Doctoral Thesis

Integrating thermodynamics-based modeling and quantitative experimental data for studying microbial metabolism

Author(s):
Kümmel, Anne Maria Rita

Publication Date:
2008

Permanent Link:
https://doi.org/10.3929/ethz-a-005713687

Rights / License:
In Copyright - Non-Commercial Use Permitted
Integrating Thermodynamics-based Modeling and Quantitative Experimental Data for Studying Microbial Metabolism

A dissertation submitted to
ETH ZURICH
for the degree of
Doctor of Sciences
presented by
Anne Maria Rita Kümmel
Dipl.-Ing. RWTH Aachen, Germany
born July, 18th, 1977
accepted on the recommendation of
Prof. Dr. Sven Panke
Prof. Dr. Uwe Sauer
Dr. Matthias Heinemann
2008
Preface

The PhD thesis in hand is built on work I did during my time at the Institute of Process Engineering and the Institute of Molecular Systems Biology at the ETH Zurich. Coming from a background of mechanical engineering, I enjoyed entering biology and learning about biological experimentation. Many people accompanied my way whom I would like to thank for their support.

I am grateful to ...

... Sven Panke for staying my advisor even after I left his lab

... Uwe Sauer for his scientific guidance but even more for his personal support.

... my colleagues in both the Panke as well as the Sauer lab for a welcoming and joyful working environment. Especially Jenny, Sarah, Roeland, Simon, and Dominik for helping me in the lab.

... Matthias who has accompanied me as a friend and a scientific supervisor, both during my diploma and my PhD thesis. He never became tired of reminding me that the glass rather was half full ...

... meinen Eltern für ihre selbstverständliche und bedingungslose Unterstützung, meinen Talenten und Interessen nachzugehen.

... Sandra for backing me when I was exhausted and for her patience and help when finishing this thesis.

Zurich, August 2008

Anne Kümmel
Vem kan förstå det stora i det lilla...

J. Sandén, swedish gospel
Contents

Summary ix
Zusammenfassung xiii

1. Introduction 1
   1.1. Background . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1
   1.1.1. Systems Biology . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1
   1.1.2. Systems Biology of Metabolism . . . . . . . . . . . . . . . . . . . 3
   1.2. Outline and Positioning of this Thesis . . . . . . . . . . . . . . . . . . 12

2. Systematic Reaction Direction Assignment 15
   2.1. Background . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
   2.2. Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19
      2.2.1. Thermodynamic Facts-based Assignment . . . . . . . . . . . . . 19
      2.2.2. Thermodynamic Heuristics-based Assignment . . . . . . . . . . 22
   2.3. Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 29
   2.4. Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34
   2.5. Methods . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 35

3. Network-embedded Thermodynamic Analysis 39
   3.1. Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 40
   3.2. Network-embedded Thermodynamic Analysis . . . . . . . . . . . . . . 42
   3.3. Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 46
      3.3.1. Thermodynamic Consistency of Metabolome Data . . . . . . . . . 48
## Contents

3.3.2. Prediction of Metabolite Concentrations ........................................ 49  
3.3.3. Prediction of Potential Active Regulatory Sites ............................... 51  
3.3.4. General Applicability of NET Analysis ........................................ 53  
3.4. Discussion ......................................................................................... 55  
3.5. Materials and Methods ...................................................................... 58  

4. Metabolomics-based Elucidation of Metabolic Activities .................. 63  

4.1. Introduction ....................................................................................... 64  
4.2. Approach to Predict Active and Inactive Reactions ......................... 66  
4.3. Results ............................................................................................... 69  
4.3.1. Experimental Data and Metabolic Model Applied As Inputs .......... 69  
4.3.2. Inferred Active Reactions due to Mass Balances ......................... 71  
4.3.3. Inferred Inactive/Active Reactions and Restricted Directions from  
       Thermodynamically Feasible EFMs ................................................... 71  
4.3.4. Validation of Inferred Activities and Direction Restrictions Obtained  
       from Thermodynamic Constraints .................................................. 72  
4.3.5. Evaluation of Additional Constraints for Growth on Glucose .......... 75  
4.4. Discussion ......................................................................................... 76  
4.5. Methods ............................................................................................ 77  
4.6. Appendix .......................................................................................... 82  

5. Experimental Approach to Quantitative Strain Comparison ............. 87  

5.1. Introduction ....................................................................................... 88  
5.2. Results .............................................................................................. 90  
5.2.1. Differences Between CEN.PK and FY4 ...................................... 90  
5.2.2. Potentially Different Transcription Factor Activities .................. 98  
5.2.3. Different Effects of Genetic Perturbations in CEN.PK and FY4 .... 99  
5.3. Discussion ........................................................................................ 102  
5.4. Materials and Methods ................................................................... 106
6. Concluding Remarks

6.1. Summary .............................................................. 111

6.2. Discussion ............................................................ 112
   6.2.1. Large-scale Data for Gaining a Holistic Perspective ...... 112
   6.2.2. Data Analysis for Inferring New Insights and Knowledge 113
   6.2.3. Model-based Knowledge Integration for Achieving System-level Understanding .............................................. 114

A. Appendix ............................................................... 117

   A.1. Experimental Analysis of the Diauxic Shift in S. cerevisiae ...... 117
      A.1.1. Introduction ..................................................... 117
      A.1.2. Experimental Results ......................................... 118
      A.1.3. Outlook ......................................................... 118
      A.1.4. Materials and Methods ........................................ 126

References ............................................................... 131

List of Publications .................................................... 149

Curriculum Vitae ....................................................... 151
Summary

Systems biology strives to understand how the behavior of living systems is emerging from the underlying diverse and complex molecular interactions. It builds (i) on large-scale data sets that harbor preferably the quantitative amounts of all interacting molecules, (ii) on computational tools for the analysis of such data sets, and (iii) on mathematical models to finally obtain system-level understanding. Nowadays, advances in high-throughput experimental omics techniques drives systems biology forward through providing a system-level perspective on biological networks. However, due to the large size of the generated data sets, they can hardly be understood or interpreted directly. In order to benefit from the information that is contained in these data sets, computational data analysis tools are essential. Here, models are valuable tools to enhance the computational data analysis next to their essential role in describing the system behavior to gain system-level understanding. Consequently, with the advances in experimental technologies the importance of computational tools to interpret these data and models as semantic tools to administer the gained knowledge increases.

Centered on microbial metabolism, the different chapters of this thesis cover on the one hand models and model-based data analyses and on the other hand the application of advanced experimental technologies. On the modeling side, computational tools (i) for assisting the reconstruction of genome-scale stoichiometric models of the metabolic network, (ii) for analyzing quantitative metabolomics data, and (iii) for inferring metabolic activities from measured metabolite concentrations were developed. On the experimental side, state-of-the-art quantitative omics data sets were generated and exploited (iv) to compare the metabolic state of two *S. cerevisiae* strains in detail and (v) to monitor
Summary

the metabolic reprogramming of \textit{S. cerevisiae} undergoing a substrate switch from glucose to ethanol.

Chapter 2 focuses on the reconstruction of large-scale models of metabolic networks. The genome-scale stoichiometric network models have particularly proven to be a valuable tool in systems biology. They are composed of the metabolic reactions’ stoichiometry and assignments of the reactions’ reversibility or irreversibility and thus define an organism’s metabolic capacity to convert substrates into biomass, energy, and by-products. In the reconstruction process, a first version of the model typically comprises a list of metabolic enzymes that are identified from the annotated genome. Whereas significant efforts have been put into the development of computational tools for identifying and compiling such organism-specific lists of metabolic reactions, the definition of the thermodynamic constraints in terms of reaction directionality is a tedious manual process. For an automated and systematic assignment of thermodynamically reasonable reaction direction restrictions, an algorithm based on thermodynamics, network topology, and heuristic rules was developed. It assigns reaction directions in metabolic models such that the reaction network is thermodynamically feasible with respect to the production of energy equivalents. The applicability of the algorithm was demonstrated on a genome-scale metabolic model of \textit{Escherichia coli}. Although not being fully comprehensive, the presented algorithm could define a significant number of irreversible reactions automatically with low computational effort. It can be a valuable part of a computational framework that assists the automated reconstruction of genome-scale metabolic models.

In Chapter 3, next to such stoichiometric metabolic network models thermodynamic constraints were employed to analyze large-scale quantitative metabolomics data: Here, a framework for mechanistic and model-based analysis of these data - the network-embedded thermodynamic analysis (NET analysis) - is introduced. By exploiting the second law of thermodynamics and the metabolites’ Gibbs energies of formation, NET analysis allows for inferring functional principles and identifies reactions that are most likely subject to active allosteric or genetic regulation as exemplified with quantitative
metabolome data from *E. coli* and *Saccharomyces cerevisiae*. Moreover, the optimization framework of NET analysis demonstrated to be a valuable tool to systematically investigate data sets for consistency, for the extension of sub-omic metabolome data sets, and for resolving intracompartamental concentrations from cell-averaged metabolome data. Without requiring any kind of kinetic modeling, NET analysis represents a scalable and unbiased approach to uncover insights from quantitative metabolome data.

Elucidating intracellular metabolic fluxes is crucial for gaining systems understanding of cellular processes. The standard technique to assess these fluxes experimentally, namely $^{13}$C-based flux analysis, however, has certain experimental limitations. Having a broader experimental applicability, in Chapter 4, a novel metabolomics-based approach for inferring metabolic activities based on mass balances and thermodynamic constraints was developed. Here, by combining NET analysis with elementary flux mode (EFM) analysis, the feasible fluxes defined by both, mass balances and thermodynamic constraints, are such comprehensively analyzed: All EFMs obtained from a stoichiometric metabolic model are tested via NET analysis for their thermodynamic feasibility with the measured metabolite concentrations. To demonstrate the approach, reaction activities within the pyruvate/ethanol metabolism of either glucose- or ethanol-grown *S. cerevisiae* from measured metabolite concentrations were inferred.

In Chapter 5, the glucose repression state of *S. cerevisiae* CEN.PK 113-D7 and FY4 was compared by quantitative, large-scale proteomics and metabolomics data sets as well as metabolic fluxes. The wild-types display similar degree of glucose repression on the flux level since the biomass and ethanol production yields as well as the respiratory TCA cycle activities are comparable. Nevertheless, differences in protein and metabolite concentrations were observed which indicated that Hxk2-dependent signaling and regulation is stronger in CEN.PK. Differences in signaling and regulatory strengths became also evident as the glucose repression state on flux level - indicated by the respiratory TCA cycle activity - for a genetic perturbation differs significantly: While the glucose repression was maintained in FY4, CEN.PK switched to a respiratory metabolism upon a $\Delta hxk2$ deletion. It was demonstrated that a point mutation in the *CYR1* gene is
causing this distinct behavior.

Beyond these projects (as described in the appendix), omics techniques were applied to monitor metabolic changes in *S. cerevisiae* during a diauxic shift. All phases of this metabolic reprogramming were characterized by quantitative metabolomics and proteomics data next to the extracellular physiology. Some of the acquired data were already exploited for the development of the NET/EFM analysis (cf. Chapter 4). The complete dynamic data set represents a basis for future analysis to elucidate the temporal and causal series of regulatory events acting in response to the altered substrate availability.

Today, while the technology to generate quantitative omics data sets develops quickly, the development of generally applicable model-based analysis concepts for such data lags behind. Therefore, the here developed and also applied model-based approaches for analysis of large-scale data contribute to the field of computational systems biology. As a case study, *S. cerevisiae*’s metabolism was experimentally analyzed by quantitative and large-scale data sets.
Zusammenfassung


Der mikrobielle Stoffwechsel steht im Mittelpunkt dieser Arbeit, die sich auf der einen Seite mit Modellen und modellbasierten Datenanalyse und auf der anderen Seite mit dem Einsatz moderner experimenteller Analytik befasst. Im Bereich der Modellierung wurden rechnergestützte Methoden entwickelt, um (i) die Modellbildung genomweiter stochiometrischer Modelle des metabolischen Netzwerks, um (ii) quantitative Metabolom-Datensätze zu analysieren, und um (iii) von gemessenen Metabolitkonzentrationen auf metabolische Aktivitäten zu schliessen. Im experimentellen Teil wurden quantitative
Zusammenfassung

Omik-Datensätze erzeugt und dazu genutzt, den Metabolismus zweier S. cerevisiae-Stämme detailliert zu vergleichen, und die Umstellung des Metabolismus vom Substrat Glukose zu Ethanol zu verfolgen.


Neben diesen stöchiometrischen Modellen des metabolischen Netzwerks wurden in Kapitel 3 thermodynamische Restriktionen zur Analyse grosser, quantitativer Metabolomik-Datensätze benutzt. Dieses Kapitel beschreibt eine Methode zur mechanistischen, modellbasierten Analyse dieser Daten, die thermodynamische Netzwerkanalyse (network-embedded thermodynamic analysis, kurz NET-Analyse). Unter Berücksichti-
Zusammenfassung

gung des zweiten Hauptsatzes der Thermodynamik und der Gibbs'schen Formationsenergien der Metabolite leitet die NET-Analyse funktionelle Zusammenhänge ab und identifiziert Reaktionen, auf die vermutlich aktive allosterische oder genetische Regulation wirkt. Dies wurde beispielhaft an zwei quantitativen Metabolomik-Datensätzen von *E. coli* und *Saccharomyces cerevisiae* gezeigt. Darüberhinaus ist die NET-Analyse ein wertvolles Werkzeug, um systematisch Datensätze auf Konsistenz zu überprüfen, um nicht messbare Metabolitkonzentrationen vorherzusagen, und um intrakompartmentelle Konzentrationen aus den gemessenen Konzentrationen (Zelldurchschnittswerte) zu berechnen. Die NET-Analyse ist eine skalierbare Methode, neues Wissen aus quantitativen Metabolomik-Datensätzen zu gewinnen, bei der nur wenige Annahmen und insbesondere kein kinetisches Modell notwendig sind.


In Kapitel 5 wurde die Glukoserepression in den *S. cerevisiae*-Stämmen CEN.PK 113-D7 und FY4 anhand von quantitativen und grossen Proteomik- und Metabolomik-Datensätzen sowie anhand von metabolischen Flüsse verglichen. Die Wildtypstämmen
Zusammenfassung

zeigen auf der Ebene der metabolischen Flüsse einen ähnlichen Grad an GlukosePRESSION, da sowohl Biomasse- und Ethanolausbeuten sowie die respiratorische Zitratzyklusaktivität ähnlich sind. Trotzdem wurden unterschiedliche Protein- und Metabolitkonzentrationen beobachtet, die darauf hinweisen, dass die Hxk2-abhängige Signalübertragung und Regulation in CEN.PK stärker ausgeprägt ist. Unterschiede in den Signal- und Regulationsstärken traten auch dadurch zutage, dass sich die Glukosepression auf Flussebene, die durch die respiratorische Zitratzyklusaktivität bestimmt wurde, infolge einer genetischen Mutation wesentlich verändert hat: Während in FY4 die Glukosepression aufrecht erhalten wird, wenn $HXK2$ eliminiert wird, geht CEN.PK zu einem respiratorischen Metabolismus über. Es konnte gezeigt werden, dass eine Punktmutation im $CYRI$-Gen für das unterschiedliche Verhalten verantwortlich ist.


1. Introduction

1.1. Background

1.1.1. Systems Biology

Living systems consist of complex networks of diverse interacting molecules. The effects of these local interactions propagate through the network such that - depending on their timing and quantity - they finally give rise to the global behavior, i.e. the biological phenotype. Understanding how a phenotype emerges from the various single interactions is the ultimate goal of systems biology as a new biological discipline [81, 94]. As the single molecules mutually influence each other, it is not sufficient to study each of them isolated from the network. Hence, systems biology adopts a holistic perspective on biological networks.

One prerequisite towards this perspective is the possibility to experimentally identify and also quantify all or a sufficiently large set of cellular components. Acquiring these data became possible through modern high-throughput analytical so-called omics techniques, e.g. metabolomics, proteomics, or transcriptomics. To administer such huge amounts of data, bioinformatics provides means for data quality control, raw data handling, and processing as well as storage and curation. However, the interpretation of the gathered data by human comprehension to gain biological insights is inherently difficult and insufficient due to the large size and high complexity of these data sets. Therefore, computational tools are an indispensable second prerequisite for systems biology. The so-called computational systems biology tools can be categorized into tools (i) to extract knowledge from the data and (ii) to integrate the knowledge into models to finally achieve a system-level understanding (cf. Fig. 1.1) [93].
1. Introduction

**Figure 1.1.** Systems biology approach. Computational systems biology tools mediate the path from system-level experimental data to system-level understanding. First, data analysis tools extract new biological knowledge from the complex system-level data. Second, models are employed to integrate the current knowledge to achieve a system-level understanding. Importantly - as symbolized in the horizontal arrow - the models which are build to gain the system-level understanding (right side), also assist in the extraction of knowledge from the complex data sets (left side).
In the first category, computational tools retrieve new insights from data to build up knowledge about the biological system. For instance, clustering approaches are applied to reveal patterns in gene expression data harboring thousands of data points. Or, kinetic parameters of enzymatic reactions are estimated from dynamic metabolome data. Importantly, here a mathematical model of the biological system assist the interpretation of the data by taking into account previous knowledge (cf. Fig. 1.1).

To attain a holistic perspective, the knowledge needs to be integrated into mechanistic mathematical models. The model here serves as a semantic mean to describe the relations between the data and knowledge entities such that it can be understood by computational tools. Hence, the system can be analyzed *in silico* to answer questions about the system’s global behavior. For instance, the viability of gene deletion mutants can be predicted [56]. Or, analysis of the biological networks’ architecture can unravel the system’s robustness properties [10, 153].

In summary, systems biology as a new approach to biological research is characterized by the transition from the reductionist approach to considering whole systems. Here, novel omics analytical techniques that aim at acquiring data on all cellular components go hand in hand with model-based computational tools that both extract new biological insight from complex data and integrate the obtained biological knowledge for a system-level understanding.

### 1.1.2. Systems Biology of Metabolism

Metabolism is a core function of any biological cell as it provides energy and produces all biomass constituents. As illustrated in Fig. 1.2, many signals and regulatory actions integratively influence the metabolism: Translation of RNA and regulatory action of signaling proteins effect the metabolic proteome, which catalyzes the metabolic reactions. The actual rates, with which the metabolites are converted, finally give rise to the cell’s phenotype as its system behavior. The metabolic network, which consists of enzymes and metabolites mutually influencing each other, is characterized by a high degree of interconnectivity (mostly induced through co-factors participating in many reactions)
1. Introduction

and a high degree of non-linearity (e.g. evoked by regulatory interactions and regulatory feedback loops). Due to these characteristics, a systems biology approach is imperative for ultimately understanding metabolism. Being in the focus of this work, the systems biology approach for studying microbial metabolism in particular is reviewed in the following.

Metabolic Data

Here, high-throughput analytical techniques to generate large-scale data sets that characterize the metabolic network’s operational state, namely fluxomics, proteomics and metabolomics are introduced and discussed.

Fluxomics $^{13}$C-based flux analysis is the standard experimental tool to analyze intracellular metabolic fluxes [166]. Shortly, by growing cells on a $^{13}$C-labeled carbon source, alternative metabolic pathways can be resolved due to different rearrangements of the labeled C-atoms in alternative pathways which leads to different labeling patterns of intracellular metabolites. These different labeling patterns are then detected
via mass spectrometry. Different approaches to metabolic flux analysis span from qualitative sample characterization to quantitative flux prediction. First, using comparative multivariate statistics, samples can be differentiated qualitatively and classified based on the detected labeling patterns [166]. Second, additionally applying a stoichiometric network model, metabolic flux ratio analysis determines fractional contributions of alternative routes for synthesis of metabolites within central carbon metabolism. Finally, $^{13}$C-constrained flux balance analysis and iterative isotopomer balancing project absolute fluxes on reactions by also taking into account measured extracellular rates [166, 206]. $^{13}$C-based flux analysis has provided significant insight into metabolic operation and regulation [167]. Despite recent advances [195, 207], it is however still limited to certain experimental setups and focusses on fluxes in central carbon metabolism.

**Proteomics** Today's high-throughput proteomics that provides large-scale data sets is based on mass spectrometry [4, 179]. For qualitative analysis, techniques to identify expressed proteins and to explore their interactions have been established over the past two decades [52, 157, 180]. For untargeted protein quantification, most approaches rely on labeling tags (i.e. isotope-coded affinity tags (ICAT), stable isotope labeling with amino acids in cell culture (SILAC)). Lately, improvements towards targeted quantification have been made through the establishment of a multiple reaction monitoring (MRM) based technique. This technique allows for higher sensitivity and enables accurate quantification of enzyme concentrations for large parts of metabolism [52].

**Metabolomics** Metabolomics has recently emerged to complement the set of omics data. As proteomics, it is primarily based on mass spectrometry. Commonly applied analysis platforms are GC/MS, LC/MS or CE/MS [50, 119]. Drawing on the definition in [133], several approaches to metabolome analysis can be distinguished (cf. [50, 65, 119, 133] for reviews). Metabolic fingerprinting is an untargeted approach to qualitatively characterize the metabolome. Without the need to relate the mass spectrometry peaks to the corresponding metabolites, the acquired mass spectra resulting from hundreds of metabolites can be used to classify samples in e.g. drug discovery
1. Introduction

or biomarker research [50, 204]. As a next metabolomics approach, profiling analyses that detect and also identify the metabolites are often applied in studying operation and regulation of microbial metabolism [119, 193]. By relating peak areas to a reference sample, semi-quantitative information on metabolite concentrations are obtained. Finally, absolute metabolite concentrations are determined in metabolite profiling and target analysis. Here, calibration with an internal standard with known concentration allows for determining the number of molecules per cell dry weight. Requisites of quantitative metabolomics are (i) a complete recovery of the metabolites during sample preparation and (ii) unique mappings of metabolites to mass spectrometry peaks or chromatograms. Reproducible sampling protocols for a variety of organisms have been established and many metabolites can be uniquely identified in the mass spectrometry chromatograms [21, 119]. The relevance of quantitative metabolomics for studying microbial metabolism is emphasized by the observation that often metabolite concentrations change significantly while metabolic fluxes are invariant [30, 131]. As metabolite concentrations respond with high amplitudes, they are a very sensitive readout on metabolic operation.

Data Analysis for Studying Metabolism

The above illustrated mass spectrometry-based analytical techniques provide large data sets of protein and metabolite concentrations and metabolic fluxes as readouts of a metabolism’s operational state. With exception of the fluxomics data being closely linked to the phenotype, the biological meaning of these multi-dimensional readouts, however, is not directly apparent. First, the data sets are too large to be directly interpretable by human comprehension. Second, a quantified value alone is per se not informative about the network’s operation. To address these issues, different methods are used in computational systems biology for extracting knowledge from the large-scale molecular readouts. For a systematic overview, a classification of the data analysis approaches based on data quality and the size of the analyzed biological system is given in Fig. 1.3. As introduced in the following, statistical, and mechanics-based data analysis tools are distinguished.
1.1. Background

Figure 1.3. Classification of data analysis approaches. Data analysis approaches are classified based on the network size of the analyzed biological system and the data accuracy level. The latter can range from readouts for which the corresponding molecules are not identified to absolute quantification of concentrations.

Statistical Methods  Statistical analysis are usually applied to analyze qualitative or semi-quantitative data sets containing hundreds of data points. Here, such high-dimensional, complex data sets that not directly interpretable, are characterized by measures that are comprehensible by a biologist (cf. [30] for review). Typically, the data is either projected to fewer dimensions or correlations within the data points are assessed. For reducing dimensionality, commonly used techniques are principal or independent component analysis that extract the parameters, which display most of the variation between different samples. To identify and group probes with similar trends or samples that are most similar, clustering methods like k-means or hierarchical clustering are applied. As no a priori knowledge is necessary and also data that lacks identification (e.g. mass spectrometry peaks that are not assigned to the respective molecule) can be analyzed, these methods benefit from their broad applicability. In the area of metabolism, metabolomics and proteomics data are commonly analyzed in this manner [28, 150, 193]. This way, metabolites that show common concentration changes upon perturbation can for example be detected and such responses to different environmental conditions be characterized.

Using purely statistical methods, causalities contained in the data can hardly be extracted. Here, considering known interrelationships between measured compounds via
network models improves the statistical data analysis (e.g. [36, 82, 138]). For instance, Cakir et al. [32] utilized the known metabolic network to derive reporter reactions, around which significant changes in metabolite concentration occur. Also, a variety of tools (e.g. [35, 67]) use statistical tests to infer an enrichment of transcriptional change of genes belonging to the metabolic pathway. While this and similar methods require the knowledge of the underlying (metabolic) network, they are still applicable to data sets harboring only relative quantification, for which at least the corresponding compounds are identified.

**Mechanistics-based Approaches** Mechanistic models that describe cause-effect chains, are able to elucidate causal relationships between the experimental readouts. Moreover, while the statistical approaches only comparatively analyze two or more data sets, mechanistic models analyze the significance of the measured physical quantities in a single data set.

Here, the data analysis tools can be distinguished based on whether they can be applied (i) to large networks and for many organisms and experimental conditions or (ii) to specific biological networks at a certain experimental setup. The former methods employ mathematical models that are based on only limited information on the considered organism (e.g. network topology) and/or utilize universal physical parameters like Gibbs energies (e.g. as in thermokinetic description of reaction rates [130]). Metabolic control analysis based on linlog kinetics for example reveals flux controlling reactions within the metabolism based on measured metabolite concentrations.

In contrast to these approaches, tools that analyze experimental data for specific biological systems employ models that describe mechanisms of each considered reaction or interaction. Here, measured concentrations can then be used to estimate kinetic parameters of the model. The application of these approaches is limited as detailed kinetic models are laborious to construct and are therefore often not available. Nevertheless, detailed mechanistic models of specific pathways like glycolysis or amino acid synthesis pathways were used to analyze small sets of absolute quantitative concentrations.
Current Challenges in the Analysis of Metabolic Data

The advances in metabolomics and proteomics enable a novel, quantitative perspective onto microbial metabolism which challenges the data analysis to exploit the wealth of information contained in the quantitative data. Especially, the analysis of the emerging quantitative metabolomics data sets is difficult: There exists no unambiguous cause-effect-chain such as for the linear relationship between genes, transcripts, and proteins that would assign a function to a metabolite. A metabolite is always educt and product at the same time and, moreover, can be an allosteric regulator. Furthermore, many factors (such as kinetic parameters, enzyme concentrations, etc.) influence a metabolite’s concentration. Such, metabolite concentrations are a systematic readout on the metabolic state, and therefore, data analysis tools that integratively analyze quantitative metabolomics data are necessary.

