit} + 1.80302 ATP → 0.05056 $NADH_{cyt}$ + 0.01716 $NADH_{mit}$ + 0.28658 $NADP_{cyt}$ + 0.03080 $NADP_{mit}$ + 1.80302 ADP + 0.05806 CO2 + 1 BIOM	*BIOMX-CCM (1 C-mol)*
83b)	0.00555 PYR + 0.04302 $ACCOA_{cyt}$ + 0 $ACCOA_{mit}$ + 0.00802 RB5P + 0.00889 E4P + 0.01777 PEP + 0.04400 G6P + 0.00187	*BIOMX-CCM-AA*

GOH3P + 0.04485GLT + 0.02116 GLN + 0.01545 ALA + 0.03703 ASP + 0.00343 ASN + 0.00885 SER + 0.00976 GLY + 0.00996 ILE + 0.00892 VAL + 0.00595 C1 + 0.00201 AICAR + 0.12717 $NADPH_{cyt}$ + 0.01192 $NADPH_{mit}$ + 0.02198 NAD_{cyt} + 0.01716 NAD_{mit} + 1.77156 ATP *(1 C-mol)*

$\rightarrow$

0.02198 $NADH_{cyt}$ + 0.01716 $NADH_{mit}$ + 0.12717 $NADP_{cyt}$ + 0.01192 $NADP_{mit}$ + 1.77156 ADP + 0.02922 CO_2 + 1 **BIOM** + 0.00754 OAC + 0.03171 AKG + 0.00094 P3G + 0.00650 NH_3

APPENDIX E : ALGORITHMS OF SAMPLING APPROACHES

E.1. Algorithm of Hit-and-Run Sampling

Hit-and-Run generates a sequence of points, V_k, which are inside a bounded and n-dimensional set, $S \in R^n$. The algorithm can be summarized in four steps (Kaufman and Smith, 1998; Zabinsky, 2003);

Step 0. Choose an arbitrary starting point, V_0, which is known to be inside the bounded set ($X_0 \in S$). Set k = 0.
Step 1. Generate a uniformly distributed random direction, D_k, over an n-dimensional hypersphere (termed hyperspherical direction, HD).
Step 2. Generate a random point $V_{k+1} = V_k + \lambda D_k$ uniformly distributed over the line set, L_k, which lies on the generated random direction;

$L_k = \{ v : v \in S$ and $v = V_k + \lambda D_k, \lambda$ a real scalar $\}$

If $L_k = \emptyset$, go to step 1.
Step 3. Stop if a stopping criterion is met (eg. the maximum number of points to be collected). Otherwise, increment k by one, and return to Step one.

Below is a more detailed explanation of the procedure;

Step 0. The initial starting point, V_0, must be inside the solution space defined by S. One should note that, although called as point, this is a vector having the same dimension as the solutions space. Here, for the sampling of flux space, the solution space is bounded by the lower and upper bounds of the fluxes as well the balancing of metabolites represented as the stoichiometric matrix. This initial point can be generated using FBA (Appendix A). To this aim, FBA is applied to the equation system by slightly increasing the lower bounds of fluxes and slightly decreasing upper bounds of fluxes. Thereby, the resulting solution vector of the system will not lie on the boundaries of the solution space, but will be guaranteed to be inside it.

Step 1. The direction is determined by randomly generating a vector of size equal to the dimension of V_k vector, from a uniform distribution (i.e. $D_k \in R^n$) from [-1, 1] interval. Then, this vector is scaled by its Euclidean norm in order to determine the unit direction vector, D_k;

$$D_k = \frac{(d_1, d_2, ..., d_n)}{\sqrt{\left(\sum_{i=1}^{n} d_i^2\right)}} \tag{E.1}$$

The resulting vector, D_k, is said to be uniformly distributed in hyperspherical direction.

Step 2. Generating a random point, V_{k+1}, on the line set, L_k, through V_k in the direction of D_k requires the determination of points where the line intersects S in both positive and negative directions. That is, the intersection points, λ_{min} and λ_{max}, where the direction vector hits the boundaries, are found such that $(V_k + \lambda D_k) \in S$ where λ is randomly chosen between the interval of λ_{min} and λ_{max}.

Determination of the bounds for λ is relatively straightforward for the case where S is a convex polytope defined by the linear constraints, $S = \{v : Av = 0, lb \leq x \leq ub\}$, with A being the stoichiometric matrix including exchange reactions, *lb* and *ub* being the lower and upper bound vectors for x. The procedure depends on the null space identification of the flux space S, and detailed in the following subsection.

E.1.1. Null Space Identification for Hit-and-Run Method

The direct implementation of Hit-and-Run method for linear equality constraints is not possible since the limits of λ will always be equal to zero for such a system. Therefore, an alternative approach is followed bu using the null space of the stoichiometric matrix, A. Null space of A is a space of all possible solutions of the system defined by the homogenous linear equation system, Av = 0. Null space determination can be determined in MATLAB using 'null' built-in function. If A is m x n in dimension, with m metabolites (equations) and n reactions (unknowns), the dimension of corresponding null space is equal to the degrees of freedom of the equation system defined by A. The column vectors

constituting null space are the basis vectors which define the flux solution space. That is, if the null space is defined as B, the linear combination of its columns, b_j, with random coefficients will define a point (vector) which will lie inside this space. Mathematically expressing,

$$B \times \alpha = V \tag{E.2}$$

In other words, the null space matrix, B, can be transformed into the flux space by the help of the coefficient vector, α. Therefore, by randomly generating coefficient vectors, a number of flux vectors within the null space can be obtained. However, it must be guaranteed that the resulting flux vectors must be within the lower/upper bound constraints of fluxes. That is,

$$lb \leq V \leq ub \tag{E.3}$$

Equations (E.2) and (E.3) form the basis of the null-space formulation for the sampling of flux space through Hit-and-Run method. The randomly generated flux vector through equation (E.2) must satisfy equation (E.3). In this way, the flux points can also be collected by Monte-Carlo approach, if the dimension of the system is small, by accepting-rejecting the generated points.

The following three steps summarize the procedure, and leads to the formation of an equation set which will be used in Hit-and-Run Sampling;

(1) Find null space of system matrix and form flux vectors such as;

$$B\alpha_k = V_k \tag{E.4}$$

where B is the null space matrix of A, α_k is a randomly generated coefficient vector and V_k is the resulting flux vector.

(2) Combine eqn. (E.4) with eqn. (E.3);

$$lb < B\alpha_k < ub \tag{E.5}$$

which can also be expressed as ; $B\alpha_k < ub$ & $-B\alpha_k < -lb$

(3) Represent the splitted form of eqn. (E.5) in matrix notation;

$$\begin{bmatrix} B \\ -B \end{bmatrix} \cdot \alpha_k \leq \begin{bmatrix} ub \\ -lb \end{bmatrix} \tag{E.6}$$

e.g. for 1st part: $\sum b_{ij}\alpha_j \leq ub_i$, and for the second part: $\sum -b_{ij}\alpha_j \leq -lb_i$ for the kth α vector. Therefore, the solution of equation system defined by eqn. (E.6) will give coefficient vector, α_k, which will generate flux vectors through eqn. (E.4) which will always satisfy the lower/upper bound constraints defined by eqn. (E.3). Therefore, the initial problem of generating flux points is transformed into the problem of generating coefficient points (vectors) through equation (E.6). Accordingly, the initially defined four steps for Hit-and-Run sampling must be modified.

E.1.2. Adaptation of the Original Hit-and-Run Algorithm for Null Space Representation

Step 0: Transform the initial starting flux vector, V_0, from flux space to the null space by multiplying both sides of eqn. (E.4) by the pseudo-inverse of null-space matrix, B.

$$\alpha_0 = pinv(B) \times V_0 \tag{E.7}$$

pseudo-inverse of a matrix, B, is equal to $(B^T \times B)\text{-}1 \times B^T$, and can be easily calculated in MATLAB using built-in function 'pinv'. Pseudo-inverse calculation is used if the matrix is not square and cannot be directly inversed. Thereby, the coefficient vector, α_0, will generate flux vectors which will not lie on the boundaries.

Step 1: The direction vector, D_k, subject to eqn. (E.1) is generated in the same way as in the original algorithm, but the dimension being equal to the dimension of the coefficient vector, not flux vector.

Step 2: The new coefficient vectors will be generated by Hit-and-Run method so that they will always result in flux vectors within the null space. Therefore, similar to the original formulation, the generation can be performed using the following equation;

$$\alpha_{k+1} = \alpha_k + \lambda D_k . \tag{E.8}$$

Here, the important point, as discussed before, is to determine λ so that resulting coefficient vector, α_{k+1}, will satisfy the linear inequality eqn. (E.6). The inequalities can be rewritten in the following form for i^{th} row, with *i* being the row index, j being the column index, and α_k and D_k being row vectors;

$$\sum_j b_{ij}\left(\alpha_j + \lambda d_j\right) \le ub_i \tag{E.9.a}$$

$$\sum_j -b_{ij}\left(\alpha_j + \lambda d_j\right) \le -lb_i \tag{E.9.b}$$

The aim is to determine lower and upper limits of the line passing through α_k in the direction of D_k. In other words, λ_{min}/ λ_{max} corresponding to the intersection points of the line with the coefficient space is to be identified. Therefore, λ will be left alone in left hand side of these equations;

$$\sum b_{ij}\alpha_j + \sum b_{ij}\lambda_j d_j \le ub_i \quad \& \quad \sum_j b_{ij}\alpha_j = v_i \quad \rightarrow \quad \lambda = \frac{ub_i - v_i}{\sum b_{ij} \cdot d_j} \tag{E.10.a}$$

$$\sum -b_{ik}\alpha_k - \sum b_{ik}\lambda D_k \le -lb_i \quad \& \quad \sum_j b_{ij}\alpha_j = v_i \quad \rightarrow \quad \lambda = \frac{lb_j - v_i}{\sum b_{ij} \cdot d_j} \tag{E.10.b}$$

Here, v_i, lb_i and ub_i show the ith element of the flux vector and its corresponding lower/upper bound values respectively. These equations define two λ values for the ith row, which also corresponds to the ith dimension of the flux space. However, for the specific flux space sampling problem dealt here, the upper bound of all flux variables is positive infinity. That is, the generated fluxes will satisfy upper bound constraints in any case. Therefore, Eqn. (E.10.a) will not be pursued anymore.

Focusing on Eqn. (E.10.b), it must be applied for each row of eqn. (E.6). The numerator of the equation will always be negative since $v_i \geq lb_i$. Therefore, depending on the sign of the denominator, λ will be either positive or negative. Focusing on these two possibilities separately, for the rows where the denominator is positive, Eqn. (E.10.b) will be negative. The maximum of these negative values (the one which is closest to zero) will give λ_{min}. This is the minimum distance to an intersection of the line with a constraint in the one direction. On the other hand,λ_{max} gives the minimum distance to an intersection on the other direction of the chosen line. For the rows where the denominator of eqn. (E.10.b) is negative, the value of λ will be positive. The minimum of these positive values will give λ_{max}.

Once these upper and lower bounds are computed, the real scalar λ for eqn. (E.8) is chosen randomly between λ_{min} and λ_{max}. Thereby, the new α coefficient vector is generated. Then, Eqn. (E.4) can be used to generate corresponding flux vector.

This procedure is repeated until a sufficient number of samples is collected. To prevent the cross-dependence of the generated points, the flux vectors are only stored in every 1000th direction change.

E.1.3. Adaptation of Monte-Carlo Sampling

One way to handle generation of flux points in the solution space is Monte-Carlo sampling. For this purpose, a random coefficient vector, α_j, can be generated, whose dimension is equal to the column number of the null space matrix, B. Then, corresponding flux vector can be obtained using eqn. (C.4). All elements of this vector must be checked

to see if all of them are within the boundary limits defined by eqn. (E.3). If all of the elements of the flux vector is within the limits, the generated flux vector is accepted and stored. The procedure can be repeated until the desired number of flux vectors are generated.

As the dimension of the system increases, the probability of the flux vector to satisfy all boundary conditions decreases. Therefore, Monte-Carlo approach for the null space is only applicable for flux systems with small dimensions.

E.2. Algorithm of multi-LP Sampling

Here, A high number of optimizations with different, random objective functions are performed in order to generate as many flux distributions as possible. That is, the row vector defining the objective function to be optimized is randomly generated in each iteration. The dimension of the row vector is equal to the column number of stoichiometric matrix (Appendix A).The negative binary values in this vector shows the fluxes to be maximized while the positive binary values indicates the fluxes to be maximized. For all other fluxes, the corresponding entry is zero. After the number of desired flux distributions is set, linear programming (FBA) is performed until this number is reached, with a different randomly generated objective function in each FBA.

APPENDIX F: REACTIONS AND CORRESPONDING Z-SCORES OF METABOLIC NETWORK, ENZSUB-3

F.1. Reactions of ENZSUB-3 Model

First column gives the genes responsible for the corresponding reaction, and the second column includes short names which are used in the article to refer to reactions. The superscript ES means that the corresponding reaction is an enzyme subset consisting of combination of more than one reaction. The first of the numbers in parenthesis next to pathway names gives the number of the reactions in that pathway; the second number gives the number of reactions prior to enzyme-subset reduction. Reaction names are shown in capital non-italic letters, and gene names are, as conventional, in capital italic letters.