For the analysis of quantitative data sets both statistical tools and model-based approaches can be applied. However, the analysis of the quantitative data sets of yet rather small scale by purely statistical approaches, that are valuable tools to analyze data sets of very large scale, to yet smaller scale, suffers too low statistical significance. Moreover, as they do not consider cause-effect-chains and do not exploit the provided quantitative information in terms of molecules per cell they would not exploit the wealth of information provided by the quantitative data sets.

Among the model-based approaches, on the contrary, generally applicable or specific models as described above could be utilized. Here, however, the availability of specific models is hampered as recent models often span only small parts of metabolism and the model reconstruction is laborious. They are also posing high requirements on the quality and coverage of the analyzed data. For instance, parameter estimation is only feasible with dynamic data with sufficient time-resolution.

Hence, tools are necessary in the field between large-scale qualitative statistical and specific mechanistic-model based analysis approaches: They should analyze quantitative
1. Introduction

metabolic data in a mechanistic manner and should be general enough to allow for widespread application. Today, these tools are not sufficiently available and novel data analysis approaches need to be developed.

Model-based Knowledge Integration for Achieving System-level Understanding

The data analysis tools in systems biology but also other biological research areas provide knowledge about metabolism on molecular level (e.g. identification of all metabolic proteins or metabolic reactions). Next, this metabolic knowledge on components and interactions has to be transformed into system-level understanding of the metabolic operation. For this challenging task, mathematical models are essential tools (cf. 1.1) and various kinds have been reconstructed that integrate our current knowledge. These models are then used to analyze the metabolic network’s characteristics and its operation to achieve a system-level understanding. As introduced in the following, static and dynamic models can be distinguished (cf. [181] for review on - not exclusively metabolic - models).

Static Models Due to our broad knowledge about the metabolic reactions, stoichiometric reconstructions of the metabolic network are available for a number of organisms (e.g. at [184]). Basically, these models describe the stoichiometry of metabolic reactions and are often represented as metabolic maps. The genome-scale stoichiometric models that are reconstructed based on amongst others the annotated genome are compendia on the - preferably - whole metabolic capacity of an organism. Due to their high coverage of metabolic reactions they find particular attention and are available for several organisms (e.g. [53, 54, 72, 155, 169]).

To obtain system-level insights from these models, graph-theoretical approaches have been used to analyze metabolic network structure and reveal its global organization in hierarchies and modules [85, 153]. Furthermore, not only the structure but also possible metabolic fluxes can be assessed. Here, flux balance analysis is an approach to predict metabolic fluxes - amongst others also growth rates. This capability can for instance be used to systematically assess gene essentiality by predicting maximal biomass yields in
1.1. Background

In silico knockout mutants [56]. Or, reaction that typically carry high metabolic fluxes were identified and flux organization analyzed by sampling the flux solution space [8]. As rarely metabolic regulation is considered in these models, their predictive quality is very limited [173].

Dynamic Models In the field of metabolism, dynamic models are typically composed of ordinary differential equations (ODEs) that describe concentration changes over time as a result of production or consumption rates. These rates are defined in mathematical reaction terms. Dynamic models give insight into the modeled system via simulation studies, i.e. in silico experiments, or by assessing model characteristics, e.g. by sensitivity analysis [97]. Here, metabolic control analysis, a conceptual framework based on sensitivity analysis [58], employs kinetic models to elucidate how metabolic fluxes are controlled. So-called flux control coefficients reveal the key compounds and reactions that have much influence on the metabolic fluxes [76, 99, 156].

Dynamic models can be classified based on the mathematical terms that determine the reaction rates. Here, either terms that describe the molecular mechanism of the reaction kinetics (e.g. Michaelis-Menten kinetics) or terms that do not describe a certain reaction mechanism but are suitable to fit experimental data can be employed. Using reaction mechanism-based terms, dynamic models often cover only rather small metabolic (sub-)networks, e.g. glycolysis [38], mitochondrial respiratory system [11], or the penicillin biosynthesis pathway [145]. Also, most of today’s models only focus on one level (i.e. only metabolite reactions or only signaling and regulation pathways). One of the few exceptions where multiple levels were considered is the model of yeast’s osmotic shock response [98].

The extension of mechanism-based dynamic models to larger networks is difficult due to lacking knowledge of the correct reaction mechanisms and due to difficult estimation of large sets of parameters from measured data. There exists only rare examples of whole-cell kinetic models as for instance for the human red blood cell that was realized through extensive modeling efforts over the last three decades [83]. As one reason for
1. Introduction

tediousness of the modeling process, the terms, that describe the reaction mechanisms, have to be adopted for each organism while for instance the reaction stoichiometry is conserved.

As an alternative, models based on generic reaction terms (linlog kinetics [101], thermokinetic descriptions [130] or convenience kinetics [110]) can be applied to every reaction. As these models disregard the specific enzymatic reaction mechanisms, their prediction and simulation capacity beyond the experimental conditions used for parametrization is limited. However, such models still have great value for metabolic control analysis: The model parameters (i.e. elasticity coefficients) of linlog kinetics and thermokinetic descriptions are directly linked to flux control coefficients [101, 130]. Again, these model do not span multiple cellular levels.

Current Challenges in Constructing Metabolic Models  While today static stoichiometric models cover an organism’s metabolic network almost completely, dynamic models are often only available on small scale. As also recognized in recent articles, the reconstruction of dynamic models on large-scale is necessary [84, 178]. In these publications, the challenges to build large-scale dynamic models of metabolism are discussed and approaches to accomplish this task in efficient manner are suggested. Also, as metabolic models often only consider the metabolic reaction level, more efforts are necessary to link these to models on signaling and metabolic regulation.

1.2. Outline and Positioning of this Thesis

This thesis centers on microbial metabolism. It addresses challenges in the area of mechanism-based analysis of quantitative omics data. In particular, computational tools to analyze quantitative metabolome data and for reconstruction of stoichiometric metabolic models were developed. Also, state-of-the-art metabolomics and proteomics data were applied to characterize the metabolic state and to infer the underlying regulatory mechanisms.

Thermodynamic principles that enjoy general applicability represents the unifying
1.2. Outline and Positioning of this Thesis

theme throughout Chapters 2-4 in which tools are described that link these principles to stoichiometric models of the metabolic network. In Chapter 2, a tool is introduced that assists in the reconstruction of genome-scale metabolic models. While the list of metabolic reactions can, to large extent, automatically be derived from the annotated genome sequence, the genome contains no information on possible reaction directions. In this work, a systematic thermodynamics-based assignment of reaction directionalities was developed and demonstrated on the a genome-scale reconstruction of *Escherichia coli*’s metabolism [155]. In Chapter 3, a tool for mechanism-based analysis of quantitative metabolome data is presented: The network-embedded thermodynamic (NET) analysis systematically integrates a stoichiometric model of the metabolic network and known flux directions with metabolite concentrations via the second law of thermodynamics. Such, the thermodynamic feasibility of measured metabolite concentration can be tested and non-measured metabolites can be predicted. Moreover, putative regulatory sites within the metabolic network can be identified. As described in Chapter 4, the NET analysis was combined with elementary flux mode analysis to analyze metabolic fluxes. These are usually experimentally determined by $^{13}$C-based flux analysis which however has limitations in the scope of applicable experimental conditions and in the ability to resolve fluxes. Here, a metabolomics-based approach to infer metabolic activities was developed that is not restricted with respect to the experimental setup. And, as demonstrated on a *Saccharomyces cerevisiae* data set, metabolic activity was inferred for reactions which are not accessible by $^{13}$C-based flux analysis.

Next to the development of modeling and data analysis tools, *S. cerevisiae*’s metabolism was analyzed by drawing on state-of-the-art quantitative metabolomics and proteomics data sets. Specifically, two common yeast strains were compared by assessing protein and metabolite concentrations as well as metabolic fluxes for glucose exponential growth (Chapter 5). Based on the observed differences in the metabolic state of the wild type and genetically perturbed strains, differences in the signaling and regulatory strength of signaling and regulatory proteins and transcription factors was inferred. The appendix presents the experimental analysis of *S. cerevisiae*’s metabolic reprogramming
1. Introduction
during diauxic growth which provided the experimental basis to develop the combined
NET/EFM analysis described in Chapter 4. Here, the distinct phases of the diauxic shift
were characterized by the intra- and extracellular metabolite and protein concentrations.
Currently only the metabolites concentration are available.
2. Systematic Assignment of Thermodynamic Constraints in Metabolic Network Models

Kümmel A., Panke S., Heinemann M.

Systematic assignment of thermodynamic constraints in metabolic network models

*BMC Bioinformatics, 7:512, 2006*
Abstract

The availability of genome sequences for many organisms enabled the reconstruction of several genome-scale metabolic network models. Currently, significant efforts are put into the automated reconstruction of such models. For this, several computational tools have been developed that particularly assist in identifying and compiling the organism-specific lists of metabolic reactions. In contrast, the last step of the model reconstruction process, which is the definition of the thermodynamic constraints in terms of reaction directionalities, still needs to be done manually. No computational method exists that allows for an automated and systematic assignment of reaction directions in genome-scale models.

We present an algorithm that - based on thermodynamics, network topology and heuristic rules - automatically assigns reaction directions in metabolic models such that the reaction network is thermodynamically feasible with respect to the production of energy equivalents. It first exploits all available experimentally derived Gibbs energies of formation to identify irreversible reactions. As these thermodynamic data are not available for all metabolites, in a next step, further reaction directions are assigned on the basis of network topology considerations and thermodynamics-based heuristic rules. Briefly, the algorithm identifies reaction subsets from the metabolic network that are able to convert low-energy co-substrates into their high-energy counterparts and thus net produce energy. Our algorithm aims at disabling such thermodynamically infeasible cyclic operation of reaction subnetworks by assigning reaction directions based on a set of thermodynamics-derived heuristic rules. We demonstrate our algorithm on a genome-scale metabolic model of E. coli. The introduced systematic direction assignment yielded 130 irreversible reactions (out of 920 total reactions), which corresponds to about 70% of all irreversible reactions that are required to disable thermodynamically infeasible energy production.

Although not being fully comprehensive, our algorithm for systematic reaction direction assignment could define a significant number of irreversible reactions automatically with low computational effort. We envision that the presented algorithm is a valu-
able part of a computational framework that assists the automated reconstruction of genome-scale metabolic models.

2.1. Background

Nowadays, high-throughput experimental omics techniques are being developed and are generating large-scale data sets and information bases that can hardly be intuitively understood. Models that enable mathematical analysis and simulation are essential to benefit from the knowledge that is contained in these data sets. Consequently, the importance of models increases along with the advances in experimental technologies.

One class of models that has particularly proven to be useful for the analysis of omics data is the class of stoichiometric metabolic models [149, 154]. Several such models - today typically available on genome-scale - were reconstructed for various organisms (e.g. [61, 72, 155]) and are used as tools in systems biology [102, 138] and metabolic engineering [139, 144]. Genome-scale stoichiometric models are composed of the metabolic reactions' stoichiometry and assignments of the reactions' reversibility or irreversibility.

In the model reconstruction process - reviewed in [41] - typically first a preliminary organism-specific metabolic network is generated by drawing on information stored in sequence databases that link coding regions to cellular functions. In the next step, the sequence-derived network is completed with knowledge from biochemistry and physiology such that a stoichiometric network is derived that reflects the cell's metabolic capabilities. For the reconstruction of metabolic reaction networks and particularly for the identification of enzymes that lack genetic evidence, a series of computational tools exist [66, 88, 92].

In contrast, reaction directions are often assigned manually, or are adopted from other existing models or databases on metabolic pathways (e.g. KEGG). Direction assignment is not only laborious but also error-prone due to manual execution. Indeed, it was shown that the direction assignments of published genome-scale models contain inconsistencies i.e. reaction directions that contradict each other (M. Terzer and J. Stelling, personal communication). Error diagnostics in these cases is difficult as the underlying reasons
2. Systematic Reaction Direction Assignment

for direction assignments are not provided in the currently available models.

Reaction directionalities are used frequently: First, they are required for analysis and simulation of metabolic phenomena by constraint-based modeling [149]. Second, the reactions’ directionality is usually reported in maps on metabolic pathways within widely-used genomic databases such as KEGG or MetaCyc [34, 87]. Third, the information on reactions’ (ir)reversibility is essential for metabolic flux analysis [206].

In principle, all biochemical reactions are reversible: A reaction can proceed in either forward or backward direction depending on the actual Gibbs energy of reaction. The Gibbs energy of reaction is determined by the reactants’ standard Gibbs energies of formation and their concentrations. A change in reactant concentrations, for example, can reverse a reaction’s direction, if the respective Gibbs energy of reaction turns from a negative to a positive value. There are, however, so-called irreversible reactions that under physiological conditions only proceed in one direction, which means that the respective reactants’ Gibbs energies of formation and concentrations are always such that only one direction is possible.

The natural approach to identify the irreversible reactions in stoichiometric models would draw on Gibbs energies of formation and physiological concentration ranges. However, experimentally determined Gibbs energies of formation are not available for all metabolites. As a workaround, a group contribution method was developed that computationally estimates Gibbs energies of formation for a large set of metabolites [121]. Using the values obtained with this method and taking into account its inherent significant uncertainties, a genome-scale thermodynamic analysis of E. coli’s metabolism showed that only five reactions (out of 873) could be classified as irreversible [74]. This very small number demonstrates that computationally estimated Gibbs energies of formation are too uncertain to be used to assign reaction directions.

An alternative approach to assign reaction directions is to apply the nonlinear constraint that arises from the fact that there must exist a thermodynamic driving force for any mass-balanced set of reaction fluxes in a reaction network [13]. For internal reaction cycles that result in no net conversion of metabolites the overall thermodynamic
driving force is zero, i.e. the cyclic operation of these reactions is infeasible. Given the specification of the directions of a subset of network fluxes (e.g. by using information about the environmental conditions to specify the exchange of metabolites with the environment), it was shown to be possible to compute the feasible direction of some of the not preset fluxes based on the nonlinear thermodynamic constraints [211]. This ab initio calculation of the reaction directions is based on an NP-complete computation [211]. As a result, a computationally effortless algorithm for the assignment of reaction directions (thermodynamics-based linear constraints) in genome-scale networks does not exist today.

Here, we present a computational method that is intended to close this gap. In a first step, our method draws on experimentally determined thermodynamic data, i.e. Gibbs energies of formation, and physiological intracellular metabolite concentrations to assign as many reaction directions as possible. Next, in order to assign further reaction directions, we draw on network topology and heuristic rules that exploit the knowledge of biochemical energy equivalents such as ATP or NADH. An algorithm that is capable to apply this procedure to genome-scale stoichiometric models was developed and implemented in Matlab. The respective script is available from the authors on request.

### 2.2. Results

In the following, the algorithm (cf. overview in Fig. 2.1) is described in detail. Each step is illustrated by applying the procedure to the genome-scale reconstruction of E. coli's metabolic network [155]. From this model, we only used the stoichiometric matrix but not the constraints that were placed on the reaction directions. In other words, we applied our algorithm to the metabolic network, in which initially all reactions were considered as reversible.

#### 2.2.1. Thermodynamic Facts-based Assignment

First, we aimed to assign as many directions as possible on solid thermodynamic grounds: A reaction can only proceed in direction of a negative Gibbs energy of reaction, \( \Delta_r G \).
2. Systematic Reaction Direction Assignment

Figure 2.1. Illustration of the algorithm for systematic assignment of reaction directions. Panel a gives an overview over the direction assignment procedure. Each step (white boxes) is described in detail in panel b.
The Gibbs energy of reaction depends on the reactants' standard Gibbs energies of formation, \( \Delta_f G_i^0 \), their concentrations, \( c_i \), and the respective stoichiometric coefficients, \( \nu_i \):

\[
\Delta_r G = \sum_i \nu_i \Delta_f G_i^0 + RT \ln(\prod_i c_i^{\nu_i}).
\]

If it turned out in our analysis that with any physiologically reasonable reactant concentrations, the Gibbs energy of reaction for a given reaction was always negative, the reaction was defined as irreversible in the respective direction as a net reaction in the other direction is thermodynamically not possible [7]. For the Gibbs energy of formation, we employed experimentally derived values, which were available for 157 out of 761 metabolites present in the network (cf. Methods). Although a computational method can roughly estimate \( \Delta_f G^0 \)-values for many more molecules [121], we preferred to employ this limited set of thermodynamic data as only a very limited set of irreversible reactions could be assigned with computationally determined \( \Delta_f G^0 \)-values due to their inherent uncertainties (cf. Background and [74]).

Furthermore, Maskow and von Stockar have shown that only with Gibbs energies of formation, that are adjusted to physiological pH and ionic strength, e.g. a flux through glycolysis is thermodynamically feasible [120]. Thus, we considered physiological pH and ionic strength values (cf. Methods) by using the respectively transformed Gibbs energies of formation/reaction [7]. For simplicity, ‘transformed Gibbs energies’ will only be referred to as ‘Gibbs energies’ in the following.

Intracellular metabolite concentrations were also required to determine the actual Gibbs energies of reaction. These are widely unknown. As, moreover, any stoichiometric model is usually applied for a variety of growth conditions and even for mutant strains where different concentration levels can be conceived, we anyhow wanted to base our analysis on concentration ranges that cover a wide spectrum of conditions. Therefore, we here assumed broad physiological ranges for intracellular metabolite concentrations, which typically are in the order of \( \mu M \) to \( mM \) [62].

Employing a respective concentration range from 0.001 to 10 \( mM \) and by using the available set of experimental values for Gibbs energies of formation, ranges of Gibbs
energies of reaction could be determined for 176 (out of 920) reactions in the model. In this set of ranges, we checked for allowed operational reaction directions: A positive (negative) direction was set if the range of Gibbs energy of reaction was exclusively negative (positive). With this approach, 43 reactions were defined as irreversible in the analyzed *E. coli* model, while 133 where defined as reversible.

As the assignment depends on the estimated Gibbs energies of formation, we performed a sensitivity analysis to assess the assignment’s reliability. We widened the allowed ranges of Gibbs energies of reaction by 1, 2, 3, 4 kJ/mol and performed assignment runs using these. Despite the broadened ranges, 40 out of 43 of our direction assignments based on thermodynamic facts were still valid. Only up to three reactions (depending on the uncertainty range used) would not had been defined as irreversible. As our earlier direction assignments are in-line with the reaction directions in the original model and also in KEGG, we believe that our irreversibility assignments are correct.

2.2.2. Thermodynamic Heuristics-based Assignment

The limited availability of experimental Gibbs energies of formation only allowed us to analyze a rather small subset of reactions. Thus, we expanded the direction assignment procedure by another approach. As shown in [211] the reaction network comprises sets of reactions whose simultaneous operation would contradict fundamental thermodynamic principles. Thus, also we aimed at identifying thermodynamically infeasible subnetworks from the metabolic network. In contrast to [211], we used a different kind of subnetwork which will be outlined below. After having identified these subnetworks, heuristic rules were employed to pinpoint the reaction(s) in the identified subnetworks which most likely are irreversible and reaction directions were set accordingly. We employed the co-substrate converting cycles to identify reactions that most likely are irreversible under all conceivable environmental conditions. The direction assignment based on topology and heuristics was also implemented in the algorithm (cf. Fig. 2.1, steps 2a-2c).

Note that it is conceivable that a direction assignment based on the topological considerations contradicts an assignment made with the thermodynamic facts. Here, this was,
2.2. Results

however, never the case. For some reactions the thermodynamics facts were only less restricting as they allowed both directions while a heuristic rule constrained the reaction into one direction. To prevent the exclusion of actually possible reaction directions, we adopted the restriction only if the heuristics-based assignment was highly reliable (see below).

Identification of Thermodynamically Infeasible Operation of Reaction Sets

First, we had to identify sets of reactions (subnetworks) whose simultaneous operation is thermodynamically infeasible. A thermodynamically infeasible operation of a subnetwork is, for example, given by a cyclic operation of a reaction set that in total results in no net conversion of metabolites. The absence of such reaction cycles is a necessary condition for thermodynamically consistent operation of reaction networks [12]. Hence, cycles in the metabolic network are a promising target to screen for thermodynamically infeasible reaction directions. Cycles can be obtained via the null space of the stoichiometric matrix.

Consider a network that consists of three reactions with the pairwise interconversion of the reactants A, B and C (cf. Fig. 2.2a). Assume a situation where A is actually converted to B, and B to C. Thus, C must have a lower Gibbs energy of formation than A. Consequently, the operation of the reaction 3 from C to A is not possible. This example shows that, if we preset a consecutive operation of two reactions, it is possible to exclude one direction of the third. Thus, here, we only can state if-then relationships for reaction directions, and consequently, an \textit{a priori} determination of reaction directions - without the assumption of other reaction directions - is not possible.

![Figure 2.2. Illustration of reaction cycles.](image)
2. Systematic Reaction Direction Assignment

Next, we extend the thought experiment and assume a reaction between A and C with a different stoichiometry that is actually able to re-cycle C to A (cf. Fig. 2.2b). This reaction would have to be driven by a “motor” that delivers the energy necessary to convert the reactant C to the higher energy state of A. In metabolism, chemical energy can be delivered by the conversion of a co-substrate (e.g. ATP) containing much energy (in the following termed as highly energetic) to its counterpart (e.g. ADP) that carries less energy (so-called low energetic). In this case, a cyclic operation of the reactions from A to B, B to C and C back to A is thermodynamically feasible as the system is supplied with energy. On the contrary, the reverse operation of this reaction cycle is thermodynamically infeasible as the “motor” would then operate in opposite direction and would become a “generator”: The cycle would produce energy (e.g. in the form of ATP). In order to exclude such thermodynamically infeasible energy production, one of the reactions in this cycle is set irreversible such that the highly energetic co-substrate cannot be produced\(^1\). The reaction that produces the highly energetic co-substrate was here the preferred target to assign a direction that only allowed energy consumption. In the context of our work, we call a model “thermodynamically feasible” if no generation of highly energetic co-substrates by a cyclic operation of metabolic reactions is possible.

Several pairs of low/highly energetic co-substrates exist (cf. Tab. 2.1). These are pairs of (i) nucleotide phosphates, of (ii) nicotinamide adenine dinucleotides, of (iii) nicotinamide adenine dinucleotide phosphates and of (iv) flavin adenine dinucleotides, and (v) intra- and extracellular protons. Due to the proton motive force over the membrane, extracellular protons are high-energy counterparts to intracellular protons.

To identify cycles that interconvert these co-substrates, again the null space of the stoichiometric matrix was calculated, however, only after having removed the co-substrates’ stoichiometric coefficients from the matrix. The respectively obtained null space then included two sets of cycles: (i) the cycles, that do not produce or consume any metabolite, and which were already described by the null space of the original stoichiometric matrix

---

\(^1\)Obviously, ATP can is produced in reaction that convert highly energetic to low energetic metabolites. However, these cannot converted back without the supply of energy as it would be the case in a reaction cycle operating in “generator” mode.

24
2.2. Results

Table 2.1. Co-substrate groups that were eliminated from the stoichiometric matrix to identify energy producing cycles. Besides these co-substrates, also the molecules water, oxygen, carbon dioxide, ammonium, and inorganic phosphate were removed from the stoichiometric matrix to identify energy producing cycles. This is necessary to elementally balance the resulting net conversion of co-substrates. NTP, NDP, and NMP denote nucleoside tri-, di- and monophosphate for adenosine, cytidine, guanosine, inosine, and uridine.

<table>
<thead>
<tr>
<th>selected co-substrate groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTP, NDP, NMP</td>
</tr>
<tr>
<td>NADH, NAD$^+$</td>
</tr>
<tr>
<td>NADPH, NADP$^+$</td>
</tr>
<tr>
<td>FADH$_2$, FAD$^+$</td>
</tr>
<tr>
<td>H$_{\text{extracellular}}^+$, H$^+$</td>
</tr>
</tbody>
</table>

and (ii) cycles that - when complemented with the removed co-substrates - interconvert these. In terms of the terminology introduced in network-based metabolic pathway analysis [136], these two sets of cycles correspond to the extreme pathways of Type III and Type II, respectively. Having complemented the cycles with the co-substrates, we determined the net conversion of co-substrates for each cycle and identified the cycles that convert low energetic co-substrates to their highly energetic counterparts. In the following steps, we worked with this set of cycles to assign reaction directions, and here, we will call these energy producing subnetworks solely ‘cycles’.

Every possible energy producing cycle is a combination of the linearly independent basis vectors of the null space of the reduced stoichiometric matrix. As the running time for the computation of all linear combinations increases exponentially with system size [12, 137], an exhaustive analysis of all possible cycles is currently not feasible (M. Terzer and J. Stelling, personal communication). For this reason, we based our assignment procedure on the cycles that are described by the basis vectors of the calculated null space matrix. As we will see below this approach was not fully comprehensive but allowed excluding thermodynamically infeasible cycling to a large extent while still being computationally reasonable.

Faced with the fact that we only obtained one possible set of linear independent basis vectors, the choice of the null space matrix calculation, however, was important for the assignment procedure. In preliminary tests, when we applied a null space matrix that included larger cycles, our algorithm assigned less reaction directions. Thus, one should apply a null space matrix with cycles that consist of the smallest possible number of
2. Systematic Reaction Direction Assignment

reactions. Here, the null space matrix was calculated from the reduced echelon form of the stoichiometric matrix by the Matlab function \textit{null}. The null space of the co-substrate reduced stoichiometric matrix was described by 227 linear independent reaction cycles with an average number of reactions of 8.85 and a median number of reactions of 4. Of all cycles within the null space matrix, 145 were energy producing cycles.

This set of cycles was now employed to assign reaction directions by thermodynamics-based heuristic rules: In three steps that are described in the following paragraphs different kinds of cycles were analyzed and reaction directions were assigned by the heuristic rules in Fig. 2.3. These rules selected for and disabled reaction steps that produce high-energy from low-energy co-substrates. Such, we could assign directions for reactions beyond the ones, for which the Gibbs energies of reaction were available.

\textbf{Analysis of Pair Cycles}

Cycles that consist of only two reactions and in total convert exclusively co-substrates occur frequently in metabolic networks. Here, the null space matrix contained 45 energy producing pair cycles. For such cycles, the direction assignment to eliminate thermodynamically infeasible energy production is straightforward due to the limited possibilities for assignment of reaction directions: There are only two reactions which can be set irreversible, and the most natural approach is to block the reaction step that produces the highly energetic co-substrate. This was the only heuristics-based assignment step that was allowed to be more restricting than the thermodynamic facts assignment step. Technically, this procedure was realized by applying heuristic rules as explained in Fig. 2.3.

Applying these heuristics to the identified energy producing pair cycles, 42 reactions were restricted to one direction. As none of these reactions was already previously defined as irreversible in the thermodynamic facts-based assignment, in summary 85 direction assignments were made until here.