Gene Name	Reaction Name	
		# CARBOHYDRATE METABOLISM
		# GLYCOLYSIS/GLUCONEOGENESIS (4/4)
GPM1-2-3	GPM	: 3-Phospho-D-glycerate <-> 2-Phospho-D-glycerate
ENO1-2, ERR1-2-3	ENO	: 2-Phospho-D-glycerate <-> Phosphoenolpyruvate
CDC19, PYK2	PYK	: ADP + Phosphoenolpyruvate -> ATP + Pyruvate
PDA1, PDB1, LAT1	PDA	: CoA + NAD+ + Pyruvate -> Acetyl-CoA + CO2 + NADH
		# CITRATE CYCLE (TCA CYCLE) (7/8)
CIT1-2-3	CIT	: Acetyl-CoA + Oxaloacetate -> Citrate + CoA
ACO1, YJL200C	ACO	: Citrate <-> Isocitrate
IDH1-2	IDH	: Isocitrate + NAD+ -> 2-Oxoglutarate + CO2 + NADH
IDP1-2-3ES	IDP ^{ES}	: Isocitrate + NADP+ -> 2-Oxoglutarate + CO2 + NADPH
KGD1-2	KGD	: 2-Oxoglutarate + CoA + NAD+ -> CO2 + NADH + Succinyl-CoA
LSC1	LSC1	: ATP + Itaconate + CoA <-> ADP + Orthophosphate + Itaconyl-CoA
LSC2	LSC2	: ATP + CoA + Succinate <-> ADP + Orthophosphate + Succinyl-CoA
		# ELECTRON TRANSPORT SYSTEM,

		COMPLEX II (3/3)
SDH1-2-3-4, YLR164W, YMR118C, YJL045W, YEL047C, OSM1	SDH	: FAD + Succinate <-> FADH2 + Fumarate
FUM1	FUM	: Fumarate <-> Malate
MDH1-2-3	MDH	: Malate + NAD+ <-> NADH + Oxaloacetate
		# ANAPLEROTIC REACTIONS (5/5)
ICL1-2	ICL	: Isocitrate -> Glyoxylate + Succinate
DAL7, MLS1	MLS	: Acetyl-CoA + Glyoxylate -> CoA + Malate
PCK1	PCK1	: ATP + Oxaloacetate -> ADP + CO2 + Phosphoenolpyruvate
PYC1-2	PYC	: ATP + CO2 + Pyruvate -> ADP + Orthophosphate + Oxaloacetate
MAE1	MAE1	: Malate + NADP+ -> CO2 + NADPH + Pyruvate
		# PYRUVATE METABOLISM (1/1)
PDC1-5-6	PDC	: Pyruvate -> Acetaldehyde + CO2
		# ENERGY METABOLISM
		# ATP SYNTHASE (1/1)
DLD1, CYB2	DLD	: 2 Ferricytochrome c + (R)-Lactate -> Pyruvate + 2 Ferrocytochrome c
		# FATTY ACIDS METABOLISM
		# FATTY ACID DEGRADATION (1/1)
POX1/FOX2/POT1	POX	: Myristic acid + ATP + 7 CoA + 7 FAD + 7 NAD+ -> AMP + Pyrophosphate + 7 FADH2 + 7 NADH + 7 Acetyl-CoA
		# PHOSPHOLIPID BIOSYNTHESIS (2/4)
CHO1	CHO1	: CDPdiacylglycerol + L-Serine <-> CMP + Phosphatidylserine
CHO2, OPI3^{ES}	OPI ^{ES}	: Phosphatidylethanolamine + 3 S-Adenosyl-L-methionine -> Phosphatidylcholine + 3 S-Adenosyl-L-homocysteine

		# STEROL BIOSYNTHESIS (2/6)
ERG6-2-3-5-4ES	ERG ^{ES}	: 3 NADPH + 2 Oxygen + S-Adenosyl-L-methionine + Zymosterol -> Ergosterol + 3 NADP+ + S-Adenosyl-L-homocysteine
*U14_*	U14_	: S-Adenosyl-L-methionine + Zymosterol -> Ergosterol + S-Adenosyl-L-homocysteine
		# NUCLEOTIDE METABOLISM
		# PURINE METABOLISM (3/11)
ADE1-2-4-5,7-6-8-13ES	AIC^{ES}	: 10-Formyltetrahydrofolate + 5-Phospho-alpha-D-ribose 1-diphosphate + 4 ATP + CO2 + Glycine + L-Aspartate + 2 L-Glutamine -> 1-(5'-Phosphoribosyl)-5-amino-4-imidazolecarboxamide + 4 ADP + Fumarate + 2 L-Glutamate + 4 Orthophosphate + Pyrophosphate + Tetrahydrofolate
ADE12-13 ^{ES}	AMP^{ES}	: GTP + IMP + L-Aspartate -> AMP + Fumarate + GDP + Orthophosphate
GUA1	GUA1	: ATP + L-Glutamine + Xanthosine 5'-phosphate -> AMP + GMP + L-Glutamate + Pyrophosphate
		# PYRIMIDINE METABOLISM (2/5)
URA2-3-4-5-10 ^{ES}	URA ^{ES}	: 5-Phospho-alpha-D-ribose 1-diphosphate + Carbamoyl phosphate + L-Aspartate + Orotate -> (S)-Dihydroorotate + CO2 + Orthophosphate + Pyrophosphate + UMP
URA7-8	URA78	: ATP + L-Glutamine + UTP -> ADP + CTP + L-Glutamate + Orthophosphate
		# AMINO ACID METABOLISM
		# GLUTAMATE METABOLISM (AMINOSUGARS METABOLISM) (10/11)
GAD1	GAD1	: L-Glutamate -> 4-Aminobutanoate + CO2
UGA1-2 ^{ES}	UGA ^{ES}	: 2-Oxoglutarate + 4-Aminobutanoate + NADP+ -> L-Glutamate + NADPH + Succinate
GFA1	GFA1	: beta-D-Fructose 6-phosphate + L-Glutamine -> D-Glucosamine 6-phosphate + L-Glutamate
PUT2	PUT2	: L-Glutamate 5-semialdehyde + NADP+ -> L-Glutamate + NADPH
*U41_*	U41_	: (S)-1-Pyrroline-5-carboxylate + NAD+ -> L-Glutamate + NADH

GLT1	GLT1	: 2-Oxoglutarate + L-Glutamine + NADH -> 2 L-Glutamate + NAD+
GDH2	GDH2	: L-Glutamate + NAD+ -> 2-Oxoglutarate + NADH + NH3
GDH1-3	GDH13	: 2-Oxoglutarate + NADPH + NH3 -> L-Glutamate + NADP+
GLN1	GLN1	: ATP + L-Glutamate + NH3 -> ADP + L-Glutamine + Orthophosphate
*U42_ -43_*	U42_43ES	: L-Glutamine -> L-Glutamate + NH3
		# ALANINE AND ASPARTATE METABOLISM (4/4)
AAT1-2	AAT	: L-Glutamate + Oxaloacetate <-> 2-Oxoglutarate + L-Aspartate
ALT1-2	ALT	: L-Glutamate + Pyruvate <-> 2-Oxoglutarate + L-Alanine
ASN1-2	ASN	: ATP + L-Aspartate + L-Glutamine -> AMP + L-Asparagine + L-Glutamate + Pyrophosphate
MHT1, SAM4	MHT	: Homocysteine + S-Adenosyl-L-methionine -> L-Methionine + S-Adenosyl-L-homocysteine
		# ASPARAGINE (1/1)
ASP(3-1)-(3-2)-(3-3)-(3-4)-1	ASP3-1	: L-Asparagine -> L-Aspartate + NH3
		# GLYCINE, SERINE AND THREONINE METABOLISM (11/15)
SER3-33-1-2ES	SERsyn ^{ES}	: 3-Phospho-D-glycerate + L-Glutamate + NAD+ -> 2-Oxoglutarate + L-Serine + NADH + Orthophosphate
SHM1-2	GLYsyn	: L-Serine + Tetrahydrofolate <-> 5,10-Methylenetetrahydrofolate + Glycine
AGX1	AGX1	: Glyoxylate + L-Alanine <-> Glycine + Pyruvate
GCV1	GCV1	: Glycine + NAD+ + Tetrahydrofolate -> 5,10-Methylenetetrahydrofolate + CO2 + NADH + NH3
HOM3-2 ^{ES}	HOM ^{ES}	: ATP + L-Aspartate + NADPH -> ADP + L-Aspartate 4-semialdehyde + NADP+ + Orthophosphate
THR1-4 ^{ES}	THRsyn ^{ES}	: ATP + L-Homoserine -> ADP + L-Threonine + Orthophosphate
CYS4	CYS4	: Homocysteine + L-Serine -> L-Cystathionine
GLY1	GLY1	: Acetaldehyde + Glycine -> L-Threonine
CHA1, ILV1	CHA1o	: L-Threonine -> 2-Oxobutanoate + NH3
CHA1, SDL1	CHA1p	: L-Serine -> NH3 + Pyruvate
*U46_*		: L-Threonine + NAD+ -> Acetate + Glycine + NADH

		# METHIONINE METABOLISM (5/5)
YFR055W		: L-Cystathionine -> Homocysteine + NH3 + Pyruvate
SAH1	SAH1	: S-Adenosyl-L-homocysteine -> Adenosine + Homocysteine
*U47_*	METsyn	: 5-Methyltetrahydrofolate + Homocysteine -> L-Methionine + Tetrahydrofolate
CYS3	CYS2	: L-Cystathionine -> 2-Oxobutanoate + L-Cysteine + NH3
SAM1-2	SAM	: ATP + L-Methionine -> Orthophosphate + Pyrophosphate + S-Adenosyl-L-methionine
		# CYSTEINE BIOSYNTHESIS (2)
*U48_*	U48_	: L-Serine + Acetyl-CoA -> CoA + O-Acetyl-L-serine
YGR012W	CYSsyn	: O-Acetyl-L-serine + Hydrogen sulfide -> Acetate + L-Cysteine
		# BRANCHED CHAIN AMINO ACID METABOLISM (VALINE, LEUCINE AND ISOLEUCINE) (5/11)
LEU2, BAT1-2 [ES]	LEUsyn [ES]	: L-Glutamate + NAD+ + 2-Isopropylmalate -> 2-Oxoglutarate + CO2 + L-Leucine + NADH
LEU4	LEU4	: Acetyl-CoA + (R)-2-Oxoisovalerate -> CoA + 2-Isopropylmalate
BAT1-2, ILV2-6, ILV5, ILV3 [ES]	ILEsyn [ES]	: 2-Oxobutanoate + L-Glutamate + NADPH + Pyruvate -> 2-Oxoglutarate + CO2 + L-Isoleucine + NADP+
BAT2	VALsyn	: (R)-2-Oxoisovalerate + L-Glutamate <-> 2-Oxoglutarate + L-Valine
ILV2-6, ILV5, ILV3 [ES]	OIVsyn [ES]	: NADPH + 2 Pyruvate -> (R)-2-Oxoisovalerate + CO2 + NADP+
		# LYSINE BIOSYNTHESIS/DEGRADATION (4/10)
U49_, LYS4-12, U50_, LYS21-20 [ES]	OXAsyn [ES]	: 2-Oxoglutarate + Acetyl-CoA + NAD+ -> 2-Oxoadipate + 2 CO2 + CoA + NADH
U51_,LYS9-1 [ES]	LYSsyn [ES]	: 2-Oxoadipate + L-2-Aminoadipate 6-semialdehyde + 2 L-Glutamate + NAD+ + NADPH <-> 2 2-Oxoglutarate + L-2-Aminoadipate + L-Lysine + NADH + NADP+
LYS2-5	LYS2_1	: ATP + L-2-Aminoadipate + NADPH -> AMP + L-2-Aminoadipate 6-semialdehyde + NADP+ + Pyrophosphate
LYS2-5	LYS2_2	: ATP + L-2-Aminoadipate + NADH -> AMP + L-