\textbf{Analysis of Remaining Energy Producing Cycles}

The following assignment step (2b in Fig. 2.1) aimed at defining reaction directions in the remaining energy producing cycles. As these consist of more than two reactions,
2.2. Results

Figure 2.3. Illustration of the procedure to assign reaction directions by heuristic rules. For the assignment steps 2a-2c (cf. Fig. 2.1), the applied heuristic rules are displayed. Generally, the rules defined a reaction as irreversible in the direction of consumption of a high-energy co-substrate. The rules, however, were not applied if the respective reaction simultaneously produced CO$_2$. The vertical arrows indicate the consecutive application of the rules: if no assignment was possible with a particular heuristic rule, the next rule along the arrow was employed. The consumption of a co-substrate with a higher energetic content was preferred over the consumption of a co-substrate with a lower energetic content. In case the cycle contains more than one reaction producing the same highly energetic co-substrate, all these reaction steps are defined as irreversible.

In pair cycles (2a), the reaction that produces the only generated co-substrate was defined as irreversible as long as it did not consume CO$_2$. In case of CO$_2$ consumption, however, it follows that the other reaction also produces CO$_2$ and we define this reaction direction as irreversible. As only one co-substrate pair is converted in each pair cycle, the assignment was achieved by applying the heuristic rules consecutively while omitting the first rule as indicated in the figure.

In the analysis of the remaining energy producing cycles (2b), the cycles can contain reaction steps that produce different kinds of co-substrates. Here, in the first place we restricted CO$_2$ consumption, which is very energy consuming and therefore likely to render a reaction step thermodynamically infeasible. If no CO$_2$ consuming reaction was preset, the production of highly energetic co-substrates were disabled with the indicated priorities. Note that NADPH and NADH producing reactions, here, were assigned with the same priority (not illustrated in the figure).

In the bypass analysis (2c), reaction directions were assigned for CO$_2$ consuming or nucleotide triphosphates producing reactions. Preliminary studies showed that only these heuristic rules were fully reliable in this assignment step, and thus, we only applied these two rules.
several conceivable options to disable energy producing cycling typically exist. Hence, it is important to note that this step of the heuristics-based assignment is less reliable.

Applying heuristic rules as depicted in Fig. 2.3 to the remaining 45 energy producing cycles in the null space matrix, 26 reactions were suggested to be irreversible. Five of these, however, were identified to be reversible in the thermodynamic facts-based assignment. In these cases, we preferred to follow the thermodynamic facts-based assignment for the following reasons: First, the Gibbs energy of reaction is the hard physical ground for a reaction’s directionality. Second, by setting a reaction reversible we do not exclude directions that indeed are possible under some physiological conditions. Therefore, only the 21 directions that do not further constrain the thermodynamic facts-based assignments were adopted, and as a result, at this point 106 reaction directions were defined in total.

**Analysis of Bypasses**

So far, only the energy producing cycles of the initially calculated null space matrix were analyzed and blocked by the outlined procedure in case the heuristic rules were applicable. As the calculation of all possible cycles is currently not feasible, in the next step (2c in Fig. 2.1), we at least investigated pairwise combinations of the complete set of available cycles - including also the non-energy producing cycles: In case a second cycle could act as a bypass for an already identified infeasible reaction step of a first cycle, we aimed to exclude the operation of the bypass.

The bypasses were identified as follows. Each reaction, which was defined as irreversible and disabled an energy producing cycle, was analyzed. Among all cycles in null space matrix (also including non-energy producing cycles), we selected for those in which the analyzed reaction occurs. This subset of cycles is capable to bridge the particular reaction of the first cycle, i.e. to form bypasses that start at the reaction’s educts and ends at its products. In consequence, the initial cycle, whose thermodynamically infeasible operation was already disabled, and the bypass build a new - potentially energy producing - cycle. For each identified bypass, we first checked whether it was an actually
2.3. Discussion

Achieved Direction Assignment

Tab. 2.2 summarizes all assignments that were made by our systematic procedure. While the thermodynamic facts-based assignment yielded 43 irreversible reactions, 87 further reaction directions were assigned based on network topology and thermodynamic heuristics. Altogether, 130 reactions were restricted to one direction, which disabled the operation of 129 of the 145 energy producing cycles present in the employed null space matrix.

<table>
<thead>
<tr>
<th>assignment step</th>
<th>number of assigned directions per step</th>
<th>in total</th>
</tr>
</thead>
<tbody>
<tr>
<td>thermodynamic facts</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>thermodynamic heuristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pair cycles</td>
<td>42</td>
<td>85</td>
</tr>
<tr>
<td>remaining energy producing cycles</td>
<td>21</td>
<td>106</td>
</tr>
<tr>
<td>bypasses</td>
<td>24</td>
<td>130</td>
</tr>
</tbody>
</table>

functional bypass given the previously made direction assignments. If the bypass was already blocked, there was no need for any action. Otherwise, we checked whether the co-substrate conversion of the resulting new cycle was thermodynamically infeasible by calculating the cycle’s Gibbs energy of reaction. If it was infeasible, a reaction direction within the bypass was assigned by applying the heuristic rules illustrated in Fig. 2.3.

Analyzing the bypasses that bridge the 106 previously assigned reactions, in a first iteration step 24 additional irreversible reaction directions were defined by the heuristic rules. In a second iteration step, in which we analyzed bypasses for the reactions that were defined as irreversible in the first iteration step, no further directions could be assigned: The bypasses were either already inhibited or no further reaction directions could be identified with the employed heuristics.

2.3. Discussion
2. Systematic Reaction Direction Assignment

Our algorithm did not completely disable thermodynamically infeasible energy production: The heuristics failed in blocking all energy producing cycles and the bypass analysis was not able to identify all possible energy producing cycles. In order to assess the completeness achieved with our approach, we estimated how many additional direction assignments had to be made to completely prohibit infeasible co-substrate conversion. For this, an iterative procedure was applied: A possible energy producing cycle was identified using flux balance analysis, and then, reaction directions were assigned manually to block this cycle (cf. Methods section). When no further energy producing cycles were found, the reactions' directionalities were assumed to reflect thermodynamic feasibility with respect to energy generation. At this point, the direction assignment was considered to be complete.

With this procedure, 59 additional assignments of reaction directions were required until infeasible energy production was excluded. Simulating aerobic growth on glucose by flux balance analysis, ATP was then produced via the respiratory chain. Importantly, the production of energy equivalents such as ATP by metabolic reactions was not generally rendered impossible by our linear constraints as our algorithm only selectively disables the generation of highly energetic co-substrates. In summary, the 189 irreversible reactions (of which 130 were assigned by our algorithm) were sufficient to yield a thermodynamically reasonable model with respect to the production of energy equivalents.

At this point, we checked whether the application of general biochemical rules such as defining all kinase reactions as irreversible would have been a much simpler and also valid alternative to our approach. A close inspection of the 74 kinase reactions in the model revealed that this would not have resulted in a correct model: For instance, the phosphoglycerate kinase reaction is known to operate in both directions and it is correctly defined as reversible in our assignment. This demonstrates that employing heuristic rules in combination with analyzing co-substrate converting cycles is superior to simple general biochemical rules.

With the model analyzed here, the calculation time required for the assignment pro-
procedure was roughly two minutes on a Pentium 3GHz PC, if the calculation of the null space matrix and generation of a Excel file for output documentation is included. The assignment algorithm itself required about 30 to 40 s. Such, the computational effort is small and the algorithm can be efficiently executed on a commonly used PC.

### Comparison to Original Model

The introduced systematic direction assignment yielded 130 reactions that were restricted in one direction. Together with the 59 manual assignments that eventually eliminated any thermodynamically infeasible cycling, we obtained 189 reactions that are irreversible in our model. In comparison to the 676 irreversible reactions in the original model from Palsson and co-workers [155], this is a rather small number and indicates a much less constrained model.

From a constraint-based modeling viewpoint, a direct comparison of the number of irreversible reactions, however, is misleading as one assigned reaction direction can practically render impossible the reversible operation for a set of other reactions. For example, one irreversible reaction that is part of an unbranched linear pathway restricts the operation of the whole pathway to one direction. Hence, in effect it is no difference if the direction of only one or all reactions of the pathway are defined as irreversible.

To allow for assessment of model flexibility due to different direction assignments, we had to identify correlated sets of reactions (cf. Methods). Using the identified correlated sets, the number of *de facto* irreversible reactions was assessed. We found that the stochiometric network of *E. coli* comprises 175 sets of correlated reactions. If one reaction in such a set is defined as irreversible, mass balance constraints rule out one particular direction for each of the other reactions in the set. In the original model, 749 reactions are practically irreversible. In comparison, our direction assignment eventually resulted in 292 reactions that practically can operate only in one direction.

We found that only in one case - namely the UTP-glucose-1-phosphate uridylyltransferase reaction - our algorithm defines a reaction as irreversible which is reversible in the original model. Remarkably, our assignment is in agreement with the EcoCyc database.
2. Systematic Reaction Direction Assignment

[1] which also states that this reaction is irreversible.

As the predicted maximal biomass yield on glucose is increased by about 20% using our reaction directions in comparison to the original, the model with our reaction directions is much less constrained and there are more possibilities to distribute the mass flux through the reaction network. Therefore, it is envisioned that it covers a larger range of metabolic scenarios, e.g. knockout mutants or different environmental conditions. As an example, a fnlA deletion mutant (in vivo viable when grown anaerobically on glucose [43]) is in silico non-viable with the original reaction directions while it is viable with our reaction directions.

Extension of Heuristic Rules

Next, we evaluated whether we could complement the employed heuristic assignment rules to increase the number of reactions that are automatically defined as irreversible. Additional or modified heuristic rules should eliminate the energy producing cycles that were not yet disabled by our algorithm.

First, we closely inspected the additional manual direction assignments that were required to eliminate all the remaining energy producing cycles (cf. Additional file 1). In this reaction set, we found reactions, which potentially could have been made irreversible by the heuristics already used in the algorithm, i.e. reactions that produce/consume high-energy/low-energy co-substrates, but for several reasons (as outlined above), the respective directions were not assigned. There are, however, groups of reactions (e.g. quinone pool reducing/oxidizing reactions) whose common attributes could be exploited by new heuristics that specifically assign directions to such sets of reactions (cf. Tab. 2.3).

As an example for such an extension of the heuristic rules, the quinone pool converting reactions were set as irreversible such that the electrons are transferred from the reduced metabolites to the final electron acceptors. Having defined the final electron acceptors, it was possible to assign 43 reaction directions in the E. coli model. When we incorporated this rule into the assignment algorithm, in total 26 more reactions were restricted in one direction (cf. Fig. 2.4). Fourteen out of the 43 reactions had been already assigned.
2.3. Discussion

Table 2.3. Number of additional direction assignments required to eliminate remaining thermodynamically infeasible energy production.

<table>
<thead>
<tr>
<th>common attributes</th>
<th>standard procedure</th>
<th>standard procedure with consideration of final electron acceptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>quinone pool reductions</td>
<td>15</td>
<td>-</td>
</tr>
<tr>
<td>transporters</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>NTP production</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>NADH/NADPH production</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>O₂ production</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>CO₂ consumption</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>NMP synthesis</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>other</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>sum</td>
<td>59</td>
<td>39</td>
</tr>
</tbody>
</table>

by the thermodynamic facts, and the bypass analysis assigned three reactions less. In summary, 156 instead of 130 reactions could then be defined as irreversible by our systematic assignment procedure.

Figure 2.4. Comparison of the assignment where final electron acceptors are considered to the default assignment. The numbers of made direction assignments of the standard assignment procedure (A) and the assignment procedure, which additionally drew on the direction of electron transfer within the respiratory chain (B) are compared. The numbers (1 - 2c) refer to the assignment steps depicted in Fig. 2.1, while step 0 represents the reaction directions that were assigned by the additional heuristic rule based on final electron acceptors.

The extension of heuristic rules by organism-specific knowledge obviously is an effective and effortless approach to increase the number of assigned directions. Similarly, one could define the directions of the transporters according to their function, which often
2. Systematic Reaction Direction Assignment

can be identified from stoichiometry alone (e.g. sugars are taken up by PTS systems).

2.4. Conclusions

This paper reports on a computational framework that - based on thermodynamic principles - systematically assigns reaction directionality in genome-scale stoichiometric metabolic models. We demonstrated its application on a metabolic reconstruction of *E. coli*. After having exploited all available thermodynamic data to define irreversible reactions, we drew on network topology and thermodynamic heuristics to assign further reaction directions: Energy producing cycles were extracted from the reaction network and thermodynamically infeasible reaction steps that produce high-energy from low-energy co-substrates were disabled.

The proposed direction assignment procedure has several advantages over other approaches. The group contribution method to computationally estimate the Gibbs energies of formation is associated with such large uncertainties that only five reactions could be identified as irreversible in a genome-scale model [74]. The method developed by Beard and co-workers for *ab initio* prediction of reaction directions [211] relies on the availability of all possible cycles in the metabolic network. Currently, these cannot be calculated with reasonable computational effort for genome-scale models and the method also does not completely disable thermodynamically infeasible cycling. In contrast, using our algorithm, we demonstrated that a large number of assignments could be made without laborious calculations: A total of 130 directions could be assigned automatically, which constitutes a large fraction of the direction assignments necessary to exclude thermodynamically infeasible energy production.

Along with the development of mathematical methods that employ genome-scale metabolic models, these models became valuable tools in systems biology and metabolic engineering. Here, our systematic assignment procedure can be used in the reconstruction of new models or in the revision of existing ones. Currently, large efforts are put into the automated reconstruction of such models [41, 175] and several computational tools exist that support the first steps of the reconstruction process [88, 132]. On the
contrary, the following steps towards finalizing the model - which include the definition of reaction directionalities - are still done manually. We envision that the here proposed algorithm could be a valuable part of a computational framework that assists the automated reconstruction process for genome-scale metabolic models.

2.5. Methods

Employed Software Package

All calculations were carried out employing Matlab (The MathWorks Inc., MI, USA) unless specified otherwise. Necessary input data are standard Gibbs energies of formation and physiological ranges of intracellular metabolite concentrations. As output, the algorithm generates a vector which specifies the assigned reaction directions, and in addition, creates a detailed report (in Microsoft Excel) on the respectively made assignments.

Applied Metabolic Network Model

For the *E. coli* data set, we employed the genome-scale model iJR904 [155]. This model is an elementally balanced stoichiometric network and such enabled the calculation of the reactions’ Gibbs energies. The model was slightly modified by eliminating one reaction of duplicate reaction pairs, i.e. reactions that occur twice in the original model’s list of reactions. Moreover, the artificial reaction that accounts for the cell’s maintenance requirements in the model was omitted. The model is supplied in Additional file 1.

Employed Gibbs Energies of Formation and Concentration Ranges

A prerequisite for the thermodynamic facts based direction assignment is the availability of standard Gibbs energies of formation for a large number of metabolites. With these and values for intracellular pH and ionic strength (see below), standard transformed Gibbs energies of formation specific for intracellular conditions were calculated using the software Mathematica (Wolfram Research Inc., IL, USA) and a Mathematica notebook provided on [2]. Standard transformed Gibbs energies of formation for the metabolites
involved in the pentose phosphate pathway and the shikimate pathway were added by drawing on data from the NIST database on thermodynamics of enzyme-catalyzed reactions [3] and from the literature [186, 187]. For the Gibbs energies of formation of the quinones in the model, the values of reduced and oxidized ubiquinone, which is the only quinone available in the database, were assumed respectively. Transformed Gibbs energies of formation were adjusted to *E. coli*'s intracellular pH of 7.6 [177] and ionic strength of 0.15 M [199] (cf. Additional file 2).

To reflect typical cytosolic concentrations, which lie in the µM to mM range [62], the intracellular concentrations' upper and lower bounds were by default set to 0.001 mM and 10 mM, respectively. Exceptions were made for oxygen, for which the upper limit was set to 0.1 mM to account for its low solubility, and carbon dioxide and inorganic phosphate, for which ranges from 1 to 50 mM were assumed.

### Manual Elimination of Energy Producing Cycles

The iterative and manual direction assignment to eliminate all remaining energy producing cycles was carried out as follows: To detect a thermodynamically infeasible cycling, a flux distribution was generated by means of flux balance analysis using maximal growth rate as optimization objective (cf. [55]). Shortly, such calculated flux distributions comprise the rate of each reaction such that (i) the conversion of each metabolite is balanced, and (ii) glucose (as the employed carbon source) is converted to as much biomass as possible. As the production of biomass requires energy, a part of the glucose has to be metabolized to CO\(_2\) to yield the necessary chemical energy. Energy producing cycles render the investment of glucose into energy dispensable, and essentially all glucose is converted to biomass. In this case, the calculated flux distribution comprises at least one thermodynamically infeasible energy producing cycle, and can be used to identify the reactions that make up this cycle.

Having identified these cycles, we manually defined directions for one or more reaction in the set of reactions such that the identified infeasible cycling is disabled. The employed rationales for the assignment were similar to the heuristics used in the presented
algorithm. Essentially, reactions that consume low-energy or produce high-energy co-substrates were selected. In some cases, this was not possible as also low-energy metabolites were produced or high-energy metabolites were consumed concomitantly. Then, we determined reaction directions according to the metabolic function of the respective enzyme.

**Calculation of Sets of Correlated Reactions**

Two reactions are correlated if the ratio of their reaction rates is identical under any conceivable condition. To identify sets of correlated reactions, in a first step, the stoichiometric matrix was extended by exchange reactions and a reaction describing biomass formation to determine mass balancing sets of reaction rates, i.e. flux distributions. Here, exchange reactions were coupled to all extracellular metabolites and enabled their interchange with the environment. Next, we calculated the null space matrix for this extended stoichiometric matrix. Rows of this null space matrix that are linearly dependent indicate that the corresponding reactions are correlated. Sets of correlated reactions were determined by an all-against-all comparison of the rows.

**Authors’ Contributions**

AK and MH designed the research. AK developed the algorithm. AK and MH analyzed the results. All authors read and approved the final manuscript.

**Acknowledgements**

Authors are grateful to Daniel Beard, Lars Küpfer, Uwe Sauer, and Wolfgang Wiechert for fruitful discussions and helpful comments on the manuscript and to Jörg Stelling and Marco Terzer for also providing the Matlab routine to determine reaction correlations. All authors were funded from the ETH Zurich.
2. Systematic Reaction Direction Assignment

Supplementary Files

Supplementary files are available on http://www.biomedcentral.com/1471-2105/7/512 and on the CD provided with this document.

**Assigned Reaction Directions** Spreadsheets (Supp1_AssignedDirections.xls) listing the assigned reaction directions. It contains three sheets with (i) the list of reactions that were assigned by the standard assignment procedure, (ii) the list of reactions in case the final electron acceptors are additionally considered, and (iii) the lists of reactions that were assigned manually to obtain a thermodynamically reasonable model. In the first two sheets, we also report in which assignment step a direction was defined by our algorithm.

**Gibbs Energies of Formation** Spreadsheet (Supp2_GibbsEnergies.xls) listing the applied Gibbs energies of formation. It contains the list of the model’s metabolites and - if available - the respective transformed Gibbs energy of formation at a pH of 7.6 and an ionic strength of 0.15 M.
3. Putative Regulatory Sites Unraveled by Network-embedded Thermodynamic Analysis of Metabolome Data

Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data

*Kümmel A., Panke S., Heinemann M.*

*Molecular Systems Biology, 2, 2006*
Abstract

As one of the most recent members of the omics family, large-scale quantitative metabolomics data are currently complementing our systems biology data pool and offer the chance to integrate the metabolite level into the functional analysis of cellular networks. Network-embedded thermodynamic analysis (NET analysis) is presented as a framework for mechanistic and model-based analysis of these data. By exploiting the second law of thermodynamics and the metabolites’ Gibbs energies of formation, NET analysis allows inferring functional principles and identifies reactions that are subject to active allosteric or genetic regulation from the coupling of metabolome data and an operating metabolic network as exemplified with quantitative metabolome data from *E. coli* and *S. cerevisiae*. Moreover, the optimization framework of NET analysis demonstrated to be a valuable tool to systematically investigate data sets for consistency, for the extension of sub-omic metabolome data sets, and for resolving intracompartmental concentrations from cell-averaged metabolome data. Without requiring any kind of kinetic modeling, NET analysis represents a perfectly scalable and unbiased approach to uncover insights from quantitative metabolome data.

3.1. Introduction

Ultimately, systems biology strives to gain a quantitative systems-level understanding of complex and highly interrelated cellular processes and phenomena. The various interactions between the cellular domains and the mere number of the components involved, however, represent a complexity beyond intuitive comprehension. For this reason, mathematical models are required as tools to integrate the ever increasing biological knowledge and the data originating from the diverse cellular domains and, in a further step, to infer novel insight from the integrated available knowledge and data [94, 181].

In order to fully exploit the wealth of information contained in genome-scale data, the mathematical model to be used to extract insight from the data should ideally have the
same dimensionality. Unfortunately, the only existing class of genome-scale models are stoichiometric models [22], whose development was pioneered by Palsson [154]. As these models only reflect the metabolic capabilities of an organism, today’s basis for model-based mechanistic integration and analysis of genome-scale data is rather limited.

Nevertheless, already the stoichiometric models demonstrated to be valuable tools for integration and analysis of a number of different omics data sets, such as fluxome data [18], high-throughput growth phenotyping data [40] and transcriptome data [40, 138]. Due to the development of affordable and powerful mass spectrometers, large-scale sets of quantitative metabolite data are currently emerging into the area of systems biology [65, 131], and it is desired to also integrate these data in order to infer novel insight [131, 165, 183].

A natural approach for model-based analysis of metabolome data would be the extension of stoichiometric models by kinetic rate expressions for each enzymatic reaction. However, there is no comprehensive knowledge about in vivo reaction mechanisms and parameters. In addition, the continuing challenges in the area of measurement [106, 210] and computational analysis [200] make it very unlikely that large-scale kinetic models will be available in the near future. For these reasons, large-scale sets of metabolome data cannot yet be assimilated into mathematical models [131]. Consequently insight, for instance into underlying regulatory mechanisms, can hardly be inferred.

In attempts to deal with the lack of detailed knowledge on parameters and mechanisms, fundamental thermodynamic principles have been increasingly applied in systems biology and metabolic engineering. In this regard, the second law of thermodynamics was incorporated into stoichiometric models [13, 147, 211] or Gibbs energies of reaction were used to analyze metabolic pathways or small networks [14, 74, 122, 146].

In this work, we also draw on these fundamental principles and present a computational thermodynamics-based framework for the analysis of quantitative metabolome data. The mapping of such data onto a stoichiometric reaction network allows extraction of novel insight without requiring any kind of kinetic modeling. In the proposed network-embedded thermodynamic analysis (NET analysis), large-scale qualitative intracellular
3. Network-embedded Thermodynamic Analysis

Fluxes derived from metabolic flux analysis and metabolite concentrations are coupled to each other via the second law of thermodynamics and the metabolites’ Gibbs energies of formation, and an optimization algorithm is employed to compute network-constrained, feasible ranges of Gibbs energies of reaction.

After illustrating the novel concept, first, we will apply the NET analysis tool to a small set of quantitative metabolite data acquired from E. coli to demonstrate (i) the practical application of the optimization framework as a tool for consistency analysis of measured metabolite concentrations, (ii) for prediction of metabolite concentrations beyond the actually taken measurements and (iii) for identification of putative active sites of genetic or allosteric regulation. Then, we analyze a larger data set obtained from S. cerevisiae in order to illustrate that the method is also applicable to more complex systems such as organisms with subcellular structures.

3.2. Network-embedded Thermodynamic Analysis

Assuming constant pressure and a closed system, according to the second law of thermodynamics a reaction occurs only in the direction of negative Gibbs energy of reaction, \( \Delta_r G \). This can be expressed in the inequalities,

\[
\Delta_r G' < 0 \quad \forall \ r > 0 \\
\Delta_r G' > 0 \quad \forall \ r < 0
\]

(3.1)

where \( r \) denotes the reaction rate, or - in other words - the net flux between metabolites participating in a reaction. Here, a negative reaction rate signifies a flux in backward direction.

The Gibbs energy of a reaction \( j \) can be calculated from the Gibbs energies of formation of the participating reactants \( i \), \( \Delta_f G_i \), and the reactants’ stoichiometric coefficients in the reaction \( j \), \( s_{ij} \),

\[
\Delta_r G_j = \sum_i s_{ij} \Delta_f G_i.
\]

(3.2)

In turn, a metabolite’s Gibbs energy of formation can be calculated from its standard Gibbs energy and its thermodynamic activity. In biochemistry, thermodynamic activ-
3.2. Network-embedded Thermodynamic Analysis

Activities are typically replaced by molar concentrations, while the effect of ionic strength is taken into account by an adequate standard Gibbs energy. Moreover, possible reactant dissociation forms are lumped into a single reactant, and thus transformed Gibbs energies of formation, $\Delta_f G'_i$, are used, which in turn are calculated from the standard transformed Gibbs energies of formation, $\Delta_f G'^0_i$, and the concentration $c_i$ of the particular metabolite $i$ [7],

$$\Delta_f G'_i = \Delta_f G'^0_i + RT \ln(c_i).$$

(3.3)

For simplicity, ‘transformed Gibbs energies’ will only be referred to as ‘Gibbs energies’ in the following.

The presented equations form the foundation for NET analysis. Metabolite concentrations and metabolic fluxes are linked via thermodynamics and the stoichiometric network. The metabolites’ Gibbs energies of formation determine, together with the stoichiometry, the Gibbs energies of reaction. These Gibbs energies of reaction and the flux directions are then coupled via the second law of thermodynamics to identify thermodynamically feasible ranges for the Gibbs energies of reaction and for concentrations of non-measured metabolites. Fig. 3.1 provides an illustration of the input data required for NET analysis (i.e. metabolite concentrations, flux directions, a metabolic network model and Gibbs energies of formation) and the various insights, which can be obtained.

In the NET analysis, as an extension to an earlier employed method [122, 146, 182], the Gibbs energies of reaction are constrained by the mutual thermodynamic interdependencies of reactions in a network (i.e. the reactions’ simultaneous action in the network). This way metabolite concentrations have to be feasible not only in view of one specific reaction but with respect to the functioning of the entire network (cf. Fig. 3.2). As we will show below, this significantly limits the feasible ranges of Gibbs energies of reaction and therefore also the feasible concentration ranges of unmeasured metabolites.

In the following, we will outline the optimization procedure that underlies the NET analysis. The optimization determines the feasible range (i.e. upper and lower bounds) of the Gibbs energy of a particular reaction $k$, $\Delta_r G'_k$, using metabolite concentrations,
3. Network-embedded Thermodynamic Analysis

![Diagram of network-embedded thermodynamic analysis (NET analysis)]

**Figure 3.1.** Illustration of the network-embedded thermodynamic analysis (NET analysis). a. b. c. denote alternatives, white and gray boxes are actions and inputs, respectively, and ellipses indicate outputs.

reaction directionalities, a metabolic network and thermodynamic data:

\[
\begin{align*}
\min / \max \ & \Delta_r G_k' \\
\text{s.t.} \ & \Delta_r G_j' < 0 \quad \forall \ r_j > 0 \quad (a) \\
\ & \Delta_r G_j' > 0 \quad \forall \ r_j < 0 \quad (b) \\
\ & \Delta_r G_j' = \sum_i s_{ij} \Delta_f G_i' \quad (c) \\
\ & \Delta_f G_i' = \Delta_f G_i^0 + RT \ln(c_i) \quad (d) \\
\ & c_i^\text{min} \leq c_i \leq c_i^\text{max} \quad (e)
\end{align*}
\]
3.2. Network-embedded Thermodynamic Analysis

![Diagram](image)

**Figure 3.2.** Illustration of the mutual thermodynamic interdependencies of reactions in a network. The presented sample network comprises the reactants $A$, $B$, $C$, and $D$, for which only ranges of concentrations are known. Possible ranges for the reactants’ Gibbs energies of formation, only taking into account these concentration ranges, are shown with confined vertical bars. Due to the provided flux directions and the co-operative action of the reactions in the network, however, the thermodynamically feasible ranges are smaller, which is highlighted by the bold parts of the bars. A flux can only flow from a higher to a lower level of Gibbs energy of formation. Thus, the planes indicating the lower and upper bounds of the thermodynamically feasible Gibbs energies of formation are not allowed to incline against the direction of the flow. The space between the displayed planes, which is defined by the thermodynamic and network-derived constraints (Eqs. 3.1-3.3), consists of the thermodynamically feasible Gibbs energies of formation and, thus, describes the feasible concentration space.

In this optimization, the Gibbs energy for the reaction $k$ is minimized and maximized under the following constraints: All reactions $j$ in the network (including the considered reaction $k$) can only proceed in direction of a negative Gibbs energy of reaction (Eqs. 3.4a/b). In turn, the Gibbs energies of reaction are determined by the reaction stoichiometries and the reactants’ Gibbs energies of formation (Eq. 3.4c), which are a function of the pre-determined standard Gibbs energies of formation and the metabolite concentrations (Eq. 3.4d). The latter are by default constrained to typical intracellular concentration ranges (see Material and Methods). These ranges should be defined cautiously such that they surely cover the possible variation in concentration under the considered experimental conditions. Measured concentration values are also considered with these constraints, which - due to the fact that measurement uncertainties cannot
be excluded - are typically allowed to vary by 10% around the measured values.

With an analogous procedure, we can determine feasible concentration ranges rather than Gibbs energies of reaction.

As mentioned above, a prerequisite for the NET analysis is the knowledge of the directions of intracellular fluxes (cf. Fig. 3.1). These can either be defined based on pre-existing knowledge (a.), determined from experimental data (b.), or (in case experimental flux data are not available) computed using flux balance analysis (FBA), (c.). FBA employs linear programming to optimize a suitable cellular objective, while assuming steady-state for the mass balances [91, 148].

In the context of assigning flux directions, it is important to note that one does not necessarily need to provide directions for all fluxes. If in doubt about a certain flux, no direction should be assigned to the particular reaction. In consequence, fewer constraints are imposed on the NET analysis optimization. This results in a larger solution space meaning that the computed ranges of feasible Gibbs energies of reaction and concentrations eventually become wider. Thus, neglecting a reaction can only lead to less insight from the measurement data but in no case will such an omission lead to wrong results. The same holds true for missing or unknown pathways: If a pathway is not considered, also only less insight is obtained.

3.3. Results

In the following, to first demonstrate the type of results which can be obtained from our method we will analyze a set of measured metabolite concentrations from E. coli as a representative from the prokaryotic domain. After having approved the method for the simple test case, we demonstrate its general applicability on a larger data set from S. cerevisiae.

Before presenting the actual results, the effect of embedding the reactions into the metabolic network is illustrated for the analyzed E. coli data. Possible ranges of Gibbs energies of reaction, only defined by the concentrations of the participating reactants, were compared to the thermodynamically feasible ranges calculated using the NET anal-
ysis. A comparison of the two ranges is shown in Fig. 3.3 for several reactions. From
this figure, it can be seen that only small parts of the concentration-defined ranges are
feasible, when also considering the operation of the reaction network.

![Figure 3.3](image)

**Figure 3.3.** Ranges of Gibbs energies of reaction with and without considering the respective reac-
tion's operation in the metabolic network. White bars indicate possible ranges due to the provided
concentration ranges (see Materials and Methods), while the black bars display thermodynamically
feasible ranges after introducing the constraints of the reaction network. Abbreviations: me(NADH),
malic enzyme (NADH dependent); pdh, pyruvate dehydrogenase; nadk, NAD kinase; akgdh, α-
ketoglutarate dehydrogenase; cs, citrate synthase; udh, cytosolic transhydrogenase; atps, ATPase;
fba, fructosebisphosphate aldolase; me(NADPH), malic enzyme (NADPH dependent); mdh, malate
dehydrogenase; rpi, ribose phosphate isomerase; pgk, phosphoglycerate kinase; gapd, glyceraldehyde
3-phosphate dehydrogenase; tpi, triosephosphate isomerase; pgm, phosphoglycerate mutase; eno,
enolase

For example, the reactions pgk, gapd, pgm, eno, rpi, and tpi are restricted to a range of
Gibbs energy of reaction close to zero when the network operation is taken into account.
The reactions udh and fba do not operate at equilibrium. Had these two reactions been
3. Network-embedded Thermodynamic Analysis

analyzed separately from the network, the employed concentration limits would have allowed for that (cf. Fig. 3.2). These examples demonstrate that in cases where only a limited set of metabolite concentrations is available, the NET analysis-based integration of the reactions leads to narrower feasible ranges of Gibbs energies of reaction, and thus to a more precise classification of the reactions.

3.3.1. Thermodynamic Consistency of Metabolome Data

Despite the recent tremendous advances in mass spectrometry, small dynamic ranges of these instruments and effects such as ion suppression still impose serious difficulties on reliable quantification, and consequently quantitative metabolomics still requires sophisticated means to circumvent these problems [106, 210]. Testing for thermodynamic feasibility of a data set can constitute a part of the necessary assessment of a measurement’s reliability.

Thermodynamically feasible sets of measured concentrations are a prerequisite for the NET analysis, as the measured concentrations need to allow for negative Gibbs energies of reaction for each preset flux, so that a solution for the optimization problem (Eqs. 3.4) exists. Here, we demonstrate that NET analysis can indeed be used as a tool to determine thermodynamic consistency of the data.

For the E. coli data set we analyzed, consistency was approved. In contrast, employing the NET analysis to other published quantitative E. coli metabolome data, we found that out of seven available data sets, only four were thermodynamically feasible [6, 57, 111, 142]. In case a data set is classified as infeasible, it is furthermore possible to identify the erroneous concentration combinations by systematically omitting measured concentrations from the data set until a feasible solution for the optimization problem is established. The measured adenylate energy charge and NADH/NAD ratio of the data set in Buchholz et al. [31] lay outside physiological ranges [17, 141]. The NET analysis revealed further that in two of the three infeasible cases flux through glycolysis was inhibited by too high a F6P/G6P\(^1\) ratio [31] and by too high a PEP/F16P ratio\(^1\).

\(^1\)Explanations for the metabolite abbreviations are provided in the legend of Fig. 4.
3.3. Results

(overarching the lower part of glycolysis) [38]. For the last one [79], the OXA/MAL, the G3P/DHAP and the R5P/RU5P ratios were too high so that the measurements were not consistent with the assumed fluxes through the tricarboxylic acid cycle, glycolysis, and pentose phosphate pathway.

Quantitative metabolome data are now used in the area of mathematical modeling. Thus, it is important that only quantitatively correct data sets are employed in these endeavors in order to prevent erroneous conclusions. For these reasons, also the Yeast Systems Biology Network working group (http://www.ysbn.org) recently stated that it would be desirable to have a tool for quality control check, that is easily applicable to quantitative metabolome data sets before their actual use in modeling efforts or their entry into databases. Due to its inherent control of thermodynamic feasibility, NET analysis can fulfill parts of the quality control check.

It is important to note that unknown pathways missing in the used metabolic network as well as unknown and thus not specified flux directions, or unavailable thermodynamic data for certain metabolites will never render a data set infeasible, and thus NET analysis is rather conservative. Further, thermodynamic feasibility is a necessary but not a sufficient condition for correct quantification of metabolite concentrations. However, NET analysis as an easy-to-apply tool can test for major experimental errors, while relying only on indisputable (i.e. thermodynamic) facts.

3.3.2. Prediction of Metabolite Concentrations

In a next step, we tested whether NET analysis can also be used to predict concentrations of unmeasured metabolites. In the E. coli data set we analyzed, besides a few actually measured metabolites, the measurements provided only pooled concentrations of the isobaric molecules F6P/G6P, DHAP/G3P, 3PG/2PG and R5P/RU5P/X5P. With the NET analysis, however, it was possible to specify the concentrations of the individual metabolites (see Fig. 3.4). Furthermore, narrow concentration ranges for the unmeasured metabolites 13DPG, 3PHP, and AMP could also be calculated.

This demonstrates that it is indeed possible to use the NET analysis to specify narrow
3. Network-embedded Thermodynamic Analysis

Figure 3.4. NET analysis of the E. coli metabolome data. Summation of the inferred concentration ranges of the pooled metabolites (e.g. 2PG and 3PG) can exceed the actually measured (pooled) concentration of these metabolites. The reason for this is the introduced uncertainty in the measured concentrations of these pooled metabolites (in order to account for potential measurement uncertainties). Abbreviations: pgm, glucose 6-phosphate isomerase; pfk phosphofructose kinase; pyk, pyruvate kinase; g6pdh, glucose 6-phosphate dehydrogenase; pgl, phosphogluconolactonase; gnd, phosphogluconate dehydrogenase; rpe, ribulose-phosphate epimerase; ttk1, transketolase; ttk2, transketolase; tala, transaldolase; edd, phosphogluconate dehydratase; eda, phosphogluconate aldolase; acon, aconitase; icdh, isocitrate dehydrogenase; succoa, succinyl-CoA synthase; sucdh, succinate dehydrogenase; fum, fumarase; icl, isocitrate lyase; mals, malate synthase; ppc, phosphoenolpyruvate carboxylase; ppck, phosphoenolpyruvate carboxykinase; pgmt, phosphoglucomutase; g3pd, glycerol 3-phosphate dehydrogenase; pgcd, phosphoglycerate dehydrogenase; sulr, sulfite reductase; G6P, glucose 6-phosphate; F6P, fructose 6-phosphate; F16P, fructose 1,6-bisphosphate; DHAP, dihydroxyacetone phosphate; G3P, glyceraldehyde 3-phosphate; 13DPG, 1,3-diphosphoglycerate; 3PG, 3-phosphoglycerate; 2PG, 2-phosphoglycerate; PEP, phosphoenolpyruvate; PYR, pyruvate; 6PGL, 6-phosphogluconolactone; 6PGC, 6-phosphogluconate; RUP5, ribulose 5-phosphate; R5P, ribose 5-phosphate; X5P, xylulose 5-phosphate; S7P, seduheptulose 7-phosphate; E4P, erythrose 4-phosphate; 2DDG6P, dehydroxy-6-phosphogluconate; ACCOA, acetyl-CoA; CIT, citrate; ICIT, isocitrate; AKG, α-ketoglutarate; SUCCOA, succinyl-CoA; SUCC, succinate; FUM, fumarate; MAL, malate; OXA, oxaloacetate; G1P, glucose 1-phosphate; GLY3P, glyceral 3-phosphate; 3PH, 3-phosphohydroxypyruvate; SO₃⁻, sulfite; H₂S, hydrogen sulfite.
3.3. Results

concentration ranges for unmeasured metabolites. This kind of prediction is particularly useful for metabolites that are difficult to resolve experimentally (e.g. isobaric molecules such as G6P/F6P). Instead of putting efforts into experimental resolution, this could also be achieved by a subsequent NET analysis. As we will show below with the \textit{S. cerevisiae} data set, in a similar manner NET analysis is even able to resolve compartmental concentrations from cell-averaged metabolome data.

### 3.3.3. Prediction of Potential Active Regulatory Sites

Measured metabolite concentrations alone hardly provide any insights into the organization of metabolism, i.e. the regulatory structure responsible for routing of matter via the different metabolic pathways, the result of which is a certain intracellular flux distribution. A flux distribution is established by the fact that in comparison with the neighboring reactions the rates of some reactions are limited by the available catalytic activity (either due to low enzyme concentration or activity), so that at branch points, mass flux is accordingly distributed into the possible pathways. A limited catalytic activity of a reaction manifests itself in a large Gibbs energy of reaction. By the NET analysis, such reactions can be identified even on the basis of incomplete metabolome data via the consideration of flux directions and of the simultaneous action of the reactions in the network.

The relationship between the thermodynamic operational mode of a reaction (far from or close to equilibrium) and its regulation was established in the area of metabolic flux control analysis: Reactions operating near equilibrium usually have a large sensitivity of the reaction rate towards variations in metabolite concentrations, and thus, such reactions have a small determining effect on the flux through the respective pathway \[130, 198\]. On the other hand, reactions with large values of Gibbs energy of reaction do not necessarily have an impact on flux control as also other parameters of the enzymatic rate expressions are influencing the flux. However, it was found that reactions operating far from equilibrium are more likely to impose flux control \[202\], and it is assumed that the corresponding enzymes are more likely to be regulated by the cell \[42\] as only in
3. Network-embedded Thermodynamic Analysis

these cases flux and thus phenotype can be affected.

The NET analysis, using the *E. coli* metabolome data, classified reactions under the applied experimental conditions whether they are potential active regulatory sites or not (Fig. 3.4). A first look at these results reveals that the pfk and pyk reactions are operating far from equilibrium and thus may represent regulatory enzymes of glycolysis, while most other glycolytic reactions display only small absolute Gibbs energies of reaction. As this assignment of regulatory enzymes is in perfect agreement with earlier studies (cf. [162] for review) and as also most of the other findings displayed in Fig. 3.4 comply with our current knowledge (cf. discussion in Supplement 4), this indicates that NET analysis of metabolome data is indeed able to provide correct regulatory insight.

Another finding obtained from the *E. coli* data is related to *E. coli*’s cytoplasmic transhydrogenase (*udh*): In glucose-limited continuous cultivations, compared to the biosynthetic demands, an excess of NADPH is produced [128]. In order to eliminate the excess NADPH, the *udh*-transhydrogenase converts NADPH into NADH. As revealed by NET analysis, the *udh* reaction operates far from equilibrium in the considered experiment signifying that this reaction may be subject to regulation (Fig. 3.4). By a sensitivity analysis, we found that a regulatory control of the *udh*-transhydrogenase is indeed required for physiological reasons: Further equilibration of the *udh* reactants, corresponding to a shift of the NAD(H) pool to the reduced state, would render the normal operation of some catabolic (i.e. NAD-dependent) dehydrogenases infeasible. Of the reactions considered in this work, glyceraldehyde 3-phosphate dehydrogenase is the most critical, because the conversion of G3P to 13DPG would stop at NADH/NAD\(^+\) ratios above 0.018.

The identification of the regulatory action being effective on the *udh*-transhydrogenase and of its underlying design principle underpins the value and power of the NET analysis in uncovering system properties. First, only by using the NET analysis based sensitivity studies it was possible to classify the *udh* reaction as operating far from equilibrium (cf. Fig. 3.3). Second, with NET analysis we were able to uncover the reasons underlying the active *udh* regulation: They root in reaction interdependencies which overarch
3.3. Results

Various parts of metabolism. Both findings could not have been obtained by solely considering of the available metabolite concentrations.

3.3.4. General Applicability of NET Analysis

So far, the conceptual idea and application of the NET analysis was demonstrated for \textit{E. coli} as a simple prokaryote. In a next step, we wanted to test the method’s applicability to more complex systems such as eukaryotes with subcellular structures. For this, we extended the method and analyzed the largest available set of quantitative data from \textit{S. cerevisiae} (cf. Tab. 3.1) [118].

<table>
<thead>
<tr>
<th>metabolite</th>
<th>concentration [mM]</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKG</td>
<td>0.12</td>
</tr>
<tr>
<td>F16P</td>
<td>0.33</td>
</tr>
<tr>
<td>F6P</td>
<td>0.82</td>
</tr>
<tr>
<td>FUM</td>
<td>0.10</td>
</tr>
<tr>
<td>G1P</td>
<td>0.61</td>
</tr>
<tr>
<td>G6P</td>
<td>3.85</td>
</tr>
<tr>
<td>GLX</td>
<td>0.02</td>
</tr>
<tr>
<td>MAL</td>
<td>0.65</td>
</tr>
<tr>
<td>PEP</td>
<td>1.97</td>
</tr>
<tr>
<td>PYR</td>
<td>0.11</td>
</tr>
<tr>
<td>SUCC</td>
<td>0.10</td>
</tr>
<tr>
<td>TRE</td>
<td>201.72</td>
</tr>
<tr>
<td>ATP</td>
<td>14.12</td>
</tr>
<tr>
<td>ADP</td>
<td>6.84</td>
</tr>
<tr>
<td>AMP</td>
<td>1.29</td>
</tr>
<tr>
<td>ALA</td>
<td>7.44</td>
</tr>
<tr>
<td>ASP</td>
<td>4.14</td>
</tr>
<tr>
<td>GLU</td>
<td>41.92</td>
</tr>
<tr>
<td>2PG + 3PG</td>
<td>2.18</td>
</tr>
<tr>
<td>CIT + ICIT</td>
<td>8.00</td>
</tr>
<tr>
<td>NAD$^+$ + NADH</td>
<td>8.58</td>
</tr>
</tbody>
</table>

First, using NET analysis we could ascertain the thermodynamic feasibility of the measured metabolite data. Also with this data set, based on the computed feasible
Gibbs energies of reaction, reactions in central carbon metabolism (e.g. pfk, icl or pdh) or the first steps in amino acid synthesis (e.g. asparagine transaminase or asparagine synthase) could be identified as potential active regulatory sites (cf. Supplement 5). Interestingly, NET analysis even was able to resolve compartmental differences: In contrast to mitochondria, the production of oxaloacetate via malate dehydrogenase and aspartate transaminase is thermodynamically infeasible in the cytosol as here the reactions display positive Gibbs energies of reaction (cf. Supplement 5). Obviously, these reactions are only needed under different environmental conditions.

An inherent problem with all current omics analyses of higher (i.e. compartmentalized) organisms is that the data obtained only represent averages over the whole cell. Several scientific questions, however, can be envisioned, for which a subcellular resolution of the data would be advantageous. Here, NET analysis revealed to be a tool that is capable to infer subcellular metabolite concentrations from cell-averaged metabolite data as it draws on additional information on compartmentalized network topology and flux distribution. Examples for subcellular metabolite concentrations resolved from the present yeast data are given in Table 3.2. These values demonstrate the power of NET analysis in inferring insight by an elegant, thermodynamic-based integration of data.

![Table 3.2. Resolved intracompartmental metabolite concentrations for the S. cerevisiae data](image)

<table>
<thead>
<tr>
<th>metabolite</th>
<th>cytosol [mM]</th>
<th>mitochondria [mM]</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-aspartate</td>
<td>9.1..10</td>
<td>&lt; 0.07</td>
</tr>
<tr>
<td>fumarate</td>
<td>&lt; 0.23</td>
<td>&gt; 0.3</td>
</tr>
<tr>
<td>L-malate</td>
<td>&lt; 1.6</td>
<td>1.56..7.15</td>
</tr>
<tr>
<td>oxaloacetate†</td>
<td>&gt; 1.56</td>
<td>&lt; 0.004</td>
</tr>
<tr>
<td>phosphoenolpyruvate</td>
<td>2.21..4.21</td>
<td>&gt; 3</td>
</tr>
<tr>
<td>pyruvate</td>
<td>0.27..0.35</td>
<td>&lt; 0.07</td>
</tr>
<tr>
<td>L-threonine†</td>
<td>0.004..2.1</td>
<td>&lt; 0.23</td>
</tr>
</tbody>
</table>

† non-measured; for the others a cell-averaged measurement value was available

Finally, we will demonstrate that NET analysis of metabolome data can also assist to comprehensively analyze functional relationships in *S. cerevisiae*’s metabolism that range over different compartments. NADH metabolism in yeast is separated between the cytosol and mitochondria: NADH produced in the cytosol (e.g. in glycolysis) has
to be re-oxidized via the respiratory chain situated in the mitochondrial membrane. As pyrimidine nucleotides cannot cross this membrane, yeast can use an NADH:ubiquinone oxidoreductase external to the mitochondrial membrane to reoxidize its cytosolic NADH. Alternatively, yeast has several redox shuttle systems which - by means of other metabolites - virtually transfer the protons into the mitochondria for subsequent oxidation of reduced metabolites like NADH or FADH$_2$ [9, 158]. To date, the true physiological role of these shuttle pathways is not fully understood and thus still subject of current research [37, 134]. In this context, it was recently stated that here indeed the resolution of compartmental concentrations could provide insight as whole-cell analyses are unable to contribute to an understanding of these processes, which - in total - have the same net outcome [9].

To investigate whether the four known shuttles [9] and the external NADH:ubiquinone oxidoreductase can be operational under the considered experimental conditions, we tested each of these options for their thermodynamic feasibility. On the basis of the present S. cerevisiae metabolite data set, NET analysis revealed that under the respective experimental conditions NADH reoxidation can only occur via the external NADH:ubiquinone oxidoreductase or the glycerol 3-phosphate shuttle while the other three shuttles (ethanol-acetaldehyde shuttle, malate-oxaloacetate shuttle, malate-aspartate shuttle) cannot be operative for thermodynamic reasons. Through laborious experimental efforts employing a series of gene knockouts studies, it was found that here the external NADH:ubiquinone oxidoreductase reoxidizes the cytosolic NADH [158]. As demonstrated here, by employing a compartmentalized model NET analysis is able to uncover functional relationships from metabolome data that are related to compartmentation and which otherwise can only be obtained by elaborate experimental efforts.

3.4. Discussion

To gain insight from large-scale quantitative metabolome data, a coupling to mechanistic models is required. Integration with kinetic models will most likely remain a major challenge for at least the near future. Consequently, we present a new methodology
that does not require the derivation of sophisticated model structures but rather applies
fundamental physical laws to the interpretation of quantitative metabolite data sets. The
framework of the NET analysis maps metabolome data to the stoichiometric network
via thermodynamics and metabolic fluxes. Insight can be obtained from even limited
data sets.

NET analysis is designed for the analysis of quantitative metabolome data as the
optimization requires quantitative concentrations (in terms of mmol/l) as input, and
thus the application is limited to respective data sets. Beyond, employing NET anal-
ysis further requires a (i) metabolic network model, (ii) flux directions that reflect the
metabolic state of the organism under the analyzed experimental conditions, and (iii)
the Gibbs energies of formation for a large number of metabolites. It is important to
note that pathways missing in the metabolic network model, not completely defined flux
directions, and incomplete thermodynamic information do not lead to wrong conclu-
sions but only limit the extent of insight that can be drawn from NET analysis. Also,
measured concentrations that are averaged over cellular compartments will not lead to
wrong conclusions as e.g. wrong classification of a data set to be feasible/infeasible; NET
analysis allows the intracompartmental concentrations to be different as long as the sum
of all compartment matches the measured value. In essence, incorrect conclusions can
be avoided if solely assured information is employed for NET analysis. As shown in
this work, NET analysis is not restricted to data from simple organisms but can also be
applied to data from more complex systems such as organisms with subcellular structure.

In the NET analysis, the measurement data first undergoes a validation of thermo-
dynamic feasibility. The necessity for such a quality control is stressed by the fact that
approximately 40% of the analyzed data sets of E. coli metabolite concentrations origi-
nating from various research groups were found to be infeasible. Actually, this percentage
is even more dramatic considering that thermodynamic feasibility is only a necessary,
but not a sufficient condition for correct data. As the number of metabolites quanti-
fied from a single experiment will constantly increase in the future and the metabolic
reactions are highly interconnected, however, the chance to uncover thermodynamic in-
feasibility will increase. Thus, the NET analysis can be used as an easy-to-apply tool to test metabolome data for major experimental errors.

The ability of NET analysis to computationally resolve intracompartamental metabolite concentrations which in general are hardly accessible experimentally allows a more detailed view of an organism with subcellular structure. This capability opens up the possibility to elucidate cellular functionalities that rely on compartmentation. Furthermore, the NET analysis based resolution of pooled metabolite concentrations enables the design of more efficient analytical methods as certain pooled metabolites do not have to be separated experimentally.

The most prominent feature of NET analysis, however, is the capability to decode valuable insights from metabolome data: Besides unraveling functional relationships overarching several parts of metabolism, NET analysis is able to identify reactions most likely to be subject to active allosteric or genetic regulation on the basis of metabolome data representing a physiological snapshot at the final level in cellular hierarchy. In this context, NET analysis could be particularly helpful for screening of high-throughput metabolome data acquired from, for example, libraries of single-ORF deletion mutants for reactions, upon which the cell exerts either genetic or allosteric regulatory action under the considered experimental conditions. We envision that the superposition of this NET analysis-derived information obtained from different mutant strains should ultimately facilitate identification of novel interrelationships in metabolic regulation. Such regulatory insight would tremendously enhance our knowledge about the regulation of metabolic pathways, where, in stark contrast to the regulation of genes encoding enzymes used to introduce specific substrates into the central metabolism, only little is known.

In contrast to detailed kinetic models, the NET analysis does not provide deterministic statements. However, as it is easy to apply, perfectly scalable to the systems level, and only relies on indisputable facts, it represents a valuable tool to computationally decipher insight from quantitative large-scale metabolome data. It can be easily envisioned that it will significantly assist systems biology research and also will support more applied
3. Network-embedded Thermodynamic Analysis

fields such as metabolic engineering.

3.5. Materials and Methods

Stoichiometric Models

For the *E. coli* data set, we employed the genome-scale model iJR904 developed by Palsson and co-workers, which is an elementally balanced stoichiometric network [155]. The fully compartmentalized, elementally and charge balanced model iND750 [54] was employed to analyze the set of metabolite concentrations from *S. cerevisiae*. For our studies on the redox shuttles, transport reactions between cytosol and mitochondria were added according to Bakker et al. [9].