		2-Aminoadipate 6-semialdehyde + NAD+ + Pyrophosphate
		# ARGININE METABOLISM (6/10)
ARG5,6-8 [ES]	*ARG5,6-8* [ES]	: ATP + 1 L-Glutamate + NADPH + N-Acetyl-L-glutamate -> 2-Oxoglutarate + ADP + NADP+ + Orthophosphate + N2-Acetyl-L-ornithine
ECM40	ECM40	: N2-Acetyl-L-ornithine + L-Glutamate -> L-Ornithine + N-Acetyl-L-glutamate
URA2, CPA1-2	CABsyn	: 2 ATP + CO2 + L-Glutamine -> 2 ADP + Carbamoyl phosphate + L-Glutamate + Orthophosphate
CAR2	CAR2	: 2-Oxoglutarate + L-Ornithine -> L-Glutamate + L-Glutamate 5-semialdehyde
ARG1-4-3 [ES]	ARGsyn [ES]	: ATP + Carbamoyl phosphate + L-Aspartate + L-Ornithine -> AMP + Fumarate + L-Arginine + Orthophosphate + Pyrophosphate
CAR1	CAR1	: L-Arginine -> L-Ornithine + Urea
		# HISTIDINE METABOLISM (2/9)
HIS1-4-6-3-5-2 [ES]	HISsyn [ES]	: 5-Phospho-alpha-D-ribose 1-diphosphate + ATP + 2 NAD+ + L-Glutamate + D-erythro-1-(Imidazol-4-yl)glycerol 3-phosphate -> 2-Oxoglutarate + L-Histidine + 2 NADH + Orthophosphate + 2 Pyrophosphate + "N-(5'-Phospho-D-1'-ribulosylformimino)-5-amino-1-(5""-phospho-D-ribosyl)-4-imidazolecarboxamide"
HIS7	HIS7	: "N-(5'-Phospho-D-1'-ribulosylformimino)-5-amino-1-(5""-phospho-D-ribosyl)-4-imidazolecarboxamide" + L-Glutamine -> L-Glutamate + 1-(5'-Phosphoribosyl)-5-amino-4-imidazolecarboxamide + D-erythro-1-(Imidazol-4-yl)glycerol 3-phosphate
		# PHENYLALANINE, TYROSINE AND TRYPTOPHAN BIOSYNTHESIS (AROMATIC AMINO ACIDS) (8/23)
ARO3-4-1-2 [ES]	CHOsyn [ES]	: ATP + D-Erythrose 4-phosphate + NADPH + 2 Phosphoenolpyruvate -> ADP + Chorismate + NADP+ + 4 Orthophosphate
ARO9, PHA2 [ES]	PHEsyn [ES]	: L-Glutamate + Prephenate -> 2-Oxoglutarate + CO2 + L-Phenylalanine
ARO8-9, AAT1-2	TYRsyn	: 3-(4-Hydroxyphenyl)pyruvate + L-Glutamate -> 2-Oxoglutarate + L-Tyrosine
TRP2-3	TRP23	: Chorismate + L-Glutamine -> Anthranilate + L-Glutamate + Pyruvate

TRP4-1-3-2-5 [ES]	TRPsyn [ES]	: 5-Phospho-alpha-D-ribose 1-diphosphate + Anthranilate + L-Serine -> CO2 + D-Glyceraldehyde 3-phosphate + L-Tryptophan + Pyrophosphate
BNA2-3 [ES]	KYNsyn [ES]	: L-Tryptophan + Oxygen -> Formate + L-Kynurenine
BNA5		: L-Kynurenine -> Anthranilate + L-Alanine
*BNA4-5-1, U54_-55_-56_* [ES]	BNA [ES]	: L-Kynurenine + NAD+ + 2 NADPH + 2 Oxygen -> 2-Oxoadipate + CO2 + L-Alanine + NADH + 2 NADP+ + NH3
		# PROLINE BIOSYNTHESIS (5/5)
PRO1	PRO1	: ATP + L-Glutamate -> ADP + alpha-D-Glutamyl phosphate
PRO3	PROsyn	: (S)-1-Pyrroline-5-carboxylate + NADPH -> L-Proline + NADP+
PRO3	PRO3_2	: L-1-Pyrroline-3-hydroxy-5-carboxylate + NADPH -> trans-4-Hydroxy-L-proline + NADP+
PRO3	PRO3_3	: L-1-Pyrroline-3-hydroxy-5-carboxylate + NADH -> trans-4-Hydroxy-L-proline + NAD+
PUT1	PUT1	: L-Proline + NAD+ -> (S)-1-Pyrroline-5-carboxylate + NADH
		# METABOLISM OF OTHER AMINO ACID
		# GLUTATHIONE BIOSYNTHESIS (1/1)
GSH1-2 [ES]	GSH [ES]	: 2 ATP + Glycine + L-Cysteine + L-Glutamate -> 2 ADP + Glutathione + 2 Orthophosphate
		# METABOLISM OF COFACTORS, VITAMINS, AND OTHER SUBSTANCES
		# THIAMINE (VITAMIN B1) METABOLISM (3/3)
*U76_-77_* [ES]	THI [ES]	: Thiamin diphosphate + ADP -> Thiamin + ATP + Orthophosphate
THI80	THI80	: ATP + Thiamin -> AMP + Thiamin diphosphate
		# PANTOTHENATE AND COA BIOSYNTHESIS (1/3)
*U86_-222_-85_* [ES]	AON [ES]	: L-Alanine + 3 Malonyl-CoA -> 3 CO2 + 3 CoA + H+EXT + AONAxt

		# FOLATE BIOSYNTHESIS (3/3)
*U88_*	U88	: 4-amino-4-deoxychorismate -> Pyruvate + 4-Aminobenzoate
FOL1	FOL1a	: 4-Aminobenzoate + 2-Amino-7,8-dihydro-4-hydroxy-6-(diphosphooxymethyl)pteridine -> Pyrophosphate + Dihydropteroate
FOL1	FOL1b	: 4-Aminobenzoate + 2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine -> Dihydropteroate
		# COENZYME A BIOSYNTHESIS (2/10)
*ECM31, ILV5, PAN5-6, U98_* [ES]	PANT [ES]	: (R)-2-Oxoisovalerate + 5,10-Methylenetetrahydrofolate + ATP + L-Aspartate + NADPH -> (R)-Pantothenate + AMP + CO2 + NADP+ + Pyrophosphate + Tetrahydrofolate
YDR531W, U92_-93_-94_-96_, PPT2 [ES]	PPT [ES]	: (R)-Pantothenate + 3 ATP + CTP + L-Cysteine -> Acyl-carrier protein + Adenosine 3',5'-bisphosphate + 2 ADP + CMP + CO2 + 2 Pyrophosphate
		# NAD BIOSYNTHESIS (3/3)
PNC1	PNC1	: Nicotinamide <-> Nicotinate + NH3
NPT1	NPT1	: Nicotinate + 5-Phospho-alpha-D-ribose 1-diphosphate -> Nicotinate D-ribonucleotide + Pyrophosphate
*U99_*	U99	: L-Aspartate + FAD -> FADH2 + a-Iminosuccinate
		# MEMBRANE TRANSPORT
		# PLASMA MEMBRANE TRANSPORT (3/3)
AGP1-3, STL1, GAP1, DIP5	GLUsc	: GLUxt <-> L-Glutamate
JEN1	LACsc	: LACxt + H+EXT <-> (R)-Lactate
*U231_*	MALsc	: H+EXT + MALxt <-> Malate
*U155_*	U155	: 2-Oxoglutarate + MALxt <-> Malate + AKGxt
		# AMINO ACIDS (21/21)
GAP1, DIP5, AGP1, TAT2, PUT4	ALAsc	: H+EXT + ALAxt <-> L-Alanine
GAP1, AGP1, GNP1, DIP5	ASNsc	: H+EXT + ASNxt <-> L-Asparagine
AGP3, GAP1,	ASPsc	: H+EXT + ASPxt <-> L-Aspartate

DIP5		
GAP1, GNP1, BAP2-3, TAT1-2	CYSsc	: H+EXT + CYSxt <-> L-Cysteine
GAP1, TAT2, DIP5, PUT4	GLYsc	: H+EXT + GLYxt <-> Glycine
GAP1, AGP1, GNP1, DIP5	GLNsc	: H+EXT + GLNxt <-> L-Glutamine
HIP1, GAP1, AGP1, TAT1	HISsc	: H+EXT + HISxt <-> L-Histidine
TAT1, GAP1, AGP1, BAP2-3	ILEsc	: H+EXT + ILExt <-> L-Isoleucine
TAT1, GAP1, AGP1, BAP2-3, GNP1	LEUsc	: H+EXT + LEUxt <-> L-Leucine
GAP1, AGP1, GNP1, BAP2-3, MUP1-3	METsc	: H+EXT + METxt <-> L-Methionine
GAP1, AGP1, TAT2, BAP2-3	PHEsc	: H+EXT + PHExt <-> L-Phenylalanine
GAP1, PUT4	PROsc	: H+EXT + PROxt <-> L-Proline
TAT1-2, GAP1, BAP2-3	TRPsc	: H+EXT + TRPxt <-> L-Tryptophan
TAT1-2, GAP1, AGP1, BAP2-3	TYRsc	: H+EXT + TYRxt <-> L-Tyrosine
GAP1, AGP1, BAP2-3, TAT1	VALsc	: H+EXT + VALxt <-> L-Valine
AGP1-3, GNP1, GAP1, DIP5	SERsc	: H+EXT + SERxt <-> L-Serine
TAT1, AGP1, GAP1, GNP1	THRsc	: H+EXT + THRxt <-> L-Threonine
LYP1, GAP1	LYSsc	: H+EXT + LYSxt <-> L-Lysine
SAM3	SAMsc	: H+EXT + SAMxt <-> S-Adenosyl-L-methionine
PUT4, UGA4	GABAsc	: H+EXT + GABAxt <-> 4-Aminobutanoate
GAP1, CAN1	ORNsc	: H+EXT + ORNxt <-> L-Ornithine
		# METABOLIC BY-PRODUCTS
*U205_*	SUCsc	: H+EXT + SUCCxt <-> Succinate
JEN1	PYRsc	: H+EXT + PYRxt <-> Pyruvate
*U206_*	CITsc	: H+EXT + CITxt <-> Citrate
		# OTHER COMPOUNDS
*U207_*	FUMsc	: H+EXT + FUMxt <-> Fumarate
*U208_*	MYRsc	: C140xt <-> Myristic acid
*U213_*	AKGsc	: H+EXT + AKGxt <-> 2-Oxoglutarate
THI7, YOR071C, YOR192C	THIsc	: THMxt + H+EXT <-> Thiamin
*U226_*	PIMsc	: PIMExt <-> Pimelic Acid

F.2. Magnitude of Regulation for the Reactions of ENZSUB-3 Model

Magnitude of regulation for the reactions of the metabolic network, ENZSUB3, at metabolic (zRE) and transcriptional (zGE) level is given in Table F.1 for the effect of very high-gravity (VHG) fermentation media on laboratory (CEN) and industrial Red Star (RS) strains. Regulation is considered to be significant for zRE, zGE > 1.28 ($p < 0.10$). In the table, regulation type is coded as follows; H means hierarchical (transcriptionally) change, M stands for metabolic change. HM shows the reactions for which there is regulation in both levels. The reactions were grouped with respect to the pathways they belong. SM is the abbreviation for standard media. ES in the reaction names stands for enzyme subset.

Table F.1. Magnitude of regulation for the reactions of ENZSUB3

	CEN (VHG vs SM)			**RS (VHG vs SM)**		
	zRE	**zGE**		**zRE**	**zGE**	
GLYCOLYSIS						
GPM	-0.95	0.32	-	-0.90	1.49	H
ENO	-1.69	2.18	H	-1.76	-1.63	-
PYK	-0.04	0.82	-	-0.05	1.26	-
PDA	-0.13	1.00	-	-0.01	1.73	H
TCA CYCLE						
CIT	-0.27	1.13	-	-0.01	-0.09	-
ACO	-0.19	1.32	H	0.00	1.46	H
IDH	-0.73	-0.83	-	-0.62	-0.98	-
IDP^{ES}	-0.08	0.41	-	0.02	1.28	H
KGD	-0.24	0.77	-	-0.02	-0.31	-
LSC1	-0.73	1.37	H	-0.56	0.84	-
LSC2	0.31	0.68	-	0.44	0.47	-
SDH	0.72	-0.27	-	-0.02	1.61	H
FUM	0.14	1.54	H	-0.78	0.72	-
MDH	-1.32	1.25	-	-1.35	1.16	-
ANAPLEROTIC REACTIONS.						
ICL	0.63	0.03	-	0.52	1.49	H
MLS	0.40	-0.84	-	0.52	0.02	-

Table F.1. continued

PCK1	-1.22	0.70	-	-1.15	1.26	-
PYC	-0.12	-1.05	-	-0.05	0.29	-
MAE1	0.68	1.52	H	0.71	1.22	-
PDC	0.57	0.07	-	0.68	1.74	H
FATTY ACID						
POX	-1.13	0.78	-	-0.92	1.28	H
CHO1	0.43	2.04	H	0.35	-1.42	-
ERG [ES]	-0.39	1.95	H	-0.31	0.85	-
PURINE- PYRIMIDINE						
AIC[ES]	0.32	1.52	H	-0.03	0.53	-
AMP[ES]	1.13	1.63	H	0.96	1.32	H
GUA1	0.76	1.54	H	0.90	1.55	H
URA [ES]	0.81	1.77	H	0.90	0.61	-
URA78	0.08	0.97	-	0.38	1.00	-
GLUTAMATE PATHWAY						
GAD1	1.29	1.46	HM	1.43	1.21	M
UGA [ES]	2.18	1.69	HM	2.23	0.38	M
GFA1	1.19	0.32	-	1.18	-0.63	-
PUT2	0.75	-0.16	-	0.74	2.15	H
GLT1	1.25	2.82	H	1.24	1.00	-
GDH2	0.56	1.83	H	0.60	1.67	H
GDH13	1.20	0.25	-	1.24	0.91	-
GLN1	0.93	1.36	H	0.95	0.70	-
ALANINE-ASPARTATE PATHWAY						
AAT	1.25	1.50	H	1.24	2.30	H
ALT	2.50	2.48	**HM**	2.48	1.80	HM
ASN	1.84	2.30	HM	1.85	0.54	M
MHT	-1.00	1.45	H	-0.99	0.35	-
ASP3-1	1.47	1.45	HM	1.41	1.34	HM
GLYCINE-SERINE-THREONINE P.						
SERsyn [ES]	0.93	1.62	H	0.80	0.43	-
GLYsyn	1.15	1.69	H	1.03	1.46	H
AGX1	2.45	0.86	M	2.31	2.21	**HM**

Table F.1. continued

GCV1	-0.17	2.53	H	0.01	0.81	-
HOM [ES]	0.41	0.30	-	0.47	1.44	H
THRsyn [ES]	0.42	1.30	H	-0.23	0.98	-
CYS4	-0.01	1.29	H	-0.31	0.43	-
GLY1	1.41	0.41	M	0.46	0.50	-
CHA1o	0.02	1.11	-	-0.96	1.52	H
CHA1p	1.47	0.93	M	1.14	0.54	-
METHIONINE-CYSTEINE PATHWAY						
YFR055W	-0.05	2.03	H	-0.05	0.44	-
SAH1	-0.83	1.74	H	-0.75	1.42	H
CYS3	-1.83	0.38	-	-1.88	0.01	-
SAM	-0.45	1.16	-	-0.34	0.81	-
CYSsyn	-1.34	-0.83	-	-1.27	1.11	-
LEUCINE-ISOLEUCINE-VALINE P.						
LEUsyn [ES]	1.07	1.54	H	1.38	0.91	M
LEU4	-0.60	0.88	-	-0.18	1.58	H
ILEsyn [ES]	1.15	1.66	H	1.42	1.90	HM
VALsyn	1.20	0.98	-	1.24	-0.33	-
OIVsyn [ES]	0.04	1.83	H	0.02	2.13	H
LYSINE PATHWAY						
OXAsyn [ES]	0.32	1.18	-	-0.06	1.17	-
LYSsyn [ES]	1.67	2.46	HM	1.16	0.90	-
LYS2_1	0.22	1.21	-	0.36	1.29	H
LYS2_2	-0.30	1.21	-	-0.17	1.29	H
ARGININE PATHWAY						
ARG5,6-8 [ES]	1.20	1.56	H	1.28	0.88	M
ECM40	1.85	2.39	HM	0.61	1.32	H
CABsyn	0.77	1.28	H	0.89	0.34	-
CAR2	1.74	0.57	M	0.62	-0.01	-
ARGsyn [ES]	0.62	0.68	-	-0.45	0.44	-
CAR1	0.56	1.74	H	-0.74	1.27	-
HISTIDINE PATHWAY						
HISsyn [ES]	0.48	1.82	H	0.75	1.25	-

Table F.1. continued

HIS7	1.00	1.32	H	1.09	2.06	H
AROMATIC AMINOACIDS P.						
CHOsyn [ES]	-0.61	1.66	H	-0.51	1.41	H
PHEsyn [ES]	1.36	1.66	HM	1.65	0.98	M
TYRsyn	1.22	1.78	H	1.84	3.41	HM
TRP23	1.67	1.31	HM	1.65	0.78	M
TRPsyn [ES]	0.44	1.74	H	0.45	2.24	H
KYNsyn [ES]	0.09	2.15	H	0.19	1.60	H
BNA5	0.65	1.97	H	0.69	-1.11	-
BNA [ES]	0.39	2.09	H	-0.06	-0.16	-
PROLINE PATHWAY						
PRO1	0.49	1.83	H	0.57	2.05	H
PROsyn	0.73	0.87	-	0.72	0.84	-
PRO3_2	-0.69	0.87	-	-0.71	0.84	-
PRO3_3	-1.44	0.87	-	-1.45	0.84	-
PUT1	-0.02	0.56	-	-0.02	1.67	H
SECRETION REACTIONS						
GLUsc	1.85	1.17	M	1.81	0.79	M
LACsc	1.43	0.81	M	1.45	0.86	M
ALAsc	1.43	1.90	HM	1.45	0.75	M
ASNsc	0.81	-0.25	-	-0.03	0.69	-
ASPsc	1.43	2.09	HM	1.29	0.86	M
CYSsc	-1.46	1.38	H	-1.47	0.53	-
GLYsc	1.43	1.66	HM	1.45	0.80	M
GLNsc	1.14	-0.25	-	-0.03	0.69	-
HISsc	0.88	-0.80	-	-0.03	0.88	-
ILEsc	0.86	1.55	H	-0.03	0.56	-
LEUsc	1.14	1.46	H	1.01	0.66	-
METsc	-1.46	1.94	H	-1.47	0.98	-
PHEsc	1.21	1.58	H	-0.03	0.21	-
PROsc	1.43	1.40	HM	1.45	0.90	M
TRPsc	-0.48	1.52	H	-0.50	0.37	-
TYRsc	-0.65	1.56	H	1.15	0.43	-

Table F.1. continued

VALsc	1.38	1.55	HM	-0.03	0.56	-
SERsc	-0.01	-0.80	-	-0.30	0.68	-
THRsc	1.14	1.00	-	-0.97	0.58	-
LYSsc	-0.01	1.84	H	-0.03	1.00	-
SAMsc	-0.47	1.54	H	-0.42	1.27	-
GABAsc	-0.15	0.80	-	-0.03	2.02	H
ORNsc	-0.06	1.45	H	-1.47	1.59	H
PYRsc	1.43	0.81	M	1.45	0.86	M
THIsc	-0.84	1.62	H	-0.74	0.87	-
GSH^{ES}	0.40	1.15	-	0.42	0.63	-
NPT1	-0.78	1.55	H	-0.65	1.20	-
OPI^{ES}	-0.46	0.48	-	-0.34	0.69	-
DLD	1.21	0.00	-	1.24	1.18	-
$PANT^{ES}$	-0.24	2.24	H	0.01	1.24	-
FOL1a	-0.76	2.05	H	0.63	0.47	-
FOL1b	-0.85	2.05	H	0.71	0.47	-
PNC1	-0.81	1.59	H	-0.73	1.60	H
THI80	-0.78	1.29	H	-0.66	0.12	-
PPT^{ES}	-0.74	1.84	H	-0.43	0.54	-

APPENDIX G: REPRXN MX SOFTWARE PACKAGE

A software package, called RepRxn MX, was developed which automates the method discussed in chapter 5. The package works under MATLAB 7.0, and consists of 13 M-files. A graphical user interface (GUI) was designed for the package (Figure G.1). Basically, a) it follows the roadmap depicted in Figure 5.2 to construct ENZSUB-3 metabolic network. Two input files are required for the this part to function: a text file including all the reactions occurring in the metabolism of interest, and another text file with the names of measured metabolites. b) it performs reporter reaction analysis based on the derived network, ENZSUB-3. This step requires external input of the p-values of metabolites within the network.

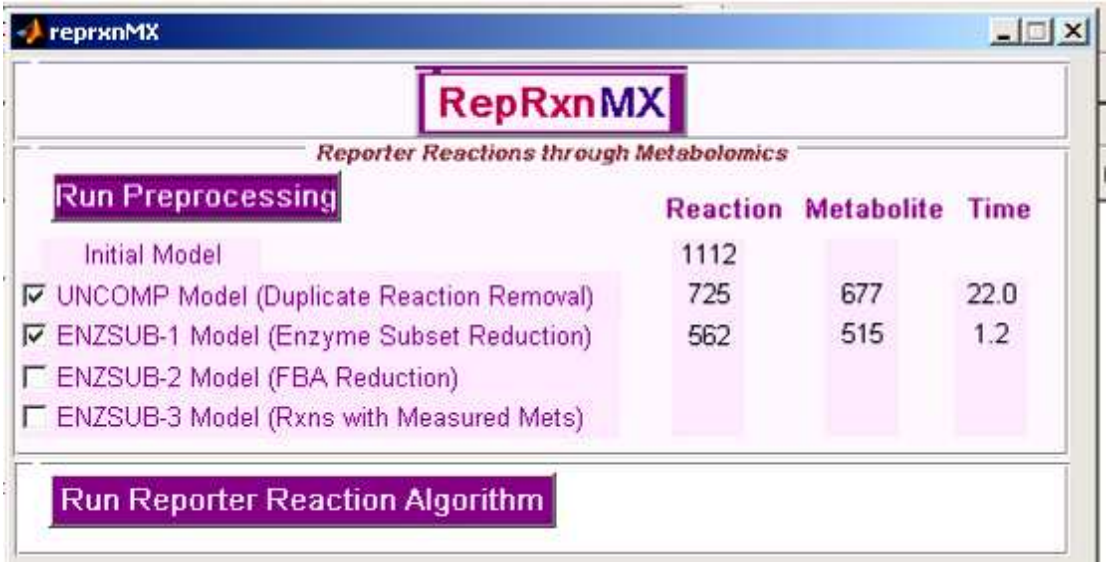


Figure G.1. Graphical user interface of the developed MATLAB package, RepRxn MX

G.1. Requirement for Model Reactions

All the metabolic reactions for the organism of interest (in this case S. cerevisiae) must be supplied in a text file in the following format;

reactionname : A + B -> C + D

In other words, a name must be assigned to each reaction, followed by colon mark ‘:’, Then, the reactants are written separated by summation mark ‘+’. The reactants and products must be separated by ‘->’ or ‘<->’ depending on if the reaction is irreversible or reversible.

Whenever required, comment lines can be introduced into the input text files by using ‘#’ sign. All lines which start with this sign will not be read and executed by the MATLAB codes.

The input metabolic network must be compatible with the nature of measured metabolite data. That is, if the data does not make any distinction between the presence of a metabolite in different compartments (such as cytosolic or mitochondrial pools), the reactions supplied must not be compartmentalized. The opposite case is also valid. In the case of present data discussed in this chapter, there is a distinction only between intracellular and extracellular compartments whereas the yeast genome-scale model includes three compartments (mitochondria, cytosol and external space). Therefore, the cytosolic/mitochondrial compartmentation of the model was removed and metabolites present in both of these compartments were represented as one, with their corresponding reactions conserved. The package assumes that this preprocessing, if necessary, is manually done before. For the present analysis, rearranged reaction list is available through a text file called “SC_forster03_uncomp.txt”.

G.2. Duplicate Reaction Removal

Duplicate reactions due to isoenzyme availability in the supplied text file of metabolic network, if any, are removed by keeping only one of them since metabolome data does not provide any special information to interpret isoenzymes. ‘A_noduplic.m’ performs this preprocessing. Below is the output of the code;

A. REMOVAL OF DUPLICATE REACTIONS: started ...
Model initially includes 1112 reactions.
Input Model was reduced by removing 387 duplicate reactions ...

New model was written into a new text file: "SC_forster03_nodup.txt"...
Metabolite List was constructed for NoDuplic Model ...
Metabolite List was written into "metlist_noduplic.txt" file..
DUPLICATE REACTION REMOVAL IS COMPLETE! :
NEW MODEL INCLUDES 725 REACTIONS AND 677 METABOLITES..
Elapsed time is 8.76 seconds.

As seen from the output, the code also constructs the list of metabolites for the new model using a function “func_metlist.m”. In this way, UNCOMP model was constructed (Figure 5.X). The execution time is 8.7 sec. in a computer with Pentium IV 3 Ghz processor and 1 GB memory.

G.3. Identification and Combination of Enzyme Subsets

UNCOMP model is processed by another MATLAB script called “B_enzsub.m” to build ENZSUB-1 model. METATOOL can also be used to perform this task. However, it is relatively inconvenient to convert the text file into METATOOL input file format in an automated fashion, and to process the output file to reconstruct the new reaction list.

G.3.1. Construction of Stoichiometric Matrix from NoDuplic Model

As detailed in the following subsection, enzyme subset combination analysis depends on the stoichiometric coefficients. Therefore, the model stored in the text file (SC_forster03_nodup.txt) is converted into a stoichiometric matrix representation by a script “text2mat_v3_sc.txt”. The script follows the following steps to construct the matrix;

- Reaction list is read from the text file.
- Reactants and products are separated into different variables.
- Reversibility information of the reactions is stored into a variable by scanning the direction of the arrows in the reactions. Thereby, lower and upper bounds of reactions are assigned.
- Metabolite names list is constructed from reaction list.

- A function “coeff_met.m” is used to construct stoichiometric coefficient matrix for reactants and products.

The output of the code is given below:

NODUPLIC model is being converted into stoichiometric representation...
Reaction List was read from the file: 725 Reactions
Metabolite List was constructed from reaction list: 677 Metabolites
Number of Irreversible Reactions: 509
Number of Reversible Reactions : 216
Matrix conversion for REACTANTS and PRODUCTS were completed ...
STOICHIOMETRIC MATRIX FOR NODUPLIC MODEL IS READY!. : 677 x 725
Elapsed time is 10.93 seconds.

G.3.2. Processing of Stoichiometric Matrix for Enzyme Subset Reduction

Next, “B_enzsub.m” script constructs ENZSUB-1 model. The following strategy is followed for the detection and further combination of linear reaction paths;

- Number of reactions in which each metabolite participates is calculated.
- The metabolites which participate in only two reactions are identified.
- Among those metabolites, the ones which have the same sign as coefficients in both reactions are discarded from the list. The reason is that combination (summation) of these two reactions will not result in the disappearance of the metabolite since both reactions either produce or consume it.
- Additionally, if the metabolite identified is among the measured metabolites, it is also discarded. The underlying reasoning behind enzyme subset combination is that the metabolites in linear pathways will not be disturbed by any other reactions; thereby the levels of these metabolites will not be perturbed, remaining constant. This is a reasonable assumption in the case of absence of data for these metabolites. However, whenever measurement is available for such metabolites,

the combination is not performed with the aim of reflecting the measurement information into the analysis.