Thermodynamic Data

A prerequisite for the NET analysis is the availability of standard Gibbs energies of formation for a large number of metabolites. With these and values for intracellular pH and ionic strength (see below), standard transformed Gibbs energies of formation specific for intracellular conditions were calculated using the software Mathematica (Wolfram Research Inc., IL, USA) and a Mathematica notebook provided on http://library.wolfram.com/inforcenter/MathSource/797 [7]. Standard transformed Gibbs energies of formation for the metabolites involved in the pentose phosphate pathway and the shikimate pathway were added by drawing on data from the NIST database on thermodynamics of enzyme-catalyzed reactions (http://xpdb.nist.gov/enzyme_thermodynamics) and from the literature [186, 187].

In order to account for potential errors in the experimentally determined equilibrium constants (from which the transformed Gibbs energies of formation were derived), standard transformed Gibbs energies of formation were allowed to vary in the NET analysis by 0.5 kJ mol\(^{-1}\) around the reported value.Converted to the concentration domain, this represents an uncertainty of approximately 20\% for the measured equilibrium concentration of a reactant.

In summary, for 137 of the 761 metabolites in the *E. coli* model, standard transformed
Gibbs energies of formation were available, and consequently, for 154 out of the 922 biochemical reactions in the metabolic model it was possible to calculate Gibbs energies of reaction. For the \textit{S. cerevisiae} model, Gibbs energies of reaction could be calculated for 232 (out of 1149) reactions with the available Gibbs energies of formation for 132 (out of 645) metabolites.

\textbf{Analyzed Metabolome Data Set and Input Data for \textit{E. coli}}

We examined a set of measured metabolite concentrations obtained from an \textit{E. coli} (W3110) chemostat culture that was operated at a dilution rate of 0.1 h$^{-1}$. The culture was fed with M9 minimal medium containing 5 g l$^{-1}$ glucose. The reactor with a working volume of 1.5 l was aerated with 1 vvm and stirred at 800 rpm in order to ensure a dissolved oxygen concentration above 60%. The concentrations of six metabolites (F13P, PEP, PYR, 6PGL, ADP, ATP) and four groups of pooled metabolites (G6P/F6P, DHAP/G3P, 2PG/3PG, R5P/RU5P/X5P) were measured on an ion-chromatography/single quadrupole mass spectrometer system (Dionex, Thermo Finnigan) following a novel integrated sampling procedure [168].

Due to the limited number of measured metabolites in the analyzed data set, from a broad literature survey metabolite data obtained from aerobic \textit{E. coli} growth experiments on glucose under various conditions (batch, chemostat and substrate pulse experiments) were gathered, and based on these data the concentration ranges of 29 metabolites were further restricted (see Supplement 2) assuming that these ranges reflect the typical variance of metabolite homeostasis. In the NET analysis, NAD(H) and NADP(H) were considered with their ratios, NADH/NAD$^+$ and NADPH/NADP$^+$, and the adenylate nucleotides, ATP, ADP and AMP, as adenylate energy charge (AEC, see Supplement 1 for definition). Upper and lower bounds for these ratios were also defined based on respective values obtained from the literature survey (see Supplement 2). Beyond, the concentrations of all other non-measured metabolites were - by default - restricted to ranges that reflect the typical levels in the cytoplasm, which usually lie within the $\mu$M to mM range [62]. Thus, for the NET analysis, default minimal and maximal concentration
limits were set to 0.001 and 10 mM, respectively.

The standard Gibbs energies of formation were calculated for an intracellular pH of 7.6 [177] and an ionic strength of 0.15 M [199] to reflect *E. coli*'s intracellular conditions. The energy required for transfer of protons from the cytosol to the extracellular environment (or energy gained in the case of reflux) was estimated using an experimentally determined proton motive force of −150 mV [159]. Multiplication of this value with the Faraday constant provided the respective Gibbs energy for proton translocation (−15 kJ mol\(^{-1}\)).

**Analyzed Metabolome Data Set and Input Data for *S. cerevisiae***

Quantified metabolite concentrations of *S. cerevisiae* CEN.PK 113-7D were obtained from an aerobic, glucose/ethanol-limited chemostat culture at a dilution rate of 0.052 h\(^{-1}\) [118]. Measurements were performed for 18 metabolites (AKG, F16P, F6P, FUM, G1P, G6P, GLX (glyoxylate), MAL, PEP, PYR, SUCC, TRE (trehalose), ATP, ADP, AMP, ALA (L-alanine), ASP (L-aspartate), GLU (L-glutamate)) and three groups of pooled metabolites (2PG/3PG, CIT/ICIT, NADH/NAD\(^+\)) either by mass spectroscopy on a LC-ESI-MS/MS system, by IE-HPLC or enzymatically. We analyzed the metabolite data obtained after eight generation cycles. By default, the measured metabolite concentrations were considered as average concentrations over all intracellular compartments using compartmental volume fractions given in Supplement 3. If, however, according to the metabolic model a metabolite could not enter a particular compartment by diffusion and was not participating in any active reaction, its concentration was assumed to be negligible in this compartment under the assumption that available transporters are not expressed. Analogous to the analysis of the *E. coli* data, default ranges that reflect physiological conditions were set for the non-measured metabolite concentrations, and ratios of pyridine nucleotides and the AEC (see Supplement 3).

A cytosolic and mitochondrial pH of 7 and 7.5, respectively, was assumed (J J Heijnen, personal communication, 2006), while an ionic strength of 0.15 M was used. To translocate ions against the mitochondrial membrane potential of −180 mV [112], the necessary energy was considered in the calculation of the respective Gibbs energies of
3.5. Materials and Methods

Determination of Flux Directions

Experimentally determined fluxes for *E. coli* were obtained from Nanchen et al. [128] and for *S. cerevisiae* from Daran-Lapujade et al. [45] and Wu et al. [209]. To obtain a flux distribution (beyond the actually measured set of fluxes) that we used as input data, we employed the optimization approach introduced by Blank et al. [18]:

\[
\min \sum |r_i|
\]

\[\text{s.t.} \quad \sum_i s_{il} r_i = 0 \quad (a)\]

\[r_m^\text{exp} \cdot (1 - \delta) \leq r_m \leq r_m^\text{exp} \cdot (1 + \delta) \quad (b)\]

\[r_i^\text{min} \leq r_i \leq r_i^\text{max} \quad (c)\]

In this optimization, the L1 norm of all fluxes \( r_i \) (including the intracellular fluxes \( r_j \) as well as the exchange fluxes with the extracellular environment) is minimized under the constraint that the mass balances hold (Eq. 3.5a; \( s_{il} \) are the stoichiometric coefficients including those of the intracellular reactions \( s_{ij} \), of the exchange fluxes and of a pseudo-reaction describing biomass synthesis). The fluxes \( r_m \), for which experimental data are available, have to lie within a range \( \delta \) (here: 10%) around the measured values \( r_m^\text{exp} \) (Eq. 3.5b). \( r_i^\text{min} \) and \( r_i^\text{max} \) are default flux boundaries defining the reversibility/irreversibility of the intracellular and exchange fluxes (Eq. 3.5c).

Optimization

The optimization problems were solved using LINDO API (LINDO Systems Inc., IL, USA) via the Matlab (The MathWorks Inc., MI, USA) interface. The actual formulations of the employed optimization procedures are presented in Supplement 1.

Acknowledgements

The authors would like to thank Matthias Reuss (University of Stuttgart) for sharing data before publication and Robert Albery (MIT), Sef Heijnen (TU Delft), Uwe Sauer
3. Network-embedded Thermodynamic Analysis

(ETH Zurich), and Jörg Stelling (ETH Zurich) for helpful discussions.

Supplementary Files

Supplementary files are available on http://www.nature.com/msb/journal/v2/n1/-suppinfo/msb4100074_S1.html and on the CD provided with this document.

Optimization Complete formulation of the optimization problem (Supplement_1.pdf).

Default Concentrations E. coli Table (Supplement_2.pdf) on the applied default concentrations for E. coli.

Inputs S. cerevisiae Tables (Supplement_3.pdf) containing the measured concentrations and applied compartmental volumes for S. cerevisiae.

Putative Regulatory Sites E. coli Discussion (Supplement_4.pdf) on putative regulatory sites in E. coli as suggested here in context of the corresponding literature.

Thermodynamically Feasible Gibbs Energies of Reaction for S. cerevisiae

Putative regulatory sites and operational compartmental differences identified for S. cerevisiae (Supplement_5.pdf).
4. Metabolomics-based Approach for Elucidating Reaction Activities in Complex Metabolic Network

Contributors:
Kümmel A. (leading author) and Heinemann M.
Abstract

We present a novel metabolomics-based approach for inferring metabolic reaction activities based on mass balances and thermodynamic constraints. By combining elementary flux mode (EFM) analysis and network-embedded thermodynamic analysis, we identified reaction activities within the pyruvate/ethanol metabolism of either glucose or ethanol grown *Saccharomyces cerevisiae* from measured metabolite concentrations. All EFMs obtained from a compartmentalized stoichiometric metabolic model were tested for thermodynamic feasibility considering measured metabolite concentrations and the second law of thermodynamics. Such, the activity of several reactions operating in a defined direction could be excluded or approved. To conclude, we demonstrate a universally applicable novel method that integrates thermodynamic with systematic mass balance analysis to reduce the flux solution space for highly flexible metabolic networks.

4.1. Introduction

Elucidating intracellular metabolic fluxes that are closely linked to the phenotype is important for gaining systems understanding of cellular processes. $^{13}$C-based flux analysis is a powerful tool to determine intracellular fluxes in central carbon metabolism of microbes [19, 59, 63]. As such, this technique has already significantly contributed to our understanding about microbial metabolism (cf. [18, 105, 167] for examples).

This technique, however, has certain limitations: First, fluxes of alternative metabolic pathways can only be resolved if the pathways lead to different labeling patterns in the common end products, e.g. measured proteogenic or free amino acids [206]. This prerequisite is often not fulfilled if the labeled carbon substrate contains only few carbon atoms (e.g. ethanol or acetate) or if the metabolic network of interest has a high degree of redundant pathways (such as for instance the pyruvate metabolism in *Saccharomyces cerevisiae* which is even distributed over different compartments). A second limitation of current $^{13}$C-labeling methods is that they can only be applied when cells are grown on a medium containing a single or very few carbon sources and when the labeling of
the measured compounds has reached a steady-state [206]. Despite ongoing attempts to extent the applicability of $^{13}$C-based flux analysis with respect to the latter point [171, 207], not every experimental setup is accessible with this technique. Also, alternative intracellular experimental data is necessary to overcome the inherent limitation in terms of resolvability of fluxes in complex networks with substrates that harbor only few possibilities for labeling.

One such alternative type of data are quantitative metabolite concentrations. Novel mass spectrometry based techniques today provide us with absolute concentrations of most central carbon metabolites, free amino acids, and redox and energy co-factors. Quantitative metabolite concentrations can in combination with the second law of thermodynamics be used to elucidate flux directions and such contribute to flux analysis [102]. In fact, metabolite concentrations and thermodynamic principles were recently integrated in the computational, constraints-based approach of flux balance analysis (FBA) [73, 78]. The validity of the fluxes predicted with these tools, however, strongly depends on an assumption of a cellular objective (e.g. maximization of biomass yield) to which the organism has adopted during evolution.

Here, we suggest a novel generally applicable metabolomics-based approach for inferring reaction activities which avoids any evolution-based assumptions. Our approach integrates thermodynamic constraints based on measured metabolite concentrations with mass balances in a combination of an elementary flux mode (EFM) analysis and a network-embedded thermodynamic (NET) analysis [102, 172]. By comprehensively analyzing the resulting space of possible flux distributions, we are able to identify reactions that were either active or inactive in the respective experiment as well as we could elucidate possible reaction directions in complex metabolic networks which cannot be resolved by $^{13}$C-based flux analysis. The few requirements for this method are a stoichiometric model of the metabolic network (such as [54]) as well as measured intra- and extracellular metabolite concentrations.

To demonstrate the capabilities of our approach, we predict active reactions and reaction directions for *S. cerevisiae* growing on either glucose or ethanol. Here, we focus on a
4. Metabolomics-based Elucidation of Metabolic Activities

part of the metabolic network, which is particularly hard to resolve with $^{13}$C-based flux analysis, i.e. the connection of Embden-Meyerhof-pathway (EMP), tricarboxylic acid (TCA) cycle, and ethanol metabolism.

4.2. Approach to Predict Active and Inactive Reactions

Our metabolomics-based approach for inferring active and inactive reactions consists of two main steps as illustrated in Fig. 4.1. First, a stoichiometric model of the considered metabolic network is used to describe all mass-balanced flux distributions via EFM\textsubscript{s}. Second, measured intracellular metabolite concentrations and NET analysis are applied to identify the subset of thermodynamically feasible EFM\textsubscript{s}. As shown in the appendix, any thermodynamically feasible and thus actually possible flux distribution can be described by linear combinations of thermodynamically feasible EFM\textsubscript{s}. Hence, by analyzing the set of thermodynamically feasible EFM\textsubscript{s}, we can identify reactions that are either inactive or active in the given experiment. While inactive reactions are inferred solely from the set of thermodynamically feasible EFM\textsubscript{s}, information about measured extracellular rates such as biomass or ethanol production are exploited to identify active reactions. In the following, the two steps of our approach are described in detail.

Step I: Exploiting Mass Balance Constraints for Inference of a Preliminary Set of Active Reactions

(a) Describing Mass-balanced Flux Distributions via EFM\textsubscript{s} At steady-state where intracellular metabolite concentrations $c_j$ do not change over time, a mass-balanced flux distribution $r$ that assigns a particular metabolic rate to each reaction $i$ is described by

$$\frac{dc_j}{dt} = \sum_i S_{ji} r_i = 0.$$  

(4.1)

Here, $S$ is the stoichiometric matrix of the known metabolic network. Additionally, a priori assumptions are typically imposed on the reactions’ reversibility:

$$r_i \in \{\text{reversible}; \text{irreversible}\}.$$  

(4.2)
4.2. Approach to Predict Active and Inactive Reactions

In most cases, Eqn. 4.1 is underdetermined and an infinite number of solutions exists. The space of possible flux solutions can be described with elementary flux modes (EFM), which are derived from Eqns. 4.1 and 4.2 [172]. Any mass balanced flux distribution \( r_i \) is a weighted sum of all EFMs \( e_j \) with non-negative weighting coefficients \( \alpha_{ij} \) [171]:

\[
r_i = \sum_j \alpha_{ij} e_j .
\]

(b) Inferring Active Reactions  By drawing on measured extracellular rates a preliminary set of active reactions can be identified from the EFMs. As only mass-balance but not thermodynamic constraints are considered at this point this is just a preliminary set which will be extended in step II. The identification is done as follows: From the EFMs, we first extract subsets that contain a non-zero flux for a certain extracellular rate which was also non-zero in the experiment (i.e. that produce or consume a certain extracellular compound). As an example, for a growing organism, one such subset would contain all biomass producing EFMs. In all of these subsets, there must exist at least
4. Metabolomics-based Elucidation of Metabolic Activities

one EFM that contributes to the actual flux distribution to be consistent with experimentally observed extracellular rates. Such, a reaction that has a forward flux in all EFMs of a particular subset will also have a forward flux in the real flux distribution. Consequently, such reactions are necessarily active in the given experiment. The such identified preliminary set of active reactions is applied as additional constraints in the following test for thermodynamic feasibility.

Step II: Exploiting Thermodynamic Constraints for Inference of Active/Inactive Reactions and Restricted Directions

(a) Imposing Thermodynamic Constraints onto EFMs. According to the second law of thermodynamics, the substrates of a reaction must have a higher Gibbs energy of formation, $\Delta_f G$, than the products. In other words, the Gibbs energy of a reaction, $\Delta_r G$, which is calculated from the reactants’ $\Delta_f G$s and the reaction stoichiometry, has to be negative in the direction $d_i$ of the proceeding reaction $i$:

$$\Delta_r G_i(c) \cdot d_i < 0 .$$  \hspace{1cm} (4.4)

The Gibbs energies depend on the reactants’ concentrations $c$, and hence, measured metabolite levels impose constraints on allowed reaction directions. Such testing for an allowed operational direction can now also be performed with EFMs. An EFM contains a defined set of reaction directions, and measured metabolite concentrations can thus be used to determine the thermodynamic feasibility of an EFM. To identify the set of thermodynamically feasible EFMs, each EFM is tested for its thermodynamic feasibility with measured metabolite concentrations by employing the NET analysis framework [102]. Finally, as is shown in the appendix, any thermodynamically feasible flux distribution $r_i^{feasible}$ can be described as a linear combination of thermodynamically feasible EFMs, $e_j^{feasible}$.

$$r_i^{feasible} = \sum_j \beta_{ij} e_j^{feasible} .$$  \hspace{1cm} (4.5)
(b) Inferring Inactive and Active Reactions  By analyzing the subset of thermodynamically feasible EFMs, we can now infer inactive and (more) active reactions. Inactive reactions are those reactions that have a zero flux in all thermodynamically feasible EFMs. Active reactions are again derived by using the information about measured extracellular rates as was described above. In addition, for reversible reactions in the model possible directions can be inferred though they were not necessarily active in the given experiment. Such reactions display both forward and backward directions in the complete set of EFMs, but in the subset of thermodynamically feasible EFMs they only occur with only positive (or only negative) or zero flux values in the EFM subsets defined based on the extracellular rates.

4.3. Results

To demonstrate the suggested approach, we aimed at elucidating reaction activities in *S. cerevisiae* when grown on either glucose or ethanol. We focused particularly on the complex part of metabolism that connects glycolysis, TCA cycle, and ethanol metabolism respectively. It is characterized by many parallel pathways and a high connectivity and moreover spread over two subcellular compartments - mitochondria and cytosol. Identifying reaction activities is highly desired in this node as particularly the fluxes during growth on ethanol cannot be resolved with $^{13}$C-labeling experiments due to the limited label information this C$_2$ molecule can carry.

4.3.1. Experimental Data and Metabolic Model Applied As Inputs

The metabolite concentrations that we used for our analysis were obtained from a *S. cerevisiae* batch experiment on glucose minimal medium (cf. A.1). Here, specifically we used data from the two consecutive growth phases on glucose and ethanol, respectively. For these steady-state periods, we determined averaged intracellular concentrations and estimated constant extracellular rates from the extracellular concentrations (cf. Methods). In contrast to these readouts the extracellular concentrations do not stay
constant and we chose a time point in the midst of each growth phase from which these concentrations were used as input data for the analysis.

A further requisite for our method is a stoichiometric model of the metabolic network that is able to describe the possible flux distributions for these two growth phases. As it is yet not possible to enumerate EFMs for genome-scale stoichiometric models such as iND750 [54, 95], we constructed a smaller model using iND750 as a compendium of *S. cerevisiae*’s reactions. The built smaller model describes the central carbon metabolism in detail and considers the compartments cytosol and mitochondria. In order to adapt the model to each respective growth phase and to thus reduce the number of EFMs to be calculated, we additionally exploited biochemical knowledge and the experimentally observed extracellular rates to eliminate reactions or constrict their directionality. For instance, we used the latter information to eliminate certain exchange fluxes from the model prior to EFM enumeration (cf. Methods and Supplementary File 1 for a detailed description of the employed model). Thus we had two models each presenting one growth phase and harboring approximately 200 reactions for which we enumerated the EFMs (cf. Tab. 4.1).

**Table 4.1.** Number of reactions in the models and the respective number of EFMs for growth on glucose and ethanol. Here, variable reactions are those for which the activity is not determined based on mass balance constraints alone.

<table>
<thead>
<tr>
<th>Reactions</th>
<th>glucose</th>
<th>ethanol</th>
</tr>
</thead>
<tbody>
<tr>
<td>biochemical reactions</td>
<td>148</td>
<td>146</td>
</tr>
<tr>
<td>transporter</td>
<td>41</td>
<td>37</td>
</tr>
<tr>
<td>lumped biomass synthesis</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>exchange fluxes with environment</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>220</td>
<td>211</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EFMs</th>
<th>glucose</th>
<th>ethanol</th>
</tr>
</thead>
<tbody>
<tr>
<td>in total</td>
<td>480858</td>
<td>287262</td>
</tr>
<tr>
<td>thermodynamically feasible</td>
<td>234507</td>
<td>14078</td>
</tr>
</tbody>
</table>
4.3. Results

4.3.2. Inferred Active Reactions due to Mass Balances

As described above, drawing on measured extracellular rates, the analysis of the whole set of EFMs yields the preliminary sets of active reactions. In the glucose growth phase, besides biomass also ethanol, acetate, and pyruvate are produced, while during growth on ethanol only biomass is produced.

Here, we find that for both substrates about 75% of all reactions need to be active solely based on mass balance constraints. Nevertheless, especially in central carbon metabolism a high degree of variability still exists as most preliminary active reactions belong to the linear but not lumped pathways for amino acid biosynthesis: There are 42 (for glucose) and 41 (for ethanol) so-called variable reactions or transporters left whose activity cannot be predicted based on mass balance considerations alone (Tab. 4.1). In the next step, we exploit the measured metabolite concentrations and the such imposed thermodynamic constraints to infer reaction activities of this set of variable reactions in particular.

4.3.3. Inferred Inactive/Active Reactions and Restricted Directions from Thermodynamically Feasible EFMs

Applying NET analysis, we tested all EFMs for the glucose and the ethanol growth phase for their thermodynamic feasibility based on the measured extra- and intracellular metabolite concentrations. Here, as outlined above, we also considered the preliminary set of active reactions as constraints in order to further restrict the possible flux solution space.

Considering the measured concentrations and the preliminary set of active reactions, we found that almost 50% of the EFMs for glucose and 5% of the EFMs for ethanol were thermodynamically feasible (Tab. 4.1). This can either be due to the fact that the measured concentrations are indeed more constraining in case of ethanol. However, as two different models were applied this difference could also be a result of different assumptions used to construct the models. The analysis of these sets of thermodynamically feasible EFMs then allowed us to identify further active reactions as well as direction
restrictions for in principle reversible reactions. A necessarily inactive reaction was not identified. For growth on glucose, we could infer one additional active reaction and one was restricted into one direction, while for ethanol, five such reactions were identified and additionally three potentially reversible reactions were restricted to one direction. As the latter three reactions still can take non-zero fluxes in biomass producing EFMs, they are not necessarily active. The predicted active reactions and the direction restrictions are summarized in Tab. 4.2 and a general overview of the predicted active reactions as provided by both constraints is shown in Fig. 4.2. The findings from the thermodynamic analysis will be discussed in the following.

<table>
<thead>
<tr>
<th>reaction</th>
<th>direction</th>
<th>enzyme</th>
<th>gene</th>
<th>growth phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACS</td>
<td>ac → acCoa</td>
<td>cytosolic acetyl-CoA synthase</td>
<td>ACS1/ACS2</td>
<td>glucose, ethanol</td>
</tr>
<tr>
<td>ALDD2y</td>
<td>acald → ac</td>
<td>cytosolic acetaldehyde dehydrogenase (NADP-dependent)</td>
<td>ALD6</td>
<td>ethanol</td>
</tr>
<tr>
<td>GLUDy</td>
<td>akg → L-glu</td>
<td>glutamate dehydrogenase (NADP-dependent)</td>
<td>GDH1</td>
<td>ethanol</td>
</tr>
<tr>
<td>HSK</td>
<td>hom → phom</td>
<td>homoserine kinase</td>
<td>THR1</td>
<td>ethanol</td>
</tr>
<tr>
<td>THRS</td>
<td>phom → L-thr</td>
<td>threonine synthase</td>
<td>THR4</td>
<td>ethanol</td>
</tr>
<tr>
<td>ACOAH</td>
<td>acCoa → ac</td>
<td>acetyl-CoA hydrolase</td>
<td>ACH1</td>
<td>glucose, ethanol</td>
</tr>
<tr>
<td>AKGt2m</td>
<td>akg[c] → akg[m]</td>
<td>mitochondrial transport of α-ketoglutarate via proton symport</td>
<td>GLY1</td>
<td>ethanol</td>
</tr>
<tr>
<td>THRA</td>
<td>acald + L-thr → gly</td>
<td>threonine aldolase</td>
<td>GLY1</td>
<td>ethanol</td>
</tr>
</tbody>
</table>

### 4.3.4. Validation of Inferred Activities and Direction Restrictions Obtained from Thermodynamic Constraints

To validate the suggested approach for inferring reaction activities, we compared our predictions with information from the literature. For a reaction, which is predicted to be active, a gene deletion mutant that is unable to catalyze this particular reaction should have a phenotype that is different from the wildtype. Hence, the fact that such a mutant is growing differently or even is non-viable serves as indication that the respective reaction is indeed active in the wildtype.
Figure 4.2. Measured metabolite concentrations and the inferred reaction activities for glucose and ethanol growth phase mapped to the metabolic network. Inferred active and inactive reactions as well as reaction directions restrictions based on both mass balance and thermodynamic constraints are displayed. Predictions that were made based on the thermodynamic constraints are indicated by also showing the corresponding reaction name.
Inferred Active Reactions  The cytosolic acetyl-CoA synthesis (ACS) is predicted to be active in both growth phases. This reaction is proposed (i) to provide the lipid metabolism with cytosolic acetyl-CoA and (ii) to assimilate acetate and ethanol as carbon sources [100, 192]. Two genes - ACS1 and ACS2 - are known to encode acetyl-CoA synthases (which both can catalyze for the ACS reaction) both of which localize to the cytosol [176]. An Δacs1 acs2 double mutant was found to be not viable on both carbon sources [192], which confirms our prediction that this reaction is active.

The cytosolic NADP-dependent aldehyde dehydrogenase (ALDD2y) reaction, which is active according to our analysis, is encoded by the constitutively expressed ALD6 gene [176]. Here, a decreased growth rate of a Δald6 mutant on ethanol in comparison to the wildtype validates also this activity prediction.

Our analysis further suggested that glutamate is produced via the cytosolic glutamate dehydrogenase (GLUDy) encoded by GDH1, while also synthesis via alternative mitochondrial reactions is possible [176]. Disruption of the GDH1 gene was found to lead to a slower growth rate on ethanol compared to the wildtype [48]. Such, also this activity prediction is in-line with experimental observation.

The homoserine kinase (HSK) and threonine synthase (THRS) reaction, which we inferred to be active for growth on ethanol, are consecutive reactions in the biosynthesis path for L-threonine. In our model, the threonine aldolase reaction constitutes an alternative path to produce L-threonine. For growth on ethanol, we did not find any information about the phenotype of the corresponding gene deletion mutants growing on ethanol. Nevertheless, in the Saccharomyces Genome Database (SGD) the synthesis of threonine via HSK and THRS is the only documented biosynthetic route [176]. Such, our prediction is at least in-line with the common biochemical knowledge.