- Biomass and the precursors in the biomass equation is also removed from the list if they are included. The precursors must be produced for the cell to function. Therefore, their disappearance must not be allowed.

As a result of these steps, the metabolite list which includes names of metabolites in the linear pathways is ready. For each metabolite in the list; the two reactions it particiapates (i.e. the corresponding columns in the stoichiometric matrix) are summed up. This new reaction is added at the end of the stoichiometric matrix. The original reactions are discarded. Since the summation leads to the disappearance of the metabolite of interest, the corresponding row is also removed from the matrix. Additionally, it is also checked whether the combination leads to the disappearance of any other measured metabolites. If this is the case, the combination is not performed. As a result, ENZSUB-1 model is reconstructed with 562 reactions and 515 metabolites. It is kept in the stoichiometric matrix form for the purpose of the following preprocessing steps.

B. DETECTION AND COMBINATION OF ENZYME SUBSETS: started ...
ENZYME-SUBSET COMBINED MODEL IS READY : ENZSUB-1.
ENZSUB-1 MODEL HAS 562 REACTIONS AND 515 METABOLITES!
Elapsed time is 1.125 seconds.

G.4. Construction of ENZSUB-2 from ENZSUB-1

G.4.1. Use of FBA to Identify Inactive Reactions

The next step is to apply linear programming (FBA) to the model to identify reactions which are not active in glucose growth. The corresponding Matlab script is “C_esr_FBA.m”. It uses the stoichiometric matrix, lower and upper bounds of reactions, and the objective function as the input to identify inactive reactions under the experimental conditions of interest. In this case, with the sole carbon source of glucose, FBA is executed

for aerobic and anaerobic conditions. After constraining the optimal biomass growth rate between its optimum value and 50% of the optimum, each flux in the model is maximized and minimized to identify the extreme values of fluxes that reaction can carry. The set of reactions identified to have only zero fluxes are saved into files called ‘rxn_omit_aer.mat’ and ‘rxn_omit_anaer.mat’. linear programming was performed by using the built-in “linprog” function of MOSEK Optimization Package.

C. USE OF FBA TO IDENTIFY INACTIVE REACTIONS: started ...
117 external metabolites were identified ...
FBA for aerobic conditions starts ...
260 inactive reactions were identified and saved for Aerobic conditions ...
FBA for anaerobic conditions starts ...
258 inactive reactions were identified and saved for Anaerobic conditions ...
Elapsed time is 175.27 seconds.

G.4.2. Omission of FBA-derived Inactive Reactions

The matlab script ‘rxn_omission.m’ reads the previously saved files which includes the names of inactive reactions; and omit them.

There are 256 inactive reactions to be omitted ...
Elapsed time is 2.99 seconds.

G.4.3. Restoring Reactions with Measured Metabolites

Another script is required to check whether the procedure led to the omission of measured metabolites or not. In other words, there may be inactive reactions which were discarded, however these reactions may involve measured metabolites. “comp_measmets.m” makes this check by identifying such metabolites. If there is any, the

reactions including these metabolites are stored back into the model, thereby leading to the final form of ENZSUB-2 model.

Measured metabolite names are read ...
There are 84 measured metabolites..
Number of metabolites in FBA reduced model is 286 ...
13 measured metabolites were identified as missing ...
ENZSUB-2 Model was finalized :
It includes 389 reactions and 303 metabolites ..
Elapsed time is 5.76 seconds.

G.5. Finalization of Preprocessing: ENZSUB-3 model

ENZSUB-2 model is scanned by “enzsub3_formn.m” code to identify reactions none of whose participating metabolites are measured. These reactions are discarded, leading to the finalization of the preprocessing step. Then, the model is converted to stoichiometric matrix in order to represent it in GML format by using the executable file, ‘GML_maker.exe’.

---

244 reactions with no measured metabolites were discarded
ENZSUB-3 Model was constructed:
Model has 183 reactions & 145 metabolites ..
Measured metabolite coverage of ENZSUB-3 is 46.4 percent ..
Elapsed time is 2.44 seconds.

ENZSUB3 model is being converted into stoichiometric representation...
Reaction List was read from the file: 145 Reactions
Metabolite List was constructed from reaction list: 183 Metabolites
Number of Irreversible Reactions: 97
Number of Reversible Reactions : 48

Matrix conversion for REACTANTS and PRODUCTS were completed ...
STOICHIOMETRIC MATRIX FOR NODUPLIC MODEL IS READY!. : 183 x 145
Elapsed time is 1.15 seconds.

Conversion of stoichiometric matrix to GML format starts ...
GML file was created ..

G.6. Reporter Reaction Algorithm

After the first part of the algorithm, preprocessing, is over, the final ENZSUB-3 model is represented as a graph. A matlab script is used to combine graph representation of ENZSUB-3 with the p-values of measured metabolites supplied in a .mat file format. Additionally, a text file including a set of p-values generated by analyzing the unknonw peaks in GC-MS spectra must be supplied, for random assignment to the unmeasured metabolites available in the model. The script applies the algorithm presented in Chapter five, and gives the list of reactions in the model and corresponding z-scores. The user can use a threshold (typically z = 1.28, corresponding to p = 0.10), to identify reporter reactions.

REFERENCES

Åkesson, M., J. Förster and J. Nielsen, 2004, "Integration of gene expression data into genome-scale metabolic models", *Metabolic Engineering,* Vol. 6, pp. 285-293.

Allen, J., H. M. Davey, D. Broadhurst, J. K. Heald, J. J. Rowland, S. G. Oliver and D. B. Kell, 2003, "High-throughput classification of yeast mutants for functional genomics using metabolic footprinting", *Nature Biotechnology,* Vol. 21, pp. 692-696.

Almaas, A., B. Kovacs, T. Vicsek, Z. N. Oltvai and A. L. Barabasi, 2004, "Global organization of metabolic fluxes in the bacterium *Escherichia coli*", *Nature,* Vol. 427, pp. 839-843.

Altıntaş M. M., B. Kırdar, Z. İ. Önsan and K. Ö. Ülgen, 2002, "Cybernetic modelling of growth and ethanol production in a recombinant *S.cerevisiae* strain secreting a bifunctional fusion protein', *Process Biochemistry*, Vol. 37, pp. 1439-1445.

Arvanitidis A and J. J. Heinisch, 1994, "Studies on the function of yeast phosphofructokinase subunits by in vitro mutagenesis", *Journal of Biological Chemistry*, Vol. 269, pp. 8911-8918.

Ataullakhanov, F. I., V. M. Vitvitsky, S. V. Komarova and E. V. Mosharov, 1996, "Energy-dependent processes and adenylate metabolism in human erythrocytes", *Biochemistry (Moscow),* Vol. 61, pp. 197-203.

Bailey, J. E., 1991, "Towards a science of metabolic engineering" *Science,* Vol. 252, pp. 1668-675.

Bakker, B. M., K. M. Overkamp, A. J. A. van Maris, P. Kötter, M. A. H. Luttik, J. P. van Dijken and J. T. Pronk, 2001, "Stoichiometry and compartmentation of NADH metabolism in *Saccharomyces cerevisiae*", *FEMS Microbiology Reviews*, Vol. 25, pp.15-37.

Barrett, C. L., N. D. Price and B. O. Palsson, 2006, "Network-level analysis of metabolic regulation in the human red blood cell using random sampling and singular value decomposition", *BMC Bioinformatics*, Vol. 7, pp. 132.

Baynes, J. and M. Dominiczak, 1999, *Medical Biochemistry*, Mosby, England.

Bino, R. J., R. D. Hall, O. Fiehn, J. Kopka, K. Saito, J. Draper, B. J. Nikolau, P. Mendes, U. Roessner-Tunali, M. H. Beale, R. N. Trethewey, B. M. Lange, E. S. Wurtele and L. W. Sumner, 2004, "Potential of metabolomics as a functional genomics tool", *Trends in Plant Sciences*, Vol. 9, pp. 418-425.

Blank, L. M., L. Kuepfer and U. Sauer, 2005, "Large-scale ^{13}C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast", *Genome Biology*, Vol. 6, R49.

Boles, E., P. De Jong-Gubbels and J. T. Pronk, 1998, "Identification and characterization of MAE1, the *Saccharomcyes cerevisiae* structural gene encoding mitochondrial malic enzyme", *Journal of Bacteriology*, Vol. 180, pp. 2875-2882.

Bolstad, B. M., R. A. Irizarry, M. Astrand and T. P. Speed, 2003, "A comparison of normalization methods for high density oligonucleotide array data based on bias and variance", *Bioinformatics*, Vol. 19, pp. 185-193.

Bonarius, H. P. J., V. Hatzimanikatis, K. P. H. Meesters, C. D. de Gooijer, G. Schmid and J. Tramper, 1996, "Metabolic Flux Analysis of Hybridoma cells in different culture media using mass balances", *Biotechnology & Bioengineering*, Vol. 50, pp. 299-318.

Borodina, I. and J. Nielsen, 2005, "From genomes to *in silico* cells via metabolic networks", *Current Opinion in Biotechnology*, Vol. 16, pp. 350-355.

Bossi, D and B. Giardina, 1996, "Red cell physiology", *Molecular Aspects in Medicine*, Vol. 17, pp. 117-128.

Boubekeur, S., O. Bunoust, N. Camougrand, M. Castroviejo, M. Rigoulet and B. Guerin, 1999, "A mitochondrial pyruvate dehyrdrogenase bypass in the yeast *Saccharomyces cerevisiae*", *Journal of Biological Chemistry*, Vol. 274, pp. 21044-21048.

Breitling R., P. Armengaud, A. Amtmann and P. Herzyk, 2004, "Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments" *FEBS Letters*, Vol. 573, pp.83-92.

Brem, R. B., G. Yvert, R. Clinton and L. Kruglyak, 2002, "Genetic dissection of transcriptional regulation in budding yeast", *Science*, Vol. 296, pp. 752-755.

Bro, C., 2003, *Transcription analysis of the yeast Saccharomyces cerevisiae*, Ph.D. Dissertation, Technical University of Denmark.

Bro, C., B. Regenberg and J. Nielsen, 2004, "Genome-wide transcriptional response of a *S. cerevisiae* strain with an altered redox metabolism" *Biotechnology & Bioengineering*, Vol. 85, pp. 269-276.

Bronk, J. R., 1999, *Human Metabolism: functional diversity and integration*, Harlow: Addison Wesley Longman, England.

Burgard, A. P., P. Pharkya and C. D. Maranas, 2003, "OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization", *Biotechnology & Bioengineering*, Vol. 84, pp. 647-657.

Camacho, D., A. Fuente and P. Mendes, 2005, "The origin of correlations in metabolomics data", *Metabolomics*, Vol. 1, pp. 53-63.

Chassagnole, C., N. Noisommit-Rizzi, J. W. Schmid, K. Mauch and M. Reuss, 2002, "Dynamic modeling of central carbon metabolism of *Escherichia coli*", *Biotechnology & Bioengineering*, Vol. 79, pp. 53-73.

Christiensen, B. and J. Nielsen, 1999, "Metabolic network analysis: a powerful tool in metabolic engineering", *Advances in Biochemical Engineering & Biotechnology*, Vol. 66, pp. 209-223.

Coleman, S. T., T. K. Fang, S. A. Rovinsky, F. J. Turano and W. S. Moye-Rowley, 2001, "Expression of a glutamate dehydrogenase homologue is required for normal oxidative stress tolerance in *Saccharomyces cerevisiae*", *Journal of Biological Chemistry*, Vol. 276, pp. 244-250.

Cornish-Bowden, A. and M. L. Cardenas, 2002, "Metabolic balance sheets", *Nature*, Vol. 420, pp. 129-130.

Cortassa, S., J. C. Aon and M. A. Aon, 1995, "Fluxes of carbon, phosphorylation, and redox intermediates during growth of *Saccharomcyes cerevisiae* on different carbon sources", *Biotechnology & Bioengineering*, Vol. 47, pp. 193-208.

Covert, M. W., E. M. Knight, J. L. Reed, M. J. Herrgard and B. O. Palsson, 2004, "Integrating high-throughput and computational data elucidates bacterial networks", *Nature*, Vol. 429, pp. 92-96.

Çakır, T., B. Kırdar and K. Ö. Ülgen, 2004. "Metabolic pathway analysis of yeast strengthens the bridge between transcriptomics and metabolic networks", *Biotechnology & Bioengineering*, Vol. 86, pp. 251-260.

Çakır, T., K. R. Patil, Z. İ. Önsan, K. Ö. Ülgen, B. Kırdar and J. Nielsen, 2006, "Integration of metabolome data with metabolic networks reveals reporter reactions". *Molecular Systems Biology*, (accepted)

Daae, E. B. and A. P. Ison, 1999, "Classification and sensitivity analysis of a proposed primary metabolic reaction network for *Streptomyces lividans*", *Metabolic Engineering*, Vol. 1, pp. 153-165.

Dandekar, T., F. Moldenhauer, S. Bulik, H. Bertram and S. Schuster, 2003, “A method for classifying metabolites in topological pathway analyses based on minimization of pathway number”, *BioSystems*, Vol. 70, pp. 255-270.

Daran-Lapujade, P., M. L. Jansen, J. M. Daran, W. van Gulik, J. H. de Winde and J. T. Pronk, 2004, “Role of transcriptional regulation in controlling fluxes in central carbon metabolism of *Saccharomyces cerevisiae*. A chemostat culture study”, *Journal of Biological Chemistry*, Vol. 279, pp. 9125-9138.

DeRisi J. L., R. I. Vishwanath and P. O. Brown, 1997, “Exploring the metabolic and genetic control of gene expression on a genomic scale”, *Science*, Vol. 278, pp. 680-686.

Devantier, R., B. Scheithauer, S. G. Villas-Bôas, S. Pedersen and L. Olsson, 2005a, “Metabolite profiling for analysis of yeast stress response during very high gravity ethanol fermentations”, *Biotechnology & Bioengineering*, Vol. 90, pp. 703-714.

Devantier, R., S. Pedersen and L. Olsson, 2005b, “Transcription analysis of *S. cerevisiae* in VHG fermentation”, *Industrial Biotechnology*, Vol. 1, pp. 51-63.

dos Santos, M. M., G. Thygesen, P. Kötter, L. Olsson and J. Nielsen, 2003, “Aerobic physiology of redox-engineered *Saccharomyces cerevisiae* strains modified in the ammonium assimilation for increased NADPH availability”, *FEMS Yeast Research*, Vol. 4, pp. 59-68.

Duarte, N. C., B. Ø. Palsson and P. Fu, 2004, "Integrated analysis of metabolic phenotypes in *Saccharomyces cerevisiae*," *BMC Genomics*, Vol.5, 63.

Ebenhöh, O. and R. Heinrich, 2003, “Stoichiometric design of metabolic networks: multifunctionality, clusters, optimization, weak and strong robustness”, *Bulletin of Mathematical Biology*, Vol. 65, pp. 323-357.

Edwards, J. S., R. Ramakrishna and B. Ø. Palsson, 2002, "Characterizing the metabolic phenotype: A phenotype phase plane analysis", *Biotechnology & Bioengineering*, Vol. 77, pp. 27-36.

Edwards, J. S., R. Ramakrishna, C. H. Schilling and B. O. Palsson, 1999, "Metabolic flux balance analysis", in S. Y. Lee and E. T. Papoutsakis (eds.), *Metabolic Engineering*, pp. 13-57, Marcel Dekker.

Edwards, J.S. and B.Ø. Palsson, 2000, "The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities", *Proceedings of the National Academy of Sciences*, Vol. 97, pp. 5528-5533.

Edwards, J.S., R. U. Ibarra and B.Ø. Palsson, 2001, "In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data", *Nature Biotechnology*, Vol. 19, pp. 125-130.

Erasmus, D. J., G. K. van der Merwe and H. J. J. van Vuuren, 2003, "Genome-wide expression analyses: metabolic adaptation of *Saccharomyces cerevisiae* to high sugar stress", *FEMS Yeast Research*, Vol. 4, pp. 375-399.

Famili, I., J. Förster, J. Nielsen and B. O. Palsson, 2003, "*Saccharomyces cerevisiae* phenotype can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network", *Proceedings of the National Academy of Sciences*, Vol. 100, pp.13134-13139.

Fell, D., 1997, *Understanding the control of metabolism*, Portland Press, London.

Fell, D.A. and J.R. Small, 1986, "Fat synthesis in adipose tissue. An examination of stoichiometric constraints", *Biochemical Journal*, Vol. 238, pp. 781-786.

Ferea, T. L., D. Botstein, P. O. Brown and R. F. Rosenzweig, 1999, "Systematic changes in gene expression patterns following adaptive evolution in yeast", *Proceedings of the National Academy of Sciences*, Vol. 96, pp. 9721-9726.

Fernie, A. R., R. N. Trethewey, A. J. Krotzky and L. Willmitzer, 2004, "Metabolite profiling: from diagnostics to systems biology", *Nature Reviews Molecular Cell Biology*, Vol. 5, pp. 1-6.

Fiehn, O. and W. Weckwerth, 2003, "Deciphering metabolic networks", *European Journal of Biochemistry*, Vol. 270, pp. 579-588.

Fiehn, O., 2001, "Combining genomics, metabolome analysis, and biochemical modeling to understand metabolic networks", *Computational & Functional Genomics*, Vol. 2, pp. 155-168.

Fischer E., N. Zamboni, and U. Sauer, 2004, "High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived ^{13}C constraints", *Analytical Biochemistry*, Vol. 325, pp.308-316.

Fischer, E. and U. Sauer, 2005, "Large-scale in vivo flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism", *Nature Genetics*, Vol. 37, pp. 636-640.

Flikweert, M. T., L. Van der Zanden, W. M. T. M. Janssen, H. Y. Steensma, J. P. Van Dijken and J. T. Pronk, 1996, "Pyruvate decarboxylase: an indispensable enzyme for growth of *Saccharomyces cerevisiae* on glucose", *Yeast*, Vol. 12, pp. 247-257.

Förster, J., 2003, *Pathway analysis of the metabolic network of Saccharomyces cerevisiae*, Ph.D Thesis, BioCentrum-DTU, Denmark.

Förster, J., A. K. Gombert and J. Nielsen, 2002, "A functional genomics approach using metabolomics and *in silico* pathway analysis", *Biotechnology & Bioengineering*, Vol. 79, pp. 704-712.

Förster, J., I. Famili, P. Fu, B. Ø. Palsson and J. Nielsen, 2003a, "Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network," *Genome Research*, Vol. 13, pp. 244-253.

Förster, J., I. Famili, B. Ø. Palsson and J. Nielsen, 2003b, "Large-scale evaluation of in silico gene deletions in *Saccharomyces cerevisiae*", OMICS, Vol. 7, pp. 193-202.

Fujii, H. and S. Miwa, 2000, "Other erythrocyte enzyme deficiencies associated with non-haematological symptoms: phosphoglycerate kinase and phosphofructokinase deficiency", *Baillieres Best Practical Research in Clinical Haematology*, Vol. 13, pp. 141-148.

Gombert, A. K. and J. Nielsen, 2000, "Mathematical modeling of metabolism", *Current Opinion in Biotechnology*, Vol. 1, pp. 18-186.

Gombert, A. K., M. M. Dos Santos, B. Christensen and J. Nielsen, 2001, "Network identification and flux quantification in the central metabolism of *Saccharomyces cerevisiae* under different conditions of glucose repression", *Journal of Bacteriology*, Vol. 183, pp. 1141-1451.

Goodacre, R., S. Vaidyanathan, W. B. Dunn, G. G. Harrigan and D. B. Kell, 2004, "Metabolomics by numbers: acquiring and understanding global metabolite data", *Trends in Biotechnology*, Vol. 22, pp. 245-252.

Granström, T. B., A. A. Aristidou, J. Jokela and M. Leisola, 2000, "Growth characteristics and metabolic flux analysis of *Candida milleri*", *Biotechnology & Bioengineering*, Vol. 70, pp. 197-207.

Griffin, T. J., S. P. Gygi, T. Ideker, B. Rist, J. Eng, L. Hood and R. Aebersold, 2002, "Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomcyes cerevisiae*", *Molecular & Cellular Proteomics*, Vol. 1, pp. 323-333.

Grotkjær, T., 2005, *Bioinformatics tools in metabolic engineering of Saccharomyces cerevisiae*, Ph. D. Dissertation, Technical University of Denmark.

Halloway, A. J., R. K. van Laar, R. W. Tothill and D. D. L. Bowtell, 2002, "Options available –from start to finish- for obtaining data from DNA microarrays II", *Supplements of Nature Genetics*, Vol. 32, pp. 481-49.

Harrington, C. A., C. Rosenow and J. Retief, 2000, "Monitoring gene expression using DNA microarrays", *Current Opinion in Microbiology*, Vol. 3, pp. 285-291.

Haselbeck, R. J., and L. McAllister-Henn, 1993, "Function and expression of yeast mitochondrial NAD- and NADP-specific isocitrate dehydrogenases", *Journal of Biological Chemistry*, Vol. 268, pp. 12116-12122.

Haurie, V., M. Perrot, T. Mini, P. Jenö, F. Sagliocco and H. Boucherie, 2001, "The transcriptional activator cat8p provides a major contribution to the reprogramming of carbon metabolism during the diauxic shift in *Saccharomyces cerevisiae*", *Journal of Biological Chemistry*, Vol. 276, pp. 76-85.

Heinrich, R. and S. Schuster, 1996, *The regulation of cellular systems*, Chapman & Hall, New York.

Hirai, M. Y., M. Yano, D. B. Goodenowe, S. Kanaya, T. Kimura, M. Awazuhara, M. Arita, T. Fujiwara and K. Saito, 2004, "Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana", *Proceedings of National Academy of Sciences*, Vol. 101, pp. 10205-10210.

Hollan, S., M. Magocsi, E. Fodor, M. Horanyi, V. Harsanyi and T. Farkas, 1997, "Search for the pathogenesis of the differing phenotype in two compound heterozygote Hungarian brothers with the same genotypic triosephosphate isomerase deficiency", *Proceedings of National Academy of Sciences*, Vol. 94, pp. 10362-10366.

Holzhütter, H. G., 2004, "The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks", *European Journal of Biochemistry*, Vol. 271, pp. 2905-2922.

Holzhutter, H.G., G. Jacobasch and A. Bisdorff, 1985, “Mathematical modelling of metabolic pathways affected by an enzyme deficiency. A mathematical model of glycolysis in normal and pyruvate- kinase-deficient red blood cells”, *European Journal of Biochemistry*, Vol. 149, pp. 101-111.

Jacobasch G., 2000, “Biochemical basis of red cell enzyme deficiencies”, *Baillieres Best Practical Research in Clinical Haematology*, Vol. 13, pp. 1-20.

Jacobasch, G. and S. M. Rapoport, 1996, “Hemolytic anemias due to erythrocyte enzyme deficiencies”, *Molecular Aspects in Medicine*, Vol.17, pp. 143-170.

Jansen, M. L. A., J. A. Diderich, M. Mashego, A. Hassane, J. H. de Winde, P. Daran-Lapujade and J. T. Pronk, 2005, “Prolonged selection in aerobic, glucose-limited chemostat cultures of *S. cerevisiae* causes a partial loss of glycolytic capacity”, *Microbiology*, Vol. 151, pp.1657-1669.

Jeong, H. B. Tombor, R. Albert, Z. N. Oltvai and A-L Barabasi, 2000, “The large- scale organization of metabolic networks, *Nature*, Vol. 407, pp. 651-654.

Jorgensen, H. S., J. Nielsen and J. Villadsen, 1995, “Metabolic flux distributions in Penicillium chrysogenum during fed-batch cultivations”, *Biotechnol. Bioeng.*, Vol. 46, pp. 117-131.

Joshi, A. and B. Palsson, 1989, “Metabolic dynamics in human red cell. A comprehensive kinetic model”, *Journal of Theoretical Biology*, Vol. 141, pp. 515-528.

Kaminiski, N. and F. Friedman, 2000, “Practical approaches to analyzing results of microarray experiments”, *American Journal of Respiratory Cell and Molecular Biology*, Vol. 27, pp. 125-132.

Kanno, H., H. Fujii, S. Miwa, 1998, “Expression and enzymatic characterization of human glucose phosphate isomerase (GPI) variants accounting for GPI deficiency”, *Blood Cells, Molecules, and Diseases*, Vol. 24, pp. 54-61.

Kauffman, K. J., P. Prakash and J. E. Edwards, 2004, "Advances in flux balance analysis" *Current Opinion in Biotechnology*, Vol. 14, pp. 491-496.

Kaufman, D. E. and R. L. Smith, 1998, "Direction choice for accelerated convergence in hit-and-run sampling", *Operations Research*, Vol. 46, pp. 84-95.

Kispal, G., C. T. Evans, C. Malloy and P. A. Srere, 1989, "Metabolic studies on citrate synthase mutants of yeast", *Journal of Biological Chemistry*, Vol. 264, pp. 11204-11210.

Kispal, G., M. Rosenkrantz, L. Guarente and P. A. Srere, 1988, "Metabolic changes in *Saccharomyces cerevisiae* strains lacking citrate synthases", *Journal of Biological Chemistry*, Vol. 263, pp. 11145-11149.

Kitano, H., 2002, "Systems biology: a brief overview", *Science*, Vol. 295, pp. 1662-1664.

Klamt, S. and E. D. Gilles, 2004, "Minimal cut sets in biochemical reaction networks", *Bioinformatics*, Vol. 20, pp. 226-234.

Klamt, S. and J. Stelling, 2003a, "Two approaches for metabolic pathway analysis?", *Trends in Biotechnology*, Vol. 21, pp. 64-69.

Klamt, S. and J. Stelling, 2002, "Combinatorial complexity of pathway analysis in metabolic networks", *Molecular Biology Reports*, Vol. 29, pp. 233–236.