Inferred Reaction Directions  Our analysis predicted that the acetyl-CoA hydrolase (ACOAH), if it was active, only could convert acetyl-CoA to acetate and CoA and not vice versa on both ethanol and glucose. While it was shown that the corresponding gene ACH1 is repressed on glucose [109], unfortunately, no information about this reaction
4.3. Results

for growth on ethanol could be found. Nevertheless, in-line with our directionality prediction, it was suggested that for acetate utilization the enzyme is indeed hydrolyzing acetyl-CoA [109].

For the threonine aldolase (THRA) reaction, we inferred that it converts L-threonine to glycine but not vice versa. Hence, it can be used for biosynthesis of the glycine but not the L-threonine. This finding is in-line with this reaction’s assignment to glycine, but not L-threonine biosynthesis in SGD.

Finally, our analysis predicted that $\alpha$-ketoglutarate can only be exported from the mitochondria to the cytosol (AKG$t2m$). This can be interpreted as a potential excess of $\alpha$-ketoglutarate in the mitochondria in comparison to the cytosol which would at least be in-line with another prediction we made: Considering the here postulated synthesis of glutarate from $\alpha$-ketoglutarate in the cytosol (cf. GDH1 discussion above), it is conceivable that $\alpha$-ketoglutarate which is produced in the mitochondria has to be transported to the cytosol. Summarizing the validation, as all inferred activities and directions are consistent with literature data, we conclude that our novel thermodynamics-based approach produces correct results.

4.3.5. Evaluation of Additional Constraints for Growth on Glucose

As only one reaction was predicted to be necessarily active during growth on glucose, we asked whether there are any means to increase the number of predictions. For this, we evaluated whether an additional constraint on the cytosolic NADH/NAD$^+$-ratio, a readout that can be determined experimentally [33], would help. Thus, we re-calculated the set of thermodynamically EFMs by assuming physiologically reasonable ratio values (i.e. 0.1, 0.02, 0.01, and 0.002 [33] with assuming a $\pm$20\% error range). For any assumed ratio value, this additional constraint resulted in one additional prediction: The cytosolic malate dehydrogenase turned out to be inactive in these scenarios. This gluconeogenetic enzyme encoded by $MDH2$ is known to be not used in for growth on glucose [124].
4. Metabolomics-based Elucidation of Metabolic Activities

4.4. Discussion

In this work, we developed a novel, metabolomics-based framework to identify active and inactive reactions suitable for complex metabolic networks. Therefore, it can aid classical $^{13}$C-flux analysis in cases where it so far cannot be used to resolve fluxes. We demonstrated our approach by analyzing data from *S. cerevisiae* growing either on glucose or ethanol and validated our predictions with knowledge from the literature. As our approach can be applied to any steady-state growth and is in particular not limited with respect to the utilized carbon substrates, it is widely applicable.

Comparing our findings with flux distributions determined by $^{13}$C-labeling experiments on glucose [20, 105, 116] and truly computationally by flux balance analysis for ethanol [105], we could locally achieve a more detailed view on the here considered metabolic node. For instance, a potential utilization of the threonine aldolase for production of L-threonine was neglected in the models in above mentioned studies. For this reaction, we inferred that the threonine aldolase can only produce glycine from L-threonine and not vice versa. Such, our approach to assess reaction activities and directions confirms assumptions *a priori* made for the $^{13}$C-based and the purely stoichiometric flux balance analysis.

The exemplary analyzed data set contained only cell averaged concentrations, and many metabolites especially in the complex pyruvate metabolism (i.e. acetaldehyde, acetate, and ethanol) were not determined. Nevertheless, we identified several reactions that have to proceed due to thermodynamical consistency with measured intracellular concentrations. We envision that a larger number of predictions can be made if more metabolites were measured. The young field of quantitative metabolomics is rapidly developing and the number of quantifiable metabolites still increases. Here, our framework can also be used to identify most informative metabolite concentrations. For instance, in a first approach to this question, we tested hypothetical cytosolic NADH/NAD-ratios. In a more systematic approach, all metabolite concentrations could be screened within their supposed physiological ranges.

Besides the restricted number of predictions due to scarce data sets, another current
limitation of our approach relates to the computational side. Computation of EFMs was so far not feasible for genome-scale networks and especially for growth scenarios with complex media or multiple extracellular products. However, very recently, an algorithm to calculate EFMs for genome-scale networks was developed (J. Stelling, pers. communication).

Summarizing, the here developed approach can be used to reveal operative metabolic reactions in complex metabolic networks where $^{13}$C-based flux analysis meets its resolution limits.

4.5. Methods

Experimental Data

All experimental data were obtained in experiments, which are described in detail in the appendix. Shortly, *S. cerevisiae* FY4 was grown on glucose minimal medium in batch mode. After glucose depletion, the culture entered a second growth phase on ethanol. During both steady-state growth phases, samples were withdrawn from the broth to determine the time courses of the biomass concentration and the concentrations of extracellular as well as intracellular metabolites. All but the intracellular concentrations obtained by LC-MS/MS experimental data were obtained from averaging over three independent experiments.

Extracellular Rates

Uptake and production rates were determined from the monitored time courses of the biomass and extracellular metabolite concentrations. By fitting spline curves to these data (via Matlab function *spaps*), we derived time-resolved production/excretion rates. Time periods, during which these rates stayed constant and concomitantly biomass production occurred, were defined as steady-state growth phases. Such, we defined the glucose and ethanol growth phase and also determined the extracellular rates in the two phases (cf. Supplementary File 2).
Metabolomics-based Elucidation of Metabolic Activities

Metabolite Concentrations for Each Growth Phase

As the intracellular metabolite concentrations remained constant during each growth phase, we calculated concentrations for each growth phase by averaging the concentrations measured at all corresponding time points. As the extracellular concentrations changed over time, we chose one time point in the midst of each growth phase and used the corresponding concentrations in the analysis: Based on the defined periods of steady-state growth, the used concentrations for glucose, ethanol, acetate, and pyruvate were those determined for 2 h before glucose depletion (for the glucose growth phase) and 5 h after glucose depletion (for the ethanol growth phase). Concentrations for extracellular inorganic phosphate and ammonia were estimated as the difference of the initial concentration and the amount already incorporated in the current amount of biomass. This calculation was based on a biomass composition of $\text{CH}_{1.769}\text{N}_{0.146}\text{O}_{0.631}\text{P}_{0.009}\text{S}_{0.00162}\text{M}_{0.019}$ [108].

Metabolic Model

Using the genome-scale model iND750 [54] as a compendium of biochemical reactions in *S. cerevisiae*, we constructed a model that describes in detail the central carbon metabolism with two compartments, namely cytosol and mitochondria. Specifically, from the iND750 model, we selected the cytosolic or mitochondrial reactions belonging to glycolysis/gluconeogenesis, pentose-phosphate pathway (PPP), TCA cycle, anaplerosis, pyruvate metabolism, and oxidative phosphorylation. A cytosolic malate synthase was added to complement the glyoxylate shunt [104].

For L-alanine, two biosynthetic routes from pyruvate were included: cytosolic and mitochondrial alanine transaminase reactions, which were assumed to solely produce but not degrade L-alanine [19, 116]. Furthermore, L-glutamate could be produced via three alternative pathways: cytosolic or mitochondrial NADP-dependent glutamate dehydrogenase from $\alpha$-ketoglutarate or mitochondrial NAD-dependent glutamate synthase from $\alpha$-ketoglutarate and glutamine [191]. For glycine synthesis, we implemented three pathways such that it could be synthesized in the mitochondria via (i) alanine-glyoxylate
transaminase [170], or in the cytosol by (ii) glycine hydroxymethyltransferase from serine [89], or (iii) from L-threonine via threonine aldolase [125]. As the latter reaction was assumed to be reversible, it could also be used to produce L-threonine, and such it constitutes a second possibility to produce L-threonine next to the linear pathway from L-aspartate. For all other amino acids, the model contains only one linear cytosolic pathway consisting of consecutive enzymatic reaction steps. Here, no alternative paths exist or are excluded based on biochemical literature as it was done also to construct models for $^{13}$C-based flux analysis [19, 116].

The model further includes transport reactions across the mitochondrial membrane for metabolites that participate in reactions in both the cytosol and the mitochondria. Additional transport reactions (i.e. for L-glutamate, α-ketoglutarate, homocitrate, glyoxylate, and 2-oxobutanoate) that were not contained in iND750 were added to properly connect additionally included alternative pathways for amino acid synthesis to the metabolic network. In the model, carbon molecules that can be exchanged with the environment are glucose, glycerol, pyruvate, acetate, ethanol, succinate, and CO$_2$.

The biomass composition was adopted from iND750 besides that trehalose and glycogen were discarded since carbohydrate storage was not considered in our model. Lumped reactions for synthesis of the remaining biomass constituents, i.e. lipids, nucleotides, and cell wall components from the corresponding precursors were determined based on the biomass composition as provided in iND750.

**EFM Enumeration**

To determine the EFMs specific for either the glucose or ethanol growth phase, we draw on additional knowledge to reduce the model for each growth phase specifically and to such also reduce computational effort in EFM enumeration. Exchange fluxes with the environment were set according to the respectively measured extracellular rates. For the glucose growth phase, acetate, pyruvate, and ethanol were only allowed to be excreted, while only glucose could be taken up. For the ethanol growth phase, acetate, pyruvate, and glucose exchange were eliminated from the model, and only ethanol was allowed to
4. Metabolomics-based Elucidation of Metabolic Activities

be taken up. For both growth phases, succinate exchange was excluded as there was only minor production of this metabolite throughout the whole experiment. We also excluded glycerol production from intermediates of the EMP’s upper part as we here focus on the fluxes in ethanol metabolism. Supplementary File 1 provides the in this way constructed model.

Furthermore, biochemical knowledge on major flux directions was exploited to also exclude intracellular reactions or restrict them into one direction. Noteworthy, applying these assumptions we only exclude flux distributions that do not comply with \textit{a priori} knowledge on the direction of metabolic fluxes. For growth on glucose, the EMP was assumed to proceed in glycolytic mode and gluconeogenetic reactions were eliminated as well as the malate synthase \cite{29}. The non-oxidative branch of the PPP was restricted to operate from the ribulose 5-phosphate to the EMP intermediates, while the oxidative branch of the PPP was eliminated. The alcohol dehydrogenases were only allowed to produce ethanol. For growth on ethanol, the EMP was restricted to the gluconeogenetic direction and exclusively glycolytic reactions were eliminated. The TCA cycle was assumed to operate in the cyclic, respiratory mode.

The two model versions specifically adapted to each carbon source are also provided in Supplementary File 1. All modification with respect to iND750 are documented in this file. To calculate the EFMs, the freely available software Metatool \cite{143} was applied via Matlab functions supplied in \cite{190}.

**Testing EFMs for Thermodynamic Feasibility**

Testing each EFM for its thermodynamic consistency with respect to the measured concentrations was performed with NET analysis (cf. 3 for details about this tool). To reduce computational effort, we applied a three step procedure to reduce the number of EFMs that had to be tested by NET analysis. Here, we make use of single reactions, for which we know that they are restricted to one direction due to measured concentrations and the preliminary set of active reactions without considering constraints imposed by neighboring reactions. If a such defined direction contradicts the corresponding direction
in an EFM the EFM can be classified as thermodynamically infeasible right away, and it is not necessary to make use of the NET analysis framework.

Specifically, we first calculate possible ranges of Gibbs energies for each reaction based on the measured concentrations and the preliminary set of active reactions using the NET analysis framework. Exclusively positive (negative) ranges indicate that the respective reaction can only proceed in backwards (forwards) direction. The such derived direction restrictions apply also for all thermodynamically feasible EFMs that are used to describe the possible flux distributions as the concentrations and active reactions apply for the actual flux distribution. Hence, in the second step, we make use of these reaction directions to identify a first set of EFMs, which cannot be thermodynamically feasible: These EFMs assign an opposite direction to a reaction compared to the one that is allowed by the direction restriction that were derived in the first step. In the third step, we finally test the remaining EFMs for thermodynamic feasibility by again using NET analysis. Here, additional EFMs can be excluded from the set of thermodynamically feasible EFMs: A reaction direction might be incompatible with measured concentrations only if other reactions are active and such impose additional constraints on the concentrations. Such cases are disclosed by NET analysis. Hence, by sorting out all thermodynamically infeasible EFMs obtained in the latter two steps, we determine the set of thermodynamically feasible EFMs.

Acknowledgments

We thank W. Wiechert, J. Stelling, and M. Terzer for fruitful discussions regarding elementary flux modes. In particular, we acknowledge M. Terzer for his enthusiasm to derive a proof for the feasibility of this approach.
4.6. Appendix - Thermodynamically Feasible Flux Distributions as Linear Combinations of Thermodynamically Feasible EFMs

In the following, we demonstrate that the subset of thermodynamically feasible EFMs is sufficient to describe all thermodynamically feasible flux distributions. In other words, we reason that Eqn. 4.5 constitutes a necessary condition for any thermodynamically feasible flux distribution. A paper on a mathematical proof to this is in preparation and we here only give an illustrated outline of the reasoning.

Before we demonstrate that Eqn. 4.5 indeed holds true, we first define terms that will be used in the following. We define the direction pattern of a flux distribution as the sign pattern of its reaction rates. Here, a direction pattern is denoted by a vector containing all indices of active reactions with the sign corresponding to the direction (cf. Fig. 4.3). Furthermore, a subset of the entries of a such defined pattern vector describes a direction pattern subset to the respective direction pattern. We call reactions bi-directional if at least one EFM assigns a forward and at least one EFMs assigns a backward flux. All others are called uni-directional. An active reaction is a reaction to which either a positive or a negative reaction rate is assigned, i.e. a net conversion of reactants in a certain direction is proceeding. Hence, a direction pattern of a particular EFM contains either positive or negative indices of all active bi-directional reactions, and positive indices for all active uni-directional reactions.

First, we consider EFMs that are solely composed of active uni-directional reactions. Here, we demonstrate that a linear combination of EFMs of which at least one is a thermodynamically infeasible EFM, that contains only active uni-directional reactions (e.g. EFMs I and II in Fig. 4.3 B), cannot be a feasible flux distribution. Such EFMs are always direction pattern subsets of a flux distribution they contribute to. For instance, for the combination of the EFMs I and II in Fig. 4.3 B both EFMs are direction pattern subsets. Each additional active reaction leads to an additional thermodynamic constraint, i.e. another instance of Eqn. 4.4. In consequence, a flux distribution, to which also a thermodynamically infeasible EFM contributes, will also be infeasible. In other
words, a thermodynamically infeasible EFM containing only uni-directional reactions cannot be combined with another EFM such that the combination is feasible.

Now, we extend the discussion to EFMs that also contain active bi-directional reactions (EFM III and IV in Fig. 4.3 B). Then, the direction pattern of an EFM that contributes to a flux distribution does not need to be a direction pattern subset of this flux distribution. For example, an equally weighted combination of EFMs III and IV in Fig. 4.3 B gives a flux distribution where the bi-directional reaction becomes inactive as the opposite fluxes in EFMs III and IV cancel each other out. Then, the thermodynamic constraint on this reaction is released as the second law of thermodynamics applies only to active reactions. Hence, the flux distribution could be thermodynamically feasible although
one or more of the contributing EFMs may be infeasible.

Thus, the key question here is whether we exclude valid thermodynamically feasible flux distributions if we only use thermodynamically feasible EFMs to describe the possible flux distributions. This is not the case as will become clear in the following: As a central property of EFMs, the number of zero entries is maximized for each EFM [171]. As a consequence of this property, for a bi-directional reaction that can have a zero flux in a flux distribution, there exists an EFM that has a zero flux for this reaction (M. Terzer, pers. communication). This EFM can then be used to construct the flux distribution in an alternative way. Using the example from above, where combining EFMs III and IV resulted in a zero flux for the bi-directional reaction, we see that the combination can also be reconstructed by EFMs I and II.

It is important to note that a combination of feasible EFMs does not necessarily have to be a feasible flux distribution: For instance, the EFMs I, II, and III all contribute to the displayed flux distribution in Fig. 4.3 C. In our approach, only the direction pattern subsets corresponding to one of the EFMs are tested for their thermodynamic feasibility but not the complete direction pattern of the flux distribution. Such, it is not tested whether the thermodynamic constraints that correspond to all active reactions of the flux distribution lead to infeasibility if applied simultaneously. Therefore, although for a feasible flux distribution all direction pattern subsets also must be feasible, the flux distribution does not have to be feasible if the tested direction pattern subsets are feasible. Noteworthy, Beard and co-workers applied a similar approach for an ab initio prediction of reaction irreversibility for constructing metabolic models, however without incorporating experimental data.

Supplementary Files

Supplementary files are available on the CD provided with this document.

**Model** Spreadsheet (Supplementary File 1 - Metabolic model.xls) containing the applied model and the reduced versions for glucose and ethanol.

**Extracellular Rates** Figure (Supplementary File 2 - Extracellular rates.pdf) displaying
the extracellular rates calculated from the measured time courses of extracellular concentrations.

**Concentration Data** Spreadsheet (Supplementary File 3 - Measured concentrations.xls) containing the applied metabolite concentrations for testing for thermodynamic feasibility.

**Thermodynamically Feasible EFMs** Spreadsheet (Supplementary File 4 - EFM analysis.xls) that gives an overview on the whole sets of EFMs and the sets of thermodynamically feasible EFMs for ethanol and glucose, respectively.
4. Metabolomics-based Elucidation of Metabolic Activities
5. Quantitative Differences in Glucose Signaling and Regulation in *Saccharomyces cerevisiae* CEN.PK 113-D7 and FY4 Cause Similar Metabolic Phenotypes

Contributors:
Kümmel A. (leading author), Ewald J., Fendt S.-M., Picotti, P., Zamboni, N. and Heinemann, M.
Abstract

To understand the effects of signalling and regulation on metabolism, glucose repression in two commonly used yeast strains was quantitatively assessed using a data driven approach to detect metabolic and regulatory differences. In particular, the glucose repression state of *S. cerevisiae* CEN.PK 113-D7 and FY4 were compared based on quantitative, large-scale proteomics and metabolomics data sets as well as metabolic fluxes. The strains display comparable degree of glucose repression on the flux level as extracellular yields and the respiratory TCA cycle activities are similar. Nevertheless, differences in protein and metabolite concentrations were observed which indicated that the Hxk2-dependent signalling is stronger in CEN.PK. Differences in effective signaling and regulation became also evident as the glucose repression state on flux level - indicated by the respiratory TCA cycle activity - for a genetic perturbation differs significantly: While glucose repression was maintained in FY4, CEN.PK switched to a respiratory metabolism upon a ∆hxk2 deletion. We could demonstrate that a point mutation in the *CYR1* gene is underlying this particular distinct behavior.

5.1. Introduction

The availability of glucose, *Saccharomyces cerevisiae*’s preferred carbon source, is to a large extent decisive for the cells’ metabolic state e.g. whether the cell grows fermentatively or respiratory. The amount of glucose is sensed and complex signaling pathways transduce this signal to regulatory pathways that eventually realize the desired metabolic adjustments. Today, a large body of knowledge about the glucose-dependent signaling and regulation is available (cf. [161, 163, 164] for review). The major signaling and regulatory pathways acting on central carbon metabolism (CCM) during high extracellular glucose concentrations are the main glucose repression pathway and the cAMP/PKA signaling pathway.

A key protein in the main glucose repression pathway is the hexose phosphorylating enzyme Hxk2 which activates the transcriptional activator Mig1 via deactivation of
the protein kinase Snf1 [96]. Besides its catalytic function to phosphorylate glucose, Hxk2 represses respiration, tricarboxylic acid (TCA) cycle activity, and utilization of alternative substrates. Such, it is a crucial regulator for glucose repression [5, 15, 115, 160, 174] (cf. [126] for review). The cAMP/PKA signaling pathway has similar effects on central carbon metabolism [127, 140, 189, 203]. The main regulator of this pathway, PKA, is activated by cAMP, which is synthesized by the adenylate cyclase Cyr1 [90]. Yet, the intracellular signal inducing cAMP synthesis has not been elucidated. Furthermore, the main glucose repression pathway and the cAMP/PKA pathways are interconnected: For instance, glucose phosphorylation catalyzed by Hxk2 is assumed to also activate the cAMP/PKA signaling [163]. Also, PKA is known to be involved in the Hxk2-dependent inhibition of Snf1 [163]. Such, both pathways can produce a similar metabolic state.

At the presence of high glucose concentrations and despite the presence of oxygen, these signaling and regulatory pathways trigger a partially fermentative metabolism via repression of a large number of genes. Physiologically, the glucose repression is signified by production of ethanol and a low intracellular respiratory metabolic flux through the TCA cycle. Glucose repressed genes have been experimentally determined by detecting their transcriptional or translational products [205]. Both, metabolic fluxes as well as expression of transcripts and proteins were exploited to determine the effectiveness of glucose repression (e.g. [77, 151, 174]).

For studying glucose signaling and repression, both S288C descendants [114, 115, 135, 160] and CEN.PK strains [77, 174, 174, 205] are often employed. Although the topology of the signaling and regulatory network is most likely identical in these strains, the interaction strengths between the molecular signaling and regulation compounds and their targets in metabolism are likely to differ between strains. One such difference is, for example, caused by a known point mutation in CEN.PK’s CYR1 gene encoding the adenylate cyclase. As a result, this strain lacks the otherwise typical cAMP increase upon sudden glucose excess [194]. Thus, unknown genetic differences that even only affect the strength of signaling and regulation, are potential reasons for apparently contradicting experimental observations of studies using different strains (e.g. [75] and [123], or [152].
and [75]). In cases where the signaling and regulation topologies are identical, divergent outcomes of the same perturbation in different strains can be explained by different signaling and regulatory interactions strengths.

Here, using a systems biology approach, we aimed deciphering quantitative differences in glucose signaling and regulation under glucose excess conditions in the two prototroph MATa yeast strains CEN.PK113-7D (representing CEN.PK strains and referred to as CEN.PK in the following) and FY4 (representing S288C isogenic strains). Specifically, we characterized the metabolic state of CEN.PK and FY4 under glucose excess conditions. For this, we determined growth rates, uptake and excretion rates, the respiratory flux through the TCA cycle, as well as the metabolic proteome and metabolome using different mass spectrometry platforms. Although the strains display comparable degree of glucose repression on metabolic flux level, the protein concentrations indicate that glucose repression on the protein level is more effective in CEN.PK, presumable due to a stronger Hxk2-dependent signaling. The metabolic fluxes in $\Delta hxk2$ deletion mutants moreover revealed that in CEN.PK, but not in FY4, the glucose repression solely depends on $HXK2$. To next assess the effect of the known genetic difference in $CYR1$, we additionally employed the strain CEN.PK JT4 in which the $CYR1$ point mutation is reverted. This strain, too, was characterized on metabolic level and we compared the effects of a $HXK2$ deletion on metabolic fluxes in all three strains. This revealed that $HXK2$'s essentiality for glucose repression in CEN.PK is due to the known defect in the cAMP/PKA pathway. Moreover, our results indicate a higher cAMP/PKA activity in CEN.PK that is potentially triggered by a Hxk2-dependent signal.

5.2. Results

5.2.1. Differences Between CEN.PK and FY4

We first comprehensively characterized the two strains when aerobically grown on high glucose concentrations. Specifically, we determined the extra- and intracellular physiology and the concentrations of about 36 metabolites and 84 proteins.
5.2. Results

<table>
<thead>
<tr>
<th>Strain</th>
<th>Relevant genotype</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEN.PK 113-7D</td>
<td>MATa</td>
<td>Euroscarf, [194]</td>
</tr>
<tr>
<td>CEN.PK Δhxk2</td>
<td>CEN.PK 113-7D hxk2::kanMX4</td>
<td>[20]</td>
</tr>
<tr>
<td>CEN.PK JT4</td>
<td>MATa LCR1</td>
<td>M. Luttik, pers. comm.</td>
</tr>
<tr>
<td>CEN.PK JT4 Δhxk2</td>
<td>CEN.PK JT4 hxk2::kanMX4</td>
<td>this study</td>
</tr>
<tr>
<td>FY4</td>
<td>MATa</td>
<td>[208]</td>
</tr>
<tr>
<td>FY4-hxk2</td>
<td>FY4 hxk2::kanMX4</td>
<td>C. Boone, pers. comm.</td>
</tr>
</tbody>
</table>

Table 5.1. Yeast strains used in this work.

Extra- and Intracellular Physiology

To assess differences between CEN.PK and FY4 on metabolic flux level, we determined the glucose repression state of metabolism from physiological measurements (growth rates, uptake and excretion rates) and the respiratory TCA cycle activity which was determined in $^{13}$C-labeling experiments. This respiratory TCA cycle activity is defined as the ratio between the cyclic flux generating the redox equivalents for the respiratory chain and the flux producing the biomass precursors that also are TCA cycle intermediates (cf. Fig. 5.1 and Materials and Methods). We employ it as an intracellular readout of the the intracellular flux distribution and degree of glucose repression on metabolic level.

Figure 5.1. Exemplary scheme on metabolic fluxes for high and low glucose repression. Under high glucose repression (left scheme), most of the flux through glycolysis is directed to ethanol excretion. Such, both, the cyclic TCA cycle flux and the respiratory TCA cycle activity as determined by the displayed equation, are low. On the contrary, a less glucose repressed metabolism (right scheme) is characterized by a higher respiratory TCA cycle activity as a higher fraction of the glycolytic flux is directed to the cyclic TCA cycle flux.

By monitoring growth rates and uptake and excretion rates in shake flask cultures, we observed higher extracellular rates of biomass and ethanol production and glucose consumption in CEN.PK. In contrast, all the various yields are almost identical in the
5. Experimental Approach to Quantitative Strain Comparison

two strains (Tab. 5.2). This suggests a similar relative intracellular flux distribution in the two strains with generally higher rates. As a readout on the intracellular flux distribution, we determined the respiratory TCA cycle activity in $^{13}$C-labeling experiments. As can be seen in Fig. 5.2, these ratios are very small in both strains which underlines that the fluxes in the two strains essentially only differ in terms of generally higher metabolic rates in CEN.PK.