Klamt S. and J. Stelling, 2003b, "Stoichiometric analysis of metabolic networks", Tutorial at ICSB 2003, the 4th international conference on systems biology.

Klamt, S., J. Stelling, M. Ginkel, E. D. Gilles, 2003, "Flux Analyzer: Exploring structure, pathways and flux distributions in metabolic networks on interactive flux maps", *Bioinformatics*, Vol. 19, pp. 261-269.

Klipp, E., R. Herwig, A. Kowald, C. Wierling, and H. Lehrach, 2005, *Systems Biology in Practice*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

Kolkman, A., M. M. A. Olsthoorn, C. E. M. Heeremans, A. J. R. Heck and M. Slijper, 2005, "Comparative proteome analysis of *Saccharomyces cerevisiae* grown in chemostat cultures limited for glucose or ethanol", Molecular & Cellular Proteomics, Vol. 4, pp. 1-11.

Krömer, J. O., O. Sorgenfrei, K. Klopprogge, E. Heinzle and C. Wittmann, 2004, "In-depth profiling of lysine-producing *Corynebacterium glutamicum* by combined analysis of the transcriptome, metabolome, and fluxome", *Journal of Bacteriology*, Vol. 186, pp. 1769-1784.

Kümmel, A., S. Panke and M. Heinemann, 2006, "Putative regulatory sites unraveled by network-based embedded thermodynamic analysis of metabolome data", *Molecular Systems Biology*, doi:10.1038/msb4100074.

Kunze, M., F. Kragler, M. Binder, A. Hartig and A. Gurvitz, 2002, "Targeting of malate synthase 1 to the peroxisomes of *Saccharomyces cerevisiae* depends on oleic acid medium", *European Journal of Biochemistry*, Vol. 269, pp. 915-922.

Kurita, O. and Y. Nishida, 1999, "Involvement of mitochondrial aldehyde dehydrogenase ALD5 in maintenance of the mitochondrial electron transport chain in *Saccharomyces cerevisiae*", *FEMS Microbiol Letters*, Vol. 181, pp. 281-287.

Lee, K., F. Berthiaume, G. N. Stephanopoulos, D. M. Yarmush and M. L. Yarmush, 2000, "Metabolic flux analysis of postburn hepatic metabolism", *Metabolic Engineering*, Vol. 2, pp. 312-327.

Lee, S., C. Phalakornkule, M. M. Domach and I. E. Grossmann, 2000, "Recursive MILP model for finding all the alternate optima in LP models for metabolic networks", *Computational and Chemical Engineering*, Vol. 2, pp. 711-716.

Lemke, N., F. Heredia, C. Barcellos, A. dos Reis, and J. C. Mombach, 2004, "Essentiality and damage in metabolic networks" *Bioinformatics*, Vol. 20, pp. 115-119.

Lipshutz, R. J., S. P. A. Fodor, T. R. Gingeras and D. J. Lockhart, 1999, "High density synthetic oligonucleotide arrays", *Supplements Nature Genetics*, Vol. 21, pp. 20-24.

Lockhart, D. J. and E. A. Winzeler, 2000, "Genomics, gene expression and DNA arrays", *Nature*, Vol. 405, pp. 827-836.

Loftus, T. M., L. V. Hall, S. L. Anderson and L. McAllister-Henn, 1994, "Isolation, characterization, and disruption of the yeast gene encoding cytosolic NADP-specific isocitrate dehydrogenase", *Biochemistry*, Vol. 33, pp. 9661-9667.

Ma, H. and A. P., 2003, "Reconstruction of metabolic networks from enom data and analysis of their global structure for various organisms", *Bioinformatics*, Vol. 19, pp. 270-277.

Mahadevan, R. and C. H. Schilling, 2000, "The effects of alternate optimal solutions in constraint-based genome-scale metabolic models", *Metabolic Engineering*, Vol. 5, pp. 264-26.

Martinov, M.V., A.G. Plotnikov, V. M. Vitvitsky and F. I. Ataullakhanov, 2000, "Deficiencies of glycolytic enzymes as a possible cause of hemolytic anemia", *Biochimica Biophysica Acta*, Vol. 1474, pp. 75-87.

McCammon, M. T. , C. B. Epstein, B. Przybyla-Zawislak, L. McAlister-Henn and R. A. Butow, 2003, "Global transcription analysis of krebs tricarboxylic acid cycle mutants reveals an alternating pattern of gene expression and effects on hypoxic and oxidative genes", *Molecular Biology of the Cell*, Vol. 14, pp. 958-972.

Meaden, P. G., F. M. Dickinson, A. Mifsud, W. Tessier, J. Westwater, H. Bussey and M. Midgley, 1997, "The ALD6 gene of *Saccharomyces cerevisiae* encodes a cytosolic, Mg^{2+} activated acetaldehyde dehydrogenase", *Yeast*, Vol. 13, pp. 1319-1327.

Mehta, A., P. J. Mason, and T. J. Vulliamy, T, 2000, "Glucose-6-phosphate dehydrogenase deficiency", *Baillieres Best Practical Research in Clinical Haematology*, Vol. 13, pp. 21-38.

Mendes, P., 2002, "Emerging bioinformatics for metabolome", *Briefings in Bioinformatics*, Vol. 3, pp. 134-145.

Messana, I., M. Orlando, L. Cassiano, L. Pennacchietti, C. Zuppi, M. Castagnola, and B. Giardina, 1996, "Human erythrocyte metabolism is modulated by O2-linked transition of hemoglobin", *FEBS Letters*, Vol. 390, pp. 25-28.

Minard, K. I., G. T. Jennings, T. M. Loftus, D. Xuan and L. McAllister-Henn, 1998, "Sources of NADP and expression of mammalian $NADP^+$ specific isocitrate dehydrogenases in *Saccharomyces cerevisiae*", *Journal of Biological Chemistry*, Vol. 273, pp. 31486-31493.

Miwa, S., 1996, "Red Cell Enzymopathies", Education Programme CD-ROM of the 26th Congress of the International Society of Haematology.

Momsen, G. and B. Vestergaard-Bogind, 1978, "Human erythrocyte 2,3-diphospoglycerate metabolism. Influence of 1,3-diphosphoglycerate and Pi", *Archives in Biochemistry and Biophysics*, Vol. 190, pp. 67-87.

Mulquiney, P. J. and P. W. Kuchel, 1999, "Model of 2,3-biphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: equations and parameter refinement", *Biochemical Journal*, Vol. 342, pp. 581-596.

Nielsen, J. and S. Oliver, 2005, "The next wave in metabolome analysis", *Trends in Biotechnology*, Vol. 23, pp. 544-546.

Nielsen, J., 2003, "It is all about metabolic fluxes", *Journal of Bacteriology*, Vol. 185, pp. 7031-7035.

Nielsen, J., 2001, "Metabolic engineering", *Appl. Microbiol. Biotechnol.*, Vol. 55, pp. 263-283.

Nissen, T. L., M. C. Kielland-Brandt, J. Nielsen and J. Villadsen, 2000, "Optimization of ethanol production in *Saccharomyces cerevisiae* by metabolic engineering of the ammonium assimilation", *Metabolic Engineering*, Vol. 2, pp. 69-77.

Nissen, T. L., U. Schulze, J. Nielsen and J. Villadsen, 1997, "Flux distributions in anaerobic, glucose-limited continuous cultures of *Saccharomyces cerevisiae*", *Microbiology*, Vol. 143, pp. 203-218.

Oh, M. K. and J. C. Liao, 2000, "Gene expression profiling by DNA microarrays and metabolic fluxes in *Escherichia coli*", *Biotechnology Progress*, Vol. 16, pp. 278-286.

Oh, M. K., L. Rohlin, K. C. Kao and J. C. Liao, 2002, "Global expression profiling of acetate-grown *Escherichia coli*", *Journal of Biological Chemistry*, Vol. 277, pp. 13175-13183.

Ohlmeier, S., A. J. Kastaniotis, J. K. Hiltunen and U. Bergmann, 2004, "The yeast mitochondrial proteome, a study of fermentative and respiratory growth", *Journal of Biological Chemistry*, Vol. 279 pp. 3956-79.

Olah, J., F. Orosz, G. M. Keserü, Z. Kovari, J. Kovacs, S. Hollan and J. Ovadi, 2002, "Triosephosphate isomerase deficiency: a neurodegenerative misfolding disease", *Biochemical Society Transactions*, Vol. 30, pp. 30-38.

Orosz, F., B.G. Vertessy, S. Hollan, M. Horany and J. Ovadi, 1996, "Triosephosphate isomerase deficiency: predictions and facts", *Journal of Theoretical Biology*, Vol. 182, pp. 437-447.

Ostergaard, S., Olsson, L. and J. Nielsen, 2000, "Metabolic engineering of *Saccharomyces cerevisiae*", *Microbiology and Molecular Biology Reviews*, Vol. 64, pp. 34-50.

Overkamp, K. M., B. M. Bakker, P. Kötter, A. van Tuijl, S. De Vries, J. P. van Dijken and J. T. Pronk, 2000, "In vivo analysis of the mechanisms for oxidation of cytosolic NADH by *Saccharomyces cerevisiae* mitochondria", *Journal of Bacteriology*, Vol. 182, pp. 2823-283.

Palsson, B. O., N. D. Price and J. A. Papin, 2003, "Development of network-based pathway definitions. The need to analyze real metabolic networks", *Trends in Biotechnology*, Vol. 21, pp.195-198.

Pandolfi, P. P., F. Sonati, R. Rivi, P. Mason, F. Grosveld and L. Luzzatto, 1995, "Targeted disruption of the housekeeping gene encoding glucose 6-phosphate dehydrogenase (G6PD): G6PD is dispensable for pentose synthesis but essential for defense against oxidative stress", *EMBO Journal*, Vol. 14, pp. 5209-5215.

Papin, J. A., N. D. Price, S. W. Wiback, D. A. Fell and B. O. Palsson, 2003, "Metabolic pathways in the post-genome era", *Trends in Biochemical Sciences*, Vol. 28**,** pp. 250-258.

Papp, B., C. Pal, and L. D. Hurst, 2004, "Metabolic network analysis of the causes and elution of enzyme dispensability in yeast", *Nature*, Vol. 429, pp. 661–664.

Patil, K. R., I. Rocha, J. Förster and J. Nielsen, 2005, "Evolutionary programming as a platform for in silico metabolic engineering", *BMC Bioinformatics*, Vol. 6, pp. 308.

Patil, K. R. and J. Nielsen, 2005, "Uncovering transcriptional regulation of metabolism using metabolic network topology", *Proceedings of National Academy of Sciences*, Vol. 102, pp. 2685-2689.

Pfeiffer, T., I. Sanchez-Valdenebro, J.C. Nuno, F. Montero and S. Schuster, 1999, "METATOOL: for studying metabolic networks", *Bioinformatics*, Vol. 15, pp. 251-257.

Phalakornkule, C., S. Lee, T. Zhu, R. Koepsel, M. Ataai, I. Grossmann and M.. Domach, 2001, "A MILP based flux alernative generation and NMR experimental design strategy for metabolic engineering", *Metabolic Engineering*, Vol. 3, pp. 124-137.

Piper, M. D. W., P. Daran-Lapujade, C. Bro, B. Regenberg, S. Knudsen, J. Nielsen and J. T. Pronk, 2002, "Reproducibility of oligonucleotide microarray transcriptome analyses", *Journal of Biological Chemistry*, Vol. 277, pp. 37001-37008.

Pramanik, J. and J. D. Keasling, 1997, "Stoichiometric model of *Escherichia coli* metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements", *Biotechnology and Bioengineering*, Vol. 56, pp. 398-421.

Price, N. D., J. L. Reed, J. A. Papin, S. J. Wiback and B. O. Palsson, 2003, "Network based analysis of metabolic regulation in the human red blood cell", *Journal of Theoretical Biology*, Vol. 225, pp. 185-194.

Price, N. D., J. Schallenberger and B. O. Palsson, 2004a, "Uniform sampling of steady state flux spaces: means to design experiments and interpret enzymopathies", *Biophysical Journal*, Vol. 88, pp.2172-2186.

Price, N. D., J. L Reed, B. O. Palsson, 2004b, "Genome-scale models of microbial cells: evaluating the consequences of constraints", *Nature Reviews Microbiology*, Vol. 2, pp. 886-897.

Prokisch, H., C. Scharfe, D. G. Camp, W. Xiao, L. David, C. Andreoli, M. E. Monroe, R. J. Moore, M. A. Gritsenko, C. Kozany, K. K. Hixson, H. M. Mottaz, H. Zischka, M. Ueffing, Z. S. Herman, R. W. Davis, T. Meitinger, P. J. Oefner, R. D. Smith and L. M. Steinmetz, 2004, "Integrative analysis of the mitochondrial proteome in yeast", *PLoS Biology*, Vol. 2, pp. e160.

Raamsdonk, L. M., B. Teusink, D. Broadhurst, N. Zhang, A. Hayes, M. C. Walsh, J. A. Berden, K. M. Brindle, D. B. Kell, J. J. Rowland, H. V. Westerhoff, K. van Dam and

S. G. Oliver, 2001, “A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations”, *Nature Biotechnology*, Vol. 19, pp. 45–50.