### Table 5.2. Physiological parameters of aerobic glucose-growing batch cultures in shake flasks.

<table>
<thead>
<tr>
<th>strain</th>
<th>biomass (g g$^{-1}$)</th>
<th>glucose (g g$^{-1}$)</th>
<th>ethanol (g g$^{-1}$)</th>
<th>acetate (g g$^{-1}$)</th>
<th>glycerol (g g$^{-1}$)</th>
<th>pyruvate (g g$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEN.PK wt</td>
<td>0.114 ±0.004</td>
<td>-</td>
<td>0.41 ±0.02</td>
<td>0.011 ±0.001</td>
<td>0.027 ±0.002</td>
<td>0.0032 ±0.0002</td>
</tr>
<tr>
<td>JT4</td>
<td>0.107 ±0.005</td>
<td>-</td>
<td>0.37 ±0.02</td>
<td>0.008 ±0.001</td>
<td>0.029 ±0.002</td>
<td>0.0029 ±0.0002</td>
</tr>
<tr>
<td>FY4 wt</td>
<td>0.113 ±0.004</td>
<td>-</td>
<td>0.40 ±0.03</td>
<td>0.008 ±0.001</td>
<td>0.033 ±0.001</td>
<td>0.0036 ±0.0002</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>strain</th>
<th>yield ($g_{glyc}$)</th>
<th>rate ($g_{Dw}$ h$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEN.PK wt</td>
<td>0.46 ±0.05</td>
<td>4.04 ±0.14</td>
</tr>
<tr>
<td>JT4</td>
<td>0.45 ±0.06</td>
<td>4.21 ±0.20</td>
</tr>
<tr>
<td>FY4 wt</td>
<td>0.33 ±0.009</td>
<td>2.92 ±0.13</td>
</tr>
</tbody>
</table>

**Figure 5.2.** Respiratory TCA cycle activities. The respiratory TCA cycle activity was determined in $^{13}$C-labeling experiments as the ratio of oxaloacetate production from L-malate via cyclic TCA cycle flux and its production from pyruvate via anaplerosis. The gray parts of the bars for the Δhxk2 mutants of the CEN.PK strains indicate the increase in respiratory TCA cycle activity that is related to the growth rate reduction caused by the gene deletion (as determined from [20]).

Hence, as (intra- and extracellular) relative rates do not differ significantly, CEN.PK higher growth is not due to internal flux redistribution. This suggests a mechanism that generally increases the metabolic fluxes in CEN.PK in comparison to FY4 which can be resident in the signaling and regulatory topology or a result of differences in other cellular processes (e.g. different amounts of ribosomes). With respect to the glucose repression state, both the low respiratory TCA cycle activity and the high ethanol yields confirm that the metabolism is similarly glucose repressed in both strains.
5.2. Results

Metabolites

Concentrations of 16 metabolites that could be determined via either GC-TOF or LC-MS/MS from the CCM, three co-substrates (i.e. ADP, AMP, and NAD$^+$), and 17 amino acids were quantified (cf. Supplementary File 1). We observe the following global concentration differences (cf. Fig. 5.3): The metabolites of the upper glycolysis (g6p, f6p, fdp, dhap, 3pg) and the pentose-phosphate pathway (PPP) show higher concentrations in CEN.PK than in FY4, while there are no common trends for the metabolites in the TCA cycle. Together with the observed increased uptake of glucose in CEN.PK, the elevated pools of upper glycolysis intermediates in this strain point to a facilitated glucose influx into glycolysis. This could be due to a higher uptake and/or glucose phosphorylation capacity in CEN.PK.

Next, looking at single reaction steps, we find reactions that display significantly different substrate-to-product ratios between CEN.PK and FY4 (cf. Fig. 5.4A). A higher substrate-to-product ratio indicates that the reaction operates further away from the thermodynamic equilibrium. Higher substrate-to-product concentration ratios indicate a more limited reaction with regard to enzyme capacity and such point to potential regulation of the respective reaction [102]. Using this assumption, the PFK and SDH reaction would have a more limiting enzyme capacity in CEN.PK.

Taken together, the metabolite concentrations indicate that CEN.PK’s higher glucose uptake rate is caused by enhanced glucose transport and/or phosphorylation capacity. The lower growth rate of FY4 could be explained by a reduced influx into glycolysis that results in a general reduction of the metabolic fluxes. Furthermore, the different substrate-to-product ratios indicate that enzyme capacities are not the same and thus different limitations in single reaction steps exist between the two strains. These limitations, though not effecting the relative metabolic fluxes, point to quantitative differences in the metabolic regulation which might be due to different allosteric or transcriptional/translational regulation.
5. Experimental Approach to Quantitative Strain Comparison

Figure 5.3. Metabolite concentration differences in CEN.PK in comparison to FY4 and CEN.PK JT4. Metabolite concentrations were determined via GC-TOF (TCA cycle intermediates and amino acids) or LC-MS/MS (EMP and PPP intermediates and co-factors) in biological duplicates. Abbreviations: g6p: glucose 6-phosphate; f1p: fructose 1-phosphate; f6p: fructose 6-phosphate; fdp: fructose 1,6-bisphosphate; dhap: dihydroxyacetonephosphate; 3pg: 3-phosphoglycerate; pep: phosphoenolpyruvate; 6pgc: 6-phosphogluconate; ru5p-D: D-ribulose 5-phosphate; r5p: ribose 5-phosphate; xu5p-D: D-xylulose 5-phosphate; cit: citrate; icit: isocitrate; akg: α-ketoglutarate; succ: succinate; fum: fumarate; mal-L: L-malate; adp: diphosphate; amp: adenosine monophosphate; nad: nicotinamide adenine dinucleotide (reduced form); his-L: L-histidine; phe-L: L-phenylalanine; tyr-L: L-tyrosine; ser-L: L-serine; gly: glycine; met-L: L-methionine; ala-L: L-alanine; val-L: L-valine; leu-L: L-leucine; ile-L: L-isoleucine; glu-L: L-glutamate; gln-L: L-glutamine; pro-L: L-proline; asp-L: L-aspartate; asn-L: L-asparagine; lys-L: L-lysine; thr-L: L-threonine.
5.2. Results

Figure 5.4. Substrate-to-product ratio differences in CEN.PK in comparison to FY4 and CEN.PK JT4. A: The substrate-to-product ratios are compared with the corresponding enzyme concentrations. Here, reactions are displayed for which both substrates and products (excluding redox or energy co-factors) were measured. Abbreviations: PGI: phosphoglucone isomerase reaction; PGTM: phosphoglucomutase reaction; PFK: phosphofructokinase reaction; ENO: enolase reaction; GND: 6-phosphogluconate dehydrogenase reaction; RPI: ribosephosphate isomerase reaction; IDH: isocitrate dehydrogenase reaction; SDH: succinate dehydrogenase reaction; FUM: fumarase reaction. B: Concentration ratios of the enzymes corresponding to the reactions shown in panel A.
5. Experimental Approach to Quantitative Strain Comparison

Proteins

Next, as the second set of compounds that are next to the metabolites directly involved in metabolic reactions, we assessed quantitative differences on the protein level: We determined amounts of 84 proteins in central carbon metabolism relative to an internal standard (cf. Materials and Methods and Supplementary File 2). The concentrations of 27 proteins are more than 2-fold higher or lower in CEN.PK than compared to FY4 (Fig. 5.5). Most striking, many proteins involved in carbohydrate storage metabolism and several in the TCA cycle are less abundant in CEN.PK. Both metabolic parts are known to be downregulated in glucose repression [161]. Moreover, proteins in other parts of metabolism that are significantly less abundant in CEN.PK (Hxk1, Glk1, Tdh1, Eno1, Pyk2, and Ald4) are also known to be glucose repressed [129, 160, 174, 176]. We thus conclude that CEN.PK has stronger glucose repression regulation with respect to protein expression. Furthermore, as many of the less abundant proteins in CEN.PK have been shown to be transcriptionally upregulated upon deletion of HXK2 [205], the here observed decreased levels are supposedly caused by a stronger Hxk2-dependent regulatory signal in this strain.

In a next step, we asked whether altered protein abundances correlate with the observed differences in the metabolite concentrations. Higher glycolytic metabolite concentrations suggested a higher capacity of either glucose transport or phosphorylation in CEN.PK. Here, the protein abundances indicate which step supposedly contributes to this higher capacity: While Hxk2 amounts are similar (data not shown), two of the glucose phosphorylating enzymes, Hxk1 and Glk1, are much less abundant in CEN.PK which suggests that the transport of glucose into the cell is the decisive step that leads to a higher glucose uptake rate in CEN.PK. This hypothesis is supported by the fact that mainly the glucose transport controls the glycolytic fluxes [23, 156].

Next, we compared the observed differences of the substrate-to-product ratios in Fig. 5.4A with the corresponding protein differences (Fig. 5.4B). Although we observed differences in the protein concentration of the ribose phosphate isomerase (RPI) and the phosphoglucone isomerase (PGI) reactions, the corresponding reactant concentration ra-
5.2. Results

Figure 5.5. Protein concentration differences in CEN.PK in comparison to FY4 and CEN.PK JT4. Here, proteins are displayed which are more than 2-fold higher or lower abundant in CEN.PK in comparison to either FY4 or CEN.PK JT4. Protein concentrations were determined via MRM-based LC-MS/MS in relation to an internal standard in biological duplicates. Proteins are labeled by the corresponding gene names. Triangles indicate minimal absolute concentration ratios.
5. Experimental Approach to Quantitative Strain Comparison

tios are comparable between the two strains. Also, the Gibbs energies of reaction for PGI and RPI are close to zero in both strains (absolute values less than $3 \frac{kJ}{molK}$). The substrate and product concentration are nearly thermodynamically equilibrated which is supposably due to large enzyme capacities that exceed the actual metabolic needs. Hence, these reactions are not flux controlling [102]. In contrast, there are four reactions where a lower enzyme concentration actually leads to less equilibrated substrate-to-product pools: enolase (ENO), 6-phosphogluconate dehydrogenase (GND), isocitrate dehydrogenase (IDH), and succinate dehydrogenase (SDH). In comparison with neighboring reactions, these reactions are more likely to control metabolic fluxes as only limited catalytic activity is provided by transcriptional or translational regulation. The different flux control for the ENO, IDH, and SDH reactions, which are all known to be influenced by glucose repression, is caused by a higher glucose repression in CEN.PK.

In summary, significantly different protein concentrations exists between CEN.PK and FY4. Generally, the differences to a more active glucose repression at protein level, potentially by a stronger Hxk2-dependent regulatory signal. Remarkably, the altered protein concentrations which locally influence the flux control do not result in significantly different metabolic flux ratios.

5.2.2. Potentially Different Transcription Factor Activities

Next, we aimed at tracing back which transcription factors (TFs) might have caused the observed differences in protein abundance. Using the information about transcriptional regulation from Yeastact [185], we aimed at determining TFs that are significantly often linked to proteins that are more than 3-fold less abundant in CEN.PK than in FY4. (cf. Materials and Methods). The most significant TFs ($p$-values $< 0.002$) are Msn2 and Msn4 (Tab. 5.3), which are activating TFs and contribute to the general stress response in yeast. As they are also mediating hyperosmotic shock resistance [117], we compared CEN.PK and FY4 hyperosmotic shock resistance. We found that CEN.PK is less resistant to hyperosmotic stress as no CEN.PK cells survived a residence in a high osmolar glycerol solution for 40 min while the survival rate was 16% ($\pm 7\%$) for FY4.
5.2. Results

(cf. Materials and Methods). This observation confirms the results of our analysis as it suggests that Msn2 and Msn4 are less active in CEN.PK which would lead to less stress resistance of this strain.

Table 5.3. Transcription factors that are potentially differently active in CEN.PK and FY4. p-values were determined based on a computational analysis using information about the transcriptional network as documented on YeastRACT and the measured protein abundances.

<table>
<thead>
<tr>
<th>TF</th>
<th>p-value</th>
<th>description</th>
<th>associated, lower abundant proteins in CEN.PK than in FY4</th>
<th>supposed activity in CEN.PK in comparison to FY4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Msn2</td>
<td>0.0008</td>
<td>activator of stress response genes</td>
<td>Hxk1, Glk1, Enol1, Ald4, Glc3, lower</td>
<td>lower</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Gsy1, Gpk1, Tps1, Tsl1, Nth1, Gdb1, Gpd1</td>
<td></td>
</tr>
<tr>
<td>Msn4</td>
<td>0.0013</td>
<td>activator of stress response genes</td>
<td>Hxk1, Glk1, Ald4, Glc3, Gsy1, lower</td>
<td>lower</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Gpk1, Tps1, Tsl1, Nth1, Gdb1, Gpd1</td>
<td></td>
</tr>
<tr>
<td>Mig1</td>
<td>0.0119</td>
<td>repressor of glucose repressed genes</td>
<td>Hxk1, Tdh1, Gsy1, Nth1, Gdb1, higher</td>
<td>higher</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nrg1</td>
<td>0.0153</td>
<td>repressor of glucose repressed genes</td>
<td>Hxk1, Ald4, Glc3, Gsy1, higher</td>
<td>higher</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nrg2</td>
<td>0.0153</td>
<td>repressor of glucose repressed genes</td>
<td>Hxk1, Ald4, Glc3, Gsy1, higher</td>
<td>higher</td>
</tr>
</tbody>
</table>

Furthermore, the list of TFs with p-values < 0.02 comprises Mig1, Nrg1, and Nrg2, all of which are involved in glucose repression [80]. In contrast to Msn2/4, they are repressing gene expression, and such we conclude that they are more active in CEN.PK. Thus, they supposedly mediate the higher glucose repression in CEN.PK.

5.2.3. Different Effects of Genetic Perturbations in CEN.PK and FY4

The quantitative comparison of CEN.PK’s and FY4’s physiology, metabolome, and proteome revealed significant differences, which point to different interaction strengths in the signaling and regulatory network: We concluded that the Mig1 and Nrg1/2-mediated glucose repression more effectively decreases the abundances of glucose repressed proteins in CEN.PK. Although on the flux level the strains display comparable degree of glucose repression (i.e. similar yield and respiratory TCA cycle activity), the metabolite concentrations indicate that the flux control differs between the strains. The SDH
reaction for example is more likely to control the TCA cycle flux in CEN.PK as the substrate-to-product ratio is further from the thermodynamic equilibrium. Such, genetic perturbations in the signaling and regulatory network are expected to result in different effects due to the differences in the interaction strengths as the respective signaling or regulation compounds are differently effective on metabolic fluxes. To test this hypothesis, we next analyzed the metabolic effects of two genetic perturbations - namely of a \( HXK2 \) deletion in both strains and the effect of a repaired point mutation in \( CYR1 \) in CEN.PK.

**Effect of Hexokinase II Deletion in CEN.PK and FY4**

Using \(^{13}\)C-labeling experiments, we determined the respiratory TCA cycle activity in CEN.PK and FY4 lacking \( HXK2 \). In FY4 \( \Delta hxk2 \), we found an increased but still low absolute TCA activity of \( 0.096 \pm 0.016 \) (Fig. 5.2) indicating that there is still significant glucose repression on flux level resulting in also yet low biomass yields (Tab. 5.4). Hence, with respect to metabolic fluxes, glucose repression is maintained to a large extent in FY4 \( \Delta hxk2 \).

<table>
<thead>
<tr>
<th>Strain</th>
<th>growth rate (h(^{-1}))</th>
<th>biomass yield (( g_{DW} g_{glucose}^{-1} ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEN.PK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wt</td>
<td>0.39 ± 0.02</td>
<td>0.13 ± 0.02</td>
</tr>
<tr>
<td>( \Delta hxk2 )</td>
<td>0.22 ± 0.06</td>
<td>0.43 ± 0.04</td>
</tr>
<tr>
<td>JT4</td>
<td>0.36 ± 0.03</td>
<td>0.13 ± 0.01</td>
</tr>
<tr>
<td>JT4 ( \Delta hxk2 )</td>
<td>0.17 ± 0.01</td>
<td>0.19 ± 0.02</td>
</tr>
<tr>
<td>FY4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wt</td>
<td>0.33 ± 0.01</td>
<td>0.12 ± 0.01</td>
</tr>
<tr>
<td>( \Delta hxk2 )</td>
<td>0.32 ± 0.03</td>
<td>0.16 ± 0.01</td>
</tr>
</tbody>
</table>

In contrast, deleting \( HXK2 \) in CEN.PK has much more pronounced effects on the respiratory TCA cycle activity. Here, it increases from \( 0.024 \pm 0.003 \) to a high value of \( 0.735 \pm 0.020 \) (Fig. 5.2). However, this increase is at least partially caused by the simultaneously provoked growth rate reduction: In CEN.PK, Blank and Sauer demonstrated that the respiratory TCA cycle activity inversely correlates with growth rate
Nevertheless, even after correcting for the growth rate reduction related increase of the respiratory TCA cycle activity (Fig. 5.2), a large portion of the observed increase in the Δhxk2 mutant can still be attributed to the HXK2 deletion itself. In agreement with the increased respiratory TCA cycle activity, the biomass yield reaches a level that resembles values of fully respiring glucose-limited cultures [194]. Hence, in-line with previous studies we noticed that in CEN.PK glucose repression heavily relies on Hxk2.

In summary, a deletion of HXK2 in CEN.PK results in significant glucose de-repression while the same deletion causes only a slight glucose de-repression in FY4. Hence in FY4, alternative regulation pathways such as the cAMP/PKA pathway either dominate the glucose repression regulation already in the wild-type or take over the regulatory role of Hxk2 upon its deletion.

Effect of the Repaired Point Mutation in CYR1 in CEN.PK

In contrast to FY4, a point mutation in CEN.PK’s CYR1, encoding adenylate cyclase, impairs glucose and acidification-induced cAMP concentration increase in de-repressed cells, while glucose repressed cells are not affected [196]. As expected for our glucose excess experiments, we do not observe any significant differences between the CEN.PK and the CEN.PK JT4, which is isogenic to CEN.PK besides that the point mutation is repaired: Metabolic fluxes, namely growth rate, biomass yield, and respiratory TCA cycle activity, as well as metabolite and protein concentrations are very similar (Tab. 5.4, Fig. 5.2, Fig. 5.3, and Fig. 5.5). For the metabolites, we only observe considerably higher concentrations for 3-phosphoglycerate and phosphoenolpyruvate in CEN.PK which is however not the case when comparing to FY4. Since CEN.PK JT4 compares to CEN.PK while hardly any similar pattern with FY4 can be recognized, the point mutation cannot explain differences of CEN.PK and FY4 wild-types growing exponentially on glucose. Nevertheless, we next asked whether the point mutation is relevant for the distinct behavior of the Δhxk2 deletion mutants. We hence determined the respiratory TCA cycle activities in CEN.PK JT4 and the corresponding Δhxk2 deletion mutant as a readout on the degree of glucose repression on metabolic fluxes.
5. Experimental Approach to Quantitative Strain Comparison

The physiology and respiratory TCA cycle activity in CEN.PK JT4 Δhxk2 shows that this strain is less glucose repressed as the wild-type CEN.PK JT4. However, the growth rate is decreased upon this deletion. Assuming a similar dependence of the respiratory TCA cycle activity on the growth rate in CEN.PK JT4 as in CEN.PK, the observed relief in glucose repression is not solely an effect of the HXK2 deletion. To estimate the degree to which the deletion de-represses the metabolism, we subtract the contribution of a slower growth rate in the deletion mutant from the observed respiratory TCA cycle activity increase in the same way as we did for CEN.PK. We find that its increase in CEN.PK JT4 is almost completely caused by the decreased growth rate alone (cf. Fig. 5.2). Hence, we conclude that the cAMP/PKA pathway of CEN.PK JT4 can make up for the loss of glucose repression due to the lack of Hxk2-dependent signaling.

Summarizing, the point mutation cannot explain the observed differences between FY4 and CEN.PK when the strains grow exponentially on glucose. Although under this experimental condition the adenylate cyclase hardly has any effect, our results suggest that its 'natural' (repaired) form (as in FY4 and CEN.PK JT4) is able to activate PKA via a Hxk2-independent signal under glucose excess conditions. However, as long as Hxk2 exerts its regulatory effect, the signal transduction via the adenylate cyclase does not become evident. As we have observed in the HXK2 deletion mutants of FY4 and CEN.PK JT4, this signaling route sustains glucose repression.

5.3. Discussion

CEN.PK and S288C isogenic strains are both commonly used to study glucose signaling and regulation. To address the question to which extent these strains and thus those studies are comparable to each other, we compared the prototrophic strains CEN.PK 113-7D and FY4. Under the assumption that their signaling and regulation network topologies are the same, we here investigated the quantitative differences within this common network to link them to known and unknown differences in the signaling and regulation strengths. Most significant differences are:

(1) Growing exponentially on glucose, CEN.PK has generally higher metabolic fluxes,
though the ratios of the fluxes to each other are similar in FY4. Hence, the degree of glucose repression with respect to metabolic fluxes is comparable in the two strains.

(2) A potential explanation for the generally higher metabolic rates is a limited substrate influx into glycolysis: Higher metabolite concentrations together with lower protein abundances of glucose phosphorylating enzymes indicate that the glucose uptake is less limited in CEN.PK.

(3) Glucose repressed enzymes (e.g. Hxk1, Glk1, Ald4 and TCA cycle enzymes) are less abundant in CEN.PK. This suggests lower Msn2/4 and higher Mig1 and Nrg1/2 activity whereas the latter set of TFs are all downstream of Snf1 and thus downstream of Hxk2.

(4) Presumably as a result of a lower Msn2/4 activity, exponentially growing CEN.PK cultures are less resistant to hyperosmotic stress.

(5) Hxk2 is essential for glucose repression in CEN.PK, but not in FY4. The glucose repression function of Hxk2 can however be compensated by a adenylate cyclase that features the glucose-induced cAMP-response (as in FY4 and CEN.PK JT4).

Taking all these findings together, we can formulate a model that is consistent with all our observations (Fig. 5.6). Our model draws on the known signaling and regulatory pathways from the literature, but adds information on interaction strength. We propose a higher regulatory strength of Hxk2 in CEN.PK having the following three effects which we inferred from our data: First, expression of the low affinity hexose transporter Hxt1 is increased via Reg1 or Grr1 (resulting in observations 1 and 2). Second, Snf1 is more repressed by Hxk2 and this relieves repression of Mig1 and Nrg1/2 (cf. observation 3). Last, Hxk2 also increases the PKA activity that leads to lower Msn2/4 activity (cf. observation 4).

This different operation of the regulatory machinery also accounts for the distinct implications of a \textit{HXXK2} deletion in the different genetic backgrounds we tested. In
5. Experimental Approach to Quantitative Strain Comparison

Figure 5.6. Model for explaining observations on strain differences.

1. Enhanced glucose transport inferred from metabolite (and protein) concentrations caused by more active Hxk2.
2. More active Nrg1/2 and Mig1 inferred from TF analysis as result of higher Hxk2 activity.
3. Although differences in the concentrations of metabolic proteins exist, both strains are similarly glucose repressed on metabolic flux level.
4. High Hxk2-dependent signal that results in higher basal PKA activity.
5. Based on the comparison of CEN.PK Δhxk2 and CEN.PK JT4 Δhxk2, the Hxk2-independent signal transduction is impaired by the point mutation in CYR1.
6. Ability of (Hxk2-independent) PKA signal to maintain glucose as demonstrated in CEN.PK JT4 Δhxk2 (which could also potentially mediated through inactivation Snf1 [70, 71, 163]).
7. Less active Msn2/4 inferred from TF analysis (and confirmed by testing for hyperosmotic stress resistance) as result of an increased PKA activity.

Note that the quantification of the signal relate to comparison between CEN.PK and FY4 and are no absolute qualification. The signaling and network topology is based on reported interaction from the literature (activation of Reg1/Grr1 via Hxk2 [161], repression of Snf1 by Hxk2 [96, 201, 215], glucose repression induced by cAMP/PKA signaling [214], Hxk2-dependent and Hxk2-independent activation of PKA [39, 161]).
5.3. Discussion

CEN.PK JT4, the Cyr1 is able to be activated by a Hxk2-independent signal, which is not the case for CEN.PK. Therefore, glucose repression can be maintained in CEN.PK JT4 Δhxk2 through a still activated PKA. The point mutation prevents sustained glucose repression in CEN.PK Δhxk2. With respect to the point mutation in CYR1, FY4 resembles CEN.PK JT4. Hence, also here Hxk2-independent PKA signaling potentially contributes to sustaining glucose repression in this strain (cf. observation 5).

One crucial element in our model is the activation of PKA via Hxk2. In literature, several mechanisms are discussed to induce PKA via Cyr1 by an intracellular signal in all of which all Hxk2 is involved: The suggested triggers of Cyr1 activation include glucose phosphorylation, acidification, or ATP/AMP ratios [163]. Our observations-based model gives further evidence on this connection. As we here used steady-state experimental conditions, the actual Hxk2-dependent mechanism is likely to increase the steady-state PKA activity.

Our line of argumentation to explain all observations is based on a hypothesized higher regulatory activity of Hxk2 in CEN.PK than FY4 despite very similar concentrations of this protein. One possible explanation for Hxk2’s different regulatory strength could be differences in the regulatory domain of the protein. It is also conceivable that the domain responsible for the protein’s nuclear localization comprises differences such that nuclear localization of Hxk2 is enhanced in CEN.PK. These hypotheses can be tested through sequence analysis, phosphorylation analysis, and/or experiments to determine subcellular localization of the protein.

On the whole, this quantitative analysis demonstrates that although both strains most likely possess the same signaling and regulation network topology, different interaction strengths cause differences in metabolic control. As demonstrated by deleting HXK2 in CEN.PK and FY4, perturbations can thus lead to very distinct results for two different strains. Even more striking, CEN.PK JT4 differs from CEN.PK only in a point mutation. Yet, we observed significant metabolic effects in the Δhxk2 mutants.

As another remarkable observation, the metabolic fluxes of the wildtype strains were similar while the protein concentrations were different. Possible explanations for this
apparent contradiction could be that differences are only located at non-decisive sites in metabolism or that the two strains have different strategies to achieve the same metabolic flux state. Either way, looking at different readouts on the metabolic state could result in different conclusions: Considering the ethanol production or the respiratory TCA cycle activity indicate a strong glucose repression in both strains. However, looking at the proteome level, higher abundant glucose repressed proteins (e.g. Hxk1 and Glk1) in FY4 suggest a less pronounced glucose repression in this strain. First, this stresses the caution indicated in comparing different strains due to unequal regulatory activities. Second, for understanding the relevance of a regulator in the signaling and regulatory network, its eventual metabolic effects have to be assessed as they report on the regulator's essentiality to achieve a certain metabolic state.