Recchia, F. A., T. R. Vogel and T. H. Hintze, 2000, “NO metabolites accumulate in erythrocytes in proportion to carbon dioxide and bicarbonate concentration”, *American Journal of Physiolology of Heart*, Vol. 279, pp. H852-H856.

Reed, J. L., T. D. Vo, C. H. Schilling, and B. Ø. Palsson, 2003, “An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR)”, *Genome Biology*, Vol. 4: pp. R54.1-R54.12.

Repiso, A., J. Boren, F. Ortega, A. Pujades, J. Centelles, J. L. Vives-Corrons, F. Climent, M. Cascante and J. Carreras, 2002, “Triosephosphate isomerase deficiency: Genetic, enzymatic, and metabolic characterization of a new case from Spain”, *Haematologica*, Vol. 87, e-cases.

Rizzi, M., M. Baltes, U. Theobald and M Reuss, 1997, “In vivo analysis of metabolic dynamics in *Saccharomyces cerevisiae*.: II. Metabolic model”, *Biotechnology Bioengineering*, Vol. 55, pp. 592-608.

Rossel, S., C. C. van der Weijden, A. L. Kruckeberg, B. M. Bakker and H. V. Westerhoff, 2005, “Hierarchical and metabolic regulation of glucose influx in starved *Saccharomyces cerevisiae*”, *FEMS Yeast Research*, Vol. 5, pp. 611-619.

Roth, S. and H. J. Schüller, 2001, “Cat8 and Sip4 mediate regulated transcriptional activation of the yeast malate dehydrogenase gene MDH2 by three carbon source-responsive promoter elements”, *Yeast*, Vol. 18, pp. 151-162.

Sanford K., P. Soucaille, G. Whited and G. Chotani, 2002, “Genomics to fluxomics and physiomics- pathway engineering”, *Current Opininon in Microbiology*, Vol. 5, pp. 318-322.

Sauer, U., 2004, “High-throughput phenomics: experimental methods for mapping fluxomes”, *Current Opinion in Biotechnology*, Vol. 15, pp. 58-63.

Schilling, C. H. and B. O. Palsson, 1998. “The underlying structure of biochemical reaction networks”, *Proceedings of National Academy of Sciences*, Vol. 95, pp. 4193-4198.

Schilling, C. H., S. Schuster, B. O. Palsson and R. Heinrich, 1999, “Metabolic pathway analysis: Basic concepts and scientific applications in the post-genomic era”, *Biotechnology Progress*, Vol. 15, pp. 296-303.

Schneider, A. S., 2000, “Triosephosphate isomerase deficiency: historical perspectives and molecular aspects”, *Baillieres Best Practical Research in Clinical Haematology*, Vol. 13, pp. 119-140.

Schuster, R. And H. G. Holzhütter, 1995, “Use of mathematical models for predicting the metabolic effect of large-scale enzyme activity alterations: Application to enzyme deficiencies of red blood cells”, *European Journal of Biochemistry*, Vol. 229, pp. 403-418.

Schuster, R., G. Jacobasch and H. G. Holzhütter, 1989, “Mathematical modelling of metabolic pathways affected by an enzyme deficiency. Energy and redox metabolism of glucose-6-phosphate-dehydrogenase-deficient erythrocytes”, European Journal of Biochemistry, Vol. 182, pp. 605-612.

Schuster, S., T. Dandekar, D. A. Fell, 1999, “Detection of elementary flux modes in biochemical networks: A promising tool for pathway analysis and metabolic engineering”, *Trends in Biotechnology*, Vol. 17, pp. 53-60.

Schuster, S., D. A. Fell and T. Dandekar, 2000, “A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks”, *Nature Biotechnology*, Vol. 18, pp. 326-332.

Schuster, S., D. A. Fell, T. Pfeiffer, T. Dandekar and P. Bork, 1998, “Elementary modes analysis illustrated with human red cell metabolism”, in C. Larsson, I. L. Pahlman and L. Gustafsson (eds.), *BioThermoKinetics in the Post Genomic Era*, pp. 332-339, Chalmers, Göteborg.

Schuster, S., C. Hilgetag, J. H. Woods and D. A. Fell, 2002a, “Reaction routes in biochemical reaction systems: Algebraic properties, validated calculation procedure and example from nucleotide metabolism”, *Journal Mathematical Biology*, Vol. 45, pp. 153-181.

Schuster, S., S. Klamt, W. Weckwerth, F. Moldenhauer and T. Pfeiffer, 2002b, “Use of network analysis of metabolic systems in bioengineering”, *Bioprocess and Biosystems Engineering*, Vol. 24, pp. 363-372.

Schuster, S., T. Pfeiffer, F. Moldenhauer, I.. Koch and T. Dandekar, 2002c, “Exploring the pathway structure of metabolism: decomposition into subnetworks and application to *Mycoplasma pneumoniae*”, *Bioinformatics*, Vol. 18, pp. 351-361.

Schwarz, R., P. Musch, A. von Kamp, B. Engels, H. Schirmer, S. Schuster and T. Dandekar, 2005, “YANA - a software tool for analyzing flux modes, gene-expression and enzyme activities”, *BMC Bioinformatics*, Vol. 6, 135.

Sheikh, K., J. Forster L. K. Nielsen, 2005, “Modeling hybridoma cell metabolism using a generic genome-scale metabolic model of *Mus musculus*”, *Biotechnology Progress*, Vol. 21, pp. 112-21.

Shi, H., J. Nkawa and K. Shimizu, 1999, “Effect of modifying metabolic network on poly-3-hydroxybutyrate biosynthesis in recombinant *Escherichia coli*”, *Journal of Bioscience and Bioengineering*, Vol. 87, pp. 666-677.

Smith, R. L., 1984, “Efficient monte carlo procedures for generating points uniformly distributed over bounded regions”, *Operations Research*, Vol. 32, pp. 1296-1308.

Stafford, D. E. and G. Stephanopoulos, 2001, "Metabolic engineering as an integrating platform for strain development", *Current Opinion in Microbiology*, Vol. 4, pp. 336-340.

Stein, S. E., 1999, "An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data", *Journal of American Society for Mass Spectrometry*, Vol. 10, pp. 770-781.

Stelling, J., S. Klamt, K. Bettenbrock, S. Schuster and E. D. Gilles, 2002, "Metabolic network structure determines key aspects of functionality and regulation", *Nature*, Vol. 420, pp. 190-193.

Stephanopoulos, G., A. A. Aristidou and J. Nielsen, 1998, *Metabolic Engineering-principles and methodologies*, Academic Press, San Diego.

Stückrath, I., H. C. Lange, P. Kötter, W. M. van Gulik, K. D. Entian and J. J. Heijnen, 2002, "Characterization of null mutants of the glyoxylate cycle and gluconeogenic enzymes in *S. cerevisiae* through metabolic network modeling verified by chemostat cultivation", *Biotechnology & Bioengineering*, Vol. 77, pp. 61-72.

Sumner, L. W., P. Mendes and R. A. Dixon, 2003, "Plant metabolomics: large-scale phytochemistry in the functional genomics era", *Phytochemistry*, Vol. 62, pp. 817-836.

Taylor, K. M., C. P. Kaplan, X. Gao and A. Baker, 1996, "Localization and targeting of isocitrate lyases in *Saccharomyces cerevisiae*", *Biochemical Journal*, Vol. 319, pp. 255-262.

ter Kuile, B. H., and H. V. Westerhoff, 2001, "Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway", *FEBS Letters*, Vol. 500, pp. 169-171.

ter Linde, J. J. M., H. Liang, R. W. Davis, H. Y. Steensma, J. P. van Dijken and J. T. Pronk, 1999, "Genome-wide transcriptional analysis of aerobic and anaerobic chemostat cultures of *Saccharomyces cerevisiae*", *Journal of Bacteriology*, Vol. 181, pp.7409-7413.

Tessier, W. D., P. G. Meaden, F. M. Dickinson and M. Midgley, 1998, "Identification and disruption of the gene encoding the K^+ activated acetaldehyde dehydrogenase of *Saccharomyces cerevisiae*", *FEMS Microbiology Letters*, Vol. 164, pp. 29-34.

Teusink, B., J. Passarge, C. A. Reijenga, E. Esgalhado, C. C. van der Weijden, M. Schepper, M. C. Walsh, B. M. Bakker, K. van Dam, H. W. Westerhoff and J. L. Snoep, 2000, "Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry", *European Journal of Biochemistry*, Vol. 267, pp. 5313–5329.

Thiele I., N. D. Price, T. D. Vo and B. O. Palsson, 2005, "Candidate metabolic network states in human mitochondria" *Journal of Biological Chemistry*, Vol. 280, pp. 11683-11695.

Townsend J. P., D. Cavalieri and D. L. Hartl, 2003, "Population genetic variation in genome-wide gene expression", *Molecular Biology and Evolution*, Vol. 20, pp. 955-963.

Tusher, V. G., R. Tibshirani, and G. Chu, 2001, "Significance analysis of microarrays applied to the ionizing radiation response", *Proceedings of National Academy of Sciences*, Vol. 98, pp. 5116–5121.

Urbanczik, R. and C. Wagner, 2005, "Functional stoichiometric analysis of metabolic networks", *Bioinformatics*, Vol. 21, pp. 4176-4180.

van den Berg, M. A. and H. Y. Steensma, 1995, "ACS2, a *Saccharomyces cerevisiae* gene encoding acetyl-coenzyme A synthetase, essential for growth on glucose", *European Journal of Biochemistry*, Vol. 231, pp. 704-713.

van Gulik, W. M. And J. J. Heijnen, 1995, “A metabolic network stoichiometry analysis of microbial growth and product formation”, *Biotechnology & Bioengineering*, Vol. 48, pp. 681-698.

Vanrolleghem, P. A., P. de Jong-Gubbels, W. M. van Gulik, J. T. Pronk J. P. van Dijken and S. Heijnen, 1996, “Validation of a metabolic network for *Saccharomcyes cerevisiae* using mixed substrate studies”, *Biotechnology Progress*, Vol. 12, pp. 434-448.

Varela, C., J. Cárdenas, F. Melo and E. Agosin, “Quantitative analysis of wine yeast gene expression profiles under winemaking conditions”, 2005, Vol. 22, pp. 369-383.

Varma, A. and B. O. Palsson, 1994, “Metabolic flux balancing: basic concepts, scientific and pratctical use” *Bio/technology*, Vol. 12, pp. 994-998.

Varner, J. and D. Ramkrishna, 1999, “Mathematical models of metabolic pathways” *Current Opinion in Biotechnology*, Vol. 10, pp.146-150.

Villas-Bôas, S. G., J. F. Moxley, M. Åkesson, G. Stephanopoulos G and J. Nielsen, 2005a, “High-throughput metabolic state analysis: the missing link in integrated functional genomics”, *Biochemical Journal*, Vol. 388, pp. 669-677.

Villas-Bôas, S. G., M. Åkesson,and J. Nielsen, 2005b, “Biosynthesis of glyoxylate from glycine in *Saccharomyces cerevisiae*”, *FEMS Yeast Research*, Vol. 5, pp. 703-709.

Villas-Bôas, S. G., S. Mas, M. Åkesson, J. Smedsgaard and J. Nielsen, 2005c, “Mass spectrometry in metabolome analysis”, *Mass Spectrometry Reviews*, Vol. 24, pp. 613-646.

Wagner A., 2000., “Robustness against mutations in genetic networks of yeast”, *Nature Genetics*, Vol. 24, pp. 355-361.

Wagner A. and D. A. Fell, 2001, "The small world inside large metabolic networks", *Proceedings of the Royal Society of London B.*, Vol. 268, pp. 1803-1810.

Weatherall, D. J. and A. B. Provan, 2000, "Red cells I: inherited aneamias", *Lancet*, Vol. 355, pp. 1169-1175.

Weckwerth, W., M. E. Loureiro, K. Wenzel, and O. Fiehn, 2004, "Differential metabolic networks unravel the effects of silent plant phenotypes", *Proceedings of National Academy of Sciences*, Vol. 101, pp.7809-7814.

Wiback, S. J. and B. O. Palsson, 2002, "Extreme pathway analysis of human red blood cell metabolism", *Biophysical Journal*, Vol. 83, pp. 808-818.

Wiechert, W., 2002, "Modeling and simulation: tools for metabolic engineering", *Journal of Biotechnology*, Vol. 94, pp. 37-63.

Williams RM, Primig M, Washburn BK, Winzeler EA, Bellis M, Sarrauste de Menthiere C., R. W. Davis and R. E. Esposito, 2002, "The Ume6 regulon coordinates metabolic and meiotic gene expression in yeast", *Proceedings of National Academy of Sciences*, Vol. 99, pp. 13431-13436.

Zabinsky, Z. B., 2003, *Stochastic adaptive search for global optimization*, Kluwyer Academic Publishers, USA

Zamboni, N. and U. Sauer, 2004, "Model-independent fluxome profiling from 2H and 13C experiments for metabolic variant discrimination", *Genome Biology*, Vol. 5, R99.

Zhu, T., C. Phalakornkule, S. Ghosh, I. E. Grosmann, R. R. Koepsel, M. M. Ataai and M. M. Domach, 2003, "A metabolic network analysis & NMR experiment design tool with user interface-driven model construction for depth-first search analysis", *Metabolic Engineering*, Vol. 5, pp.74-85.