5.4. Materials and Methods

Strains and Culture Conditions

The strains used in this study are listed in Tab. 5.1. In CEN.PK JT4, the HXK2 gene was deleted by the short flanking homology method using the loxP-kanMX4-loxP gene disruption cassette [68]. The deletion cassette was amplified from pUG6, using as a forward primer 5'-TCTTTGTTGACCTTCGCCACTGTCTTATCTACAATAAC-3', which includes 18 nucleotides complementary to pUG6 and a 40-nucleotide extension corresponding to the region -150 to -110 upstream of the start codon of the HXK2 ORF, and as a reverse primer 5'-AGTACGCAAGCTATCTAGAGGAAGTGTAGAGAGGGTTAAATAGT-GGATCTGATATCACCCTC-3', which includes 21 nucleotides complementary to pUG6 and 40 nucleotides corresponding to the region +1851 to +1811 downstream of the start codon of the HXK2 ORF. Deletion of the HXK2 ORF was confirmed by PCR on genomic DNA extracted from G418R transformants.

In all experiments, cells were grown in minimal defined medium with glucose as sole carbon source. The medium was prepared from autoclaved salt and glucose solutions and sterile filtered solutions of vitamins and trace metals to reach concentrations as
5.4. Materials and Methods

described in Verduyn et al. [197]. Liquid precultures with 10 g/l glucose were inoculated by colonies from YPD plates. In all experiments, the glucose concentration was 10 g/l and the temperature was maintained at 30 °C. Phthalate buffer (10 mM for shake flask experiments, 90 mM for \(^{13}\)C labeling experiments) was used to maintain pH at a value of 5.

**Biomass and Extracellular Metabolite Concentrations**

Biomass concentrations were monitored by measuring optical density (OD) at 600 nm with a spectrophotometer (Pharmacia Novaspec II). The biomass dry weight was calculated using a OD-biomass dry weight (DW) correlation coefficient of 0.486 \(\frac{a_{DW/l}}{OD}\) (cf. appendix). To determine the extracellular concentrations of glucose, ethanol, acetate, pyruvate, and glycerol (as described in the appendix, 1 ml samples were taken and centrifuged for 4 min at 4000 rpm at 4 °C. The growth rates, \(\mu\), and biomass yields, \(Y_{x\text{glc}}\), were calculated by linear regression using Eqs. 5.1-5.2 based on the data (i.e. biomass concentrations, \(x\), and extracellular metabolite concentrations, \(c\)) from all biological replicates using Matlab.

\[
\ln(x) = \ln(x_0) + \mu t \tag{5.1}
\]
\[
x = x_0 + Y_{x\text{glc}}(c_{\text{glc}0} - c_{\text{glc}}) \tag{5.2}
\]

The yields of ethanol, acetate, glycerol, and pyruvate were calculated by using equations analog to Egn. 5.2.

**Respiratory TCA Cycle Activity**

The respiratory TCA cycle activity was determined from \(^{13}\)C-labeling experiments where 10% of the glucose in the medium was uniformly labeled (obtained from Sigma, Switzerland). These experiments (both, preculture and culture) were carried out in deep well plates with a culture volume of 1.2 ml and biomass concentration was measured at 600 nm (Spectra MAX Plus, Bucher Biotec AG). Cultures were harvested during exponential growth at ODs between 0.8 and 1.3 and centrifuged at 4 °C and 4000 rpm for 3 min.
5. Experimental Approach to Quantitative Strain Comparison

Supernatant and cell pellets were separately frozen at -40 °C. Extracellular metabolite concentrations in the supernatant were determined via HPLC-IR/DAD methods as described above.

To determine the labeling patterns in proteogenic amino acids, samples were prepared and analyzed by GC-MS as described in Blank and Sauer [20]. Cell pellets were hydrolyzed in 180 µl 6 M HCl at 105 °C for 12 h and the hydrolysate dried at 95 °C. The free amino acids were derivatized at 85 °C for 1 h using 20 µl dimethylformamide and 20 µl N-(tert-butyl(dimethyl)silyl)-N-methyltrifluoroacetamide [60]. GC-MS analysis was carried out using a series 6890N GC (Agilent Technologies, Switzerland) in combination with an MD800 mass spectrometer (Agilent Technologies, Switzerland) [59]. The respiratory TCA cycle activity, $TCA_{\text{resp}}$, was calculated from the fraction of mitochondrial oxaloacetate from anaplerosis, $oaa^m_{\text{ana}}$, which was determined using FiatFlux ([212]) with a stoichiometric network as reported in [20].

$$TCA_{\text{resp}} = 1 - oaa^m_{\text{ana}}$$  \hspace{1cm} (5.3)

Intracellular Metabolite Concentrations

Intracellular metabolite concentrations from two biological and two technical replicates were determined on two mass spectrometry platforms (GC-TOF and LC-MS/MS) by the same quenching, sample preparation, and mass spectrometry analysis procedure as described in the appendix.

Protein Abundances

Protein abundances were determined by a multiple reaction monitoring (MRM) based mass spectrometry approach as described in the appendix. To calculate the average protein abundances from the two biological replicates, mean values of peptide amounts were calculated from the transitions, mean values for proteins from the peptide mean values, and finally the two biological replicates were averaged. A conservative estimation of the standard deviation was calculated from the variance of the pool of all measured transitions for the respective protein.
Osmotic Shock Experiments

To determine the survival rate after hyperosmotic shock, we adapted a method reported in Beney et al. [16]. Colony forming units (CFU) for an aliquot of yeast culture exposed to hyperosmotic shock and a non-treated aliquot - both from an exponentially growing culture - were counted. For hyperosmotic shock treatment, an aliquot of 1 ml was mixed with 9 ml autoclaved glycerol/water solution (116.9 g glycerol in 50 ml water, approx. 65 mPa) and left for 40 min at room temperature. Both treated and non-treated samples were diluted 33000 and 66000 times respectively in autoclaved isotonic NaCl solution (9 g/l) and 200 µl spread on YPD plates. After 2 days incubation at 30 °C, photographs were taken and the CFU counted. The survival rate was calculated as the ratio of CFU of treated versus non-treated samples.

Transcription Factor Analysis of Protein Abundances

TFs which are more often than by chance associated with the subset of proteins that are 3-times lower abundant in CEN.PK were determined by a statistical analysis adopted from [26]. We here applied documented TF-gene associations reported in [185]. For each TF that interacts with the proteins for which concentrations were determined, a p-value based on a hypergeometric distribution was calculated.

\[ p_j = 1 - \sum_{i=0}^{k_j-1} \binom{M_j}{i} \binom{N-M_j}{n-i} \binom{N}{i} \]  

(5.4)

Here, \( N \) is the total number of proteins measured and \( n \) the number of proteins that are 3-times lower abundant in CEN.PK. \( M_j \) and \( k_j \) give the number of measured proteins and measured, 3-times lower abundant proteins in CEN.PK respectively that interact with TF \( j \). The such calculated p-value gives the probability that the TF has at least the observed number of interactions with the subset of lower abundant proteins.

Acknowledgments

We acknowledge Jack Pronk for providing the strain CEN.PK JT4 and Stephanie Heux for deleting HXK2 in this strain. Furthermore, we thank Reinhard Dechant for discus-
5. Experimental Approach to Quantitative Strain Comparison

...ision on glucose signaling and regulation. Uwe Sauer and Dirk Müller were helpful by reading the manuscript and giving constructive feedback.

Supplementary Files

Supplementary files are available on the CD provided with this document.

Metabolite Concentrations Spreadsheet (Supplementary File 1 - Metabolite Concentrations.xls) containing the measured metabolite concentrations.

Protein Concentration Ratios Spreadsheet (Supplementary File 2 - Protein Abundance Differences.xls) containing the 84 measured protein concentration differences.
6. Concluding Remarks

6.1. Summary

Systems biology builds on experimental and computational (modeling) tools that capture a large-scale perspective to analyze and ultimately understand the behavior of a biological system. This work contributed to both sides - the model-based computational data analysis and experimental analysis: On the computational side, thermodynamic constraints were incorporated into metabolic models. First, thermodynamic principles were used to assist the reconstruction of genome-scale stoichiometric metabolic models. Here, reaction directions were systematically assigned based on thermodynamic feasibility. Second, the model-based analysis tool, NET analysis, has been developed for interpretation of quantitative metabolomics data. Here, besides being a quality check for quantitative metabolomics data, NET analysis exploits Gibbs energies of reaction to suggest putative regulatory sites. Third, to analyze metabolic fluxes thermodynamics and mass balances were linked by coupling NET analysis to EFM analysis. By this approach, unknown reaction activities are inferred from a quantitative metabolomics data set. The developed tools and algorithms were made publicly available [102, 103, 213] or publications are in preparation.

On the experimental side, quantitative large-scale data sets were used to characterize metabolic operation and regulation in \textit{S. cerevisiae}. First, during steady-state growth on glucose in excess, two \textit{S. cerevisiae} strains displayed significant differences in the amounts of metabolic proteins and metabolites. To explain these quantitative differences as well as additional observations on the metabolic state of mutant strains, distinct interaction strengths of signaling and regulatory compounds were proposed. Second, the changes of
metabolite and protein concentrations as well as extracellular rates that occur during the diauxic shift from glucose to ethanol growth were monitored. So far, the generated data was only partly exploited in this work (i.e. in Chapter 4). The acquired data presents a quantitative and dynamic data basis for future analysis of the yeast diauxic shift in a computational systems-biology approach.

6.2. Discussion

Systems biology aims at understanding the complex and nonlinear biological networks that give rise to the emerging phenotype. After one decade of intensive systems biology research and significant achievements, we still have only a limited understanding of biological system behavior [94]. In the background of the experiences made in each sub-project of this work, systems biology’s current standing and deficiencies are discussed in the following with regard to the single steps of the systems biology approach (Fig. 1.1)

6.2.1. Large-scale Data for Gaining a Holistic Perspective

Today, omics techniques provide a wealth of large-scale data (transcriptome, fluxome, metabolome, and proteome data) such that we are about to gain a near to system-level perspective in qualitative manner [86]. Through recent advances of the analytical techniques, also large-scale quantitative data are emerging. In particular, proteomics shifts from semi-quantitative large-scale data to albeit smaller, but truly quantitative data and in metabolomics the number of quantifiable small molecules increases. In this work, the value of this data became evident as based on the quantitative comparison of two S. cerevisiae strains an explanation of the behavior of mutant strains could be proposed in data-driven manner.

Nevertheless, the coverage of quantifiable cellular compounds is still not sufficient. In the proteomics data set to compare two yeast strains, concentrations of most metabolic proteins of the central carbon metabolism but not proteins that are involved in its regulation could be acquired as the latter are less abundant.

Incomplete coverage of the analyzed cellular part is also an obstacle in mechanistics-
based data analysis. Lacking measured metabolite concentrations are one reason that the significance of NET analysis is limited: For many reactions only broad ranges of possible Gibbs energies can be determined. In addition, the measured quantities are usually averaged over the whole cell disregarding that the concentration can be very unequally distributed over different compartments.

As a second obstacle to mechanics-based analysis, these data analysis approaches are typically sensitive to systematic errors in the experimentation since the quantified amounts translate directly to the model parameters. For instance, the NET analysis tool may not find a feasible solution due to one wrongly quantified metabolite concentration. Especially in metabolomics, rapid quenching that immediately stops any metabolic activity is a crucial step to gain correct results.

Besides there requirements to provide complete and accurate data sets, another challenge to experimentation is the acquisition of data in high time resolution to enable dynamic analyses. For studying metabolic reprogramming in yeast, metabolite concentrations were measured down to approx. 15 min resolution. As the turnover rates for metabolite concentrations are however often significantly smaller, this time resolution is not sufficient to estimate parameters of enzyme kinetics. For better dynamic resolution, automation of sampling and sample preparation is required to be able to handle the enormous amount of samples. In the field of metabolomics, automated sampling devises were already developed that allow for time-resolution of a few seconds [107, 188].

6.2.2. Data Analysis for Inferring New Insights and Knowledge

The novel quantitative data provides the opportunity but also requires to go beyond statistical analysis, and hence, analyze the data using mechanistic models. Today, we yet lack general applicable mechanics-based approaches that are suitable to analyze incomplete and cell-averaged time-snapshot data sets. Despite the insights that were gained from the quantitative data sets for the strain comparison, the information provided by the acquired data was not exploited to full extent. For example, measured protein and metabolite concentrations were analyzed integratively only to a limited,
6. Concluding Remarks

non-model based extent. Here, kinetic models that connect proteins with metabolites by mathematical terms are not suitable due to lacking or insufficient dynamic information. Also for the yeast diauxic shift data set, the time-resolution is not sufficient to estimate parameter of a kinetic model.

This thesis contributed to tackle this shortcoming of current data analysis approaches by drawing on thermodynamic principles. The NET analysis is a generally applicable mechanics-based approach to analyze quantitative metabolomics data: It translates concentrations to Gibbs energies and thus relates it not to other concentrations of a reference state but to general physical quantities. This approach is suitable for large-scale analysis and widely applicable as it requires only few organism-specific information. To limited extent, it is also able to resolve localization of metabolites to the different compartments and to infer concentrations of non-measured metabolite concentrations.

One today unresolved challenge to general applicable data analysis is that causalities can only be hypothesized since the model assumptions are limited to general or very few organism specific information. For instance, in NET analysis the regulatory sites in metabolism are only putative. Using a mechanics-based kinetic model, the quantified model parameters would pinpoint the flux controlling reactions and regulatory compounds acting on them.

Hence, the development of generally applicable model-based data analysis concepts that assist to understand the cause-effects-chains within the analyzed biological system lags behind the advances to generate quantitative data that is today provided by modern mass spectrometry techniques.

6.2.3. Model-based Knowledge Integration for Achieving System-level Understanding

Ultimately, systems biology strives to achieve a system-level understanding. For this, dynamic and quantitative descriptions, i.e. kinetic models in terms of ODEs, of all compounds and interactions are required. Today, this is only possible on small scale by formulating kinetic models. Here, time-resolved, quantitative data is necessary (i)
to determine proper model structure to describe the reaction mechanisms and (ii) to estimate the model parameters. Up to now, the coverage of the quantitative data sets is not satisfactory for this. Also the necessary time resolution challenges experimentation as already described above (i.e. sampling and sample processing should be automated). Besides these experimental difficulties, model discrimination and parameter estimation for such an envisioned large-scale model are a hard to solve computational challenge. Though being a yet available concept for mechanics-based knowledge integrations, the difficult scaling-up of models describing each reaction step with mechanics-based terms hampers their application.

Alternatively, to reduce experimental and computational effort to feasible magnitude, models can be restricted to crucial compounds and interactions that are sufficient to describe the system behavior. In contrast to a top-down approach for which concepts are available to reduce an existing model to smaller size (e.g. [44, 47]), model reconstruction that focuses only on the essential parts \textit{a priori} is incomparably more difficult. Here, already in the large-scale data analysis appropriate tools should identify the essential parts of the system that have to be included in the model. For instance, reactions that are identified to be at equilibrium using NET analysis most likely do not need to be modeled kinetically. Here, the intimate link via modeling of data analysis and knowledge integration becomes evident: The lack of general applicable data analysis approaches to suggest causal interrelationships as described above handicaps the model construction.

In conclusion for model-based knowledge integration, there are concepts for detailed dynamic, quantitative modeling available (e.g. ODEs). However, the reconstruction of such models on large-scale comes along with tremendous experimental and computational efforts. To reduce these efforts but still being able to describe the system’s behavior, novel data analysis tools that \textit{a priori} pinpoint the essential parts of the biological network which a model must include are necessary.

Summarizing, emerging analytical techniques and experimentation enable the necessary shift from qualitative to quantitative and dynamic large-scale data sets to monitor
biological processes. For the analysis of the such data sets, generally applicable tools have to be developed that take the available knowledge into account and infer biological insights from the data. Insights obtained by such concepts will assist the model-based knowledge integration by focussing on the essential compounds and interactions. Hence, the such constructed models do not consider each single interaction or reaction, but the crucial elements and events underlying the biological system’s behavior.

Since systems biology strongly relies on mathematical models for describing and understanding complex biological systems, it differs from other biological disciplines. Such an approach does not first of all focus on the experimental verification of proposed physical interactions but the construction of models that are in-line with current knowledge and experimental observations and allow for comprehension of the complex nature of life.
A. Appendix

A.1. Experimental Analysis of the Metabolic Reprogramming During Diauxic Shift in S. cerevisiae

A.1.1. Introduction

The diauxic shift of S. cerevisiae - namely the utilization of produced by-products once the primary carbon source is depleted - inherently involves a switch in metabolism from glycolytic to gluconeogenetic flux through the Emden-Meyerhoff pathway that is associated with rearrangements in carbon storage and energy metabolism [51]. These drastic metabolic changes need to be administered by regulatory mechanisms that respond to the changes in environmental conditions, e.g. availability of carbon sources.

The yeast diauxic shift has been in the focus of several studies to reveal the regulatory actions that underlie this metabolic reprogramming. Tremendous changes in the gene expression have been observed [27, 49, 69]. Also, the relevance of certain regulatory components on the diauxic shift and the expression of proteins before and after the shift have been determined [24, 25, 70]. Such, many important players in the regulatory machinery are known.

So far, the effects in metabolism during the transition, i.e. changes in metabolite and protein concentrations, have however not been assessed in time-resolved manner. In a first step towards elucidating the temporal and causal series of regulatory events acting on the metabolism, we monitored the extra- and intracellular metabolism. Specifically, biomass and extracellular metabolite concentrations were determined next to intracellular concentrations of metabolites and metabolic proteins before, during, and after the
transitions from growth on glucose to growth on ethanol.

A.1.2. Experimental Results

Extracellular Physiology and Growth Phases  We first characterized the extracellular physiology of \textit{S. cerevisiae} batch cultures that pass through the different phases of the diauxic shift. Time-courses of biomass as well as glucose, ethanol, acetate, pyruvate, glycerol, and succinate concentrations were monitored in three biological replicates (Fig. A.1). Fitting these concentrations with splines, uptake and production rates were determined throughout the experiment. Distinct phases of the experiment were determined as periods during which the rates stayed constant (Fig. A.2).

Intracellular Metabolite and Protein Concentrations  Intracellular metabolite and protein concentrations were applied to quantitatively monitor metabolic events inside the cells. As the turnover rates are higher for metabolite pools than for protein pools, metabolite concentrations were determined each 15 to 40 minutes (higher time-resolution during the transition phase) while only six samples were taken in each experiment to determine protein abundances (two in glucose growth phase, two in the transient phases and two in ethanol growth phase). The metabolite concentrations which are shown in Figs. A.3 - A.5 stay to large extent constant over the growth phases. The protein samples not processed and analyzed yet. However, the multiple reaction monitoring (MRM) transitions are already validated and the targeted proteins are shown in Fig. A.6.

A.1.3. Outlook

The here acquired dynamic and quantitative data represents a basis for a detailed model-based analysis of the yeast diauxic shift. In reproducible controlled experiments, we could obtain reproducible results which now gives a bases for detailed analysis of the metabolic reprogramming. Based on the extracellular rates, distinct phases (i.e. glucose growth phase, two transient phases, and ethanol growth phase) of the shift were distinguished. In the analysis introduced in Chapter 4, the metabolite concentrations in the two growth phases were employed to analyze the metabolic fluxes in the two
steady-state growth phases. The key question for further analysis is to elucidate the events that encompasses the transition between these growth phases. For instance, the time-course of observed events in protein and metabolite concentrations can indicate causal interrelationships. Here, a model-based computational systems biology approach can lead to a better understanding of the metabolic reprogramming as it would comprehensively capture the various interactions of metabolites, proteins and higher-level regulatory compounds.
Figure A.1. Extracellular concentrations of *S. cerevisiae* during the diauxic shift. The different colors indicate three biological replicates.
A.1. Experimental Analysis of the Diauxic Shift in \textit{S. cerevisiae}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{diauxic_shift_diagram.png}
\caption{Uptake and production rates in distinct growth phases of \textit{S. cerevisiae} during the diauxic shift.}
\end{figure}
Figure A.3. Intracellular concentrations of central carbon metabolism intermediates in *S. cerevisiae* during the diauxic shift. The different colors indicate three biological replicates.
A.1. Experimental Analysis of the Diauxic Shift in *S. cerevisiae*

Figure A.4. Intracellular concentrations of amino acids in *S. cerevisiae* during the diauxic shift. The different colors indicate three biological replicates.
Figure A.5. Intracellular concentrations of co-factors in *S. cerevisiae* during the diauxic shift. The different colors indicate three biological replicates.
Figure A.6. Targeted proteins within central carbon metabolism in *S. cerevisiae* used in MRM-based proteome analysis.
A. Appendix

A.1.4. Materials and Methods

Strain Used and Experimental Conditions

*S. cerevisiae* FY4 was grown in minimal defined medium with glucose as sole carbon source. The medium was prepared from autoclaved salt and glucose solutions and sterile filtered solutions of vitamins and trace metals to reach concentrations as described by Verduyn et al. [197]. 50 ml precultures were inoculated colonies on YPD plates and grown in 500 ml shake flasks at a glucose concentration of 10 g/l. A fermenter (Bioengineering AG) with a working volume of 2 l minimal medium containing 5 g/l glucose was inoculated by 5 ml preculture from mid-exponential phase (initial OD 0.01-0.02). The fermentation broth was stirred at 1200 rpm, aerated with 1 vvm air, and its temperature maintained at 30 °C. The pH was controlled at a value of 5 by addition of 2 M KOH solution. To reduce ethanol evaporation, the offgas stream was led through a water cooled condenser and the precipitate recycled. If necessary, a few drops of a 20% PEG 2000 solution was added as antifoaming agent. Biological independent cultures were carried out in triplicate.

Biomass and Extracellular Concentrations

Biomass concentrations were monitored by measuring optical density (OD) at 600 nm with a spectrophotometer (Pharmacia Novaspec II). The correlation of OD to biomass dry weight \(0.486 \frac{g_{DW}/l}{OD}\) had previously been determined from a batch culture grown on 10 g/l glucose minimal medium. In these experiments, 15 ml of culture broth at five ODs ranging from 0.3 to 3.3 were filtered using cellulose nitrate filters with a pore size of 0.45 μm (Sartorius AG, Goettingen, Germany) and the dried filters were weighted. To determine the extracellular concentrations of glucose, ethanol, acetate, pyruvate, succinate, and glycerol, 1 ml samples were taken and centrifuged for 4 min at 4000 rpm at 4 °C. The supernatant was analyzed with an HPLC system (Agilent HP1100), equipped with a polymer column (Aminex HPX-87H from BioRad). As eluent 5 mM H\(_2\)SO\(_4\) was used and the column was heated to 60 °C. The compounds were detected and quantified with a refractive index (RI) detector and an UV/Vis-detector (DAD). For absolute
quantification, calibration curves with external standards for the corresponding pure substance obtained from Sigma, Switzerland, were used.

**Intracellular Metabolite Concentrations**

For determination of intracellular concentrations, samples were withdrawn from culture every 15-40 min. To quench metabolism and extract intracellular metabolites, the following procedure was applied (modified from [46, 64]). Four samples of 1-4 ml at each sampling time point and quenched in methanol at -40°C. After centrifuging for 3 min at 14000 rpm and -9°C, they were frozen at -40°C. Intracellular metabolites were extracted by incubation in 75% ethanol for 3 min at 95°C. The supernatant was retained by centrifuging at -9°C, and two samples each were prepared for either LC-MS/MS of GC-TOF analysis.

For quantification by GC-TOF, two sample aliquots were derivatized with either TMS or TBDMS. The samples were separated via GC on a HP5-MS (Hewlett-Packard, length 30m x ID 0.25 x film 0.25 µm) column and injected (CIS4, Gerstel, Germany) for MS analysis to a TOF spectrometer (Pegasus III, Leco, Germany). Detailed information on process parameters are described in Zamboni et al., in preparation. Leco ChromaTOF software (version 2.32) was used for machine control. An autosampler (MPS2, Gerstel, Germany) controlled by Gerstel Maestro software (version 1.2.3.5) was used to provide samples to the GC-TOF system.

For quantification by LC-MS/MS, an Ion-Pairing LC method adapted from Luo et al. [113]. The mobile phase was composed of eluent A (aqueous solution of 10 mM tributylamine and 15 mM acetic acid) and eluent B (methanol); the gradient profile was as follows: $t = 0$ min, 0% B; $t = 15$ min, 55% B; $t = 27$ min, 66% B; $t = 28$ min, 100% B. The end-capped C18 column Synergi Hydro RP, 2.1 x 150 mm, 4 µm particles (Phenomenex, Aschaffenburg, Germany) was employed. The column was equilibrated for 20 min before each injection, the flow rate was 200 µL/min, the column temperature was controlled at 40°C, the injection volume was µL. For tandem MS analysis an Applied Biosystems/MDS Sciex Q 4000 TRAPTM linear ion trap mass Spectrometer.
A. Appendix

(AB/MDS Sciex, Concord, Canada) was coupled to the LC. Analyst software (version 1.4.2, AB/MDS Sciex, Concord, Canada) was used for both machine control and data acquisition. All analyses were performed in negative ion and selected reaction monitoring mode with Q1 and Q3 set to unit resolution. Ion spray voltage, auxiliary gas temperature, nebulizer gas (GS1), auxiliary gas (GS2), curtain gas (CUR) and collision gas (CAD) were set to -4200 V, 650°C, 65, 40, 10, 4 (arbitrary units), respectively. Nitrogen (Pangas, Dagmersellen, Switzerland) was used as curtain and collision gas. Declustering potential (DP), collision energy (CE) and collision cell exit potential (CXP) were optimized separately for each transition. To obtain temporal resolution of greater than 1 Hz for each transition, the run was divided into five segments and the dwell time for each transition was set to 50 ms.

Protein Concentrations

To determine protein concentration, duplicate samples of 25 ml were harvested on ice and washed twice with 5 ml lysis buffer (cf. below). During washing, the protein samples were centrifuged for 5 min at 5000 rpm at 4°C and the supernatant discarded afterwards. Cell pellets were frozen at -80°C.

For protein extraction, the cell pellets were thawed in an ice cold lysis buffer containing 20 mM TrisHCl pH 8.0, 2 mM DTT, 100 mM KCl, 10 mM EDTA, and complete yeast protease inhibitors cocktail (Roche, Mannheim, Germany), using 1 ml of lysis buffer per gram of yeast. Yeast cells were lysed by glass bead beating, and lysed cells were centrifuged to remove cellular debris. The supernatant was transferred to a fresh tube and the protein concentration in the extract was determined by Bradford assay. Proteins were precipitated by adding six volumes of cold (-20°C) acetone and resolubilized in a digestion buffer containing 8M urea and 0.1M NH4HCO3. A 100 µg-aliquot of each yeast protein sample was transferred to a fresh tube and mixed with an equal amount of 15N-labeled yeast proteins (cf. below). To digest the proteins, they were reduced with 12 mM dithiotreitol for 30 min at 35°C and alkylated with 40 mM iodoacetamide for 45 min at 25°C, in the dark. Samples were diluted with 0.1 M NH4HCO3 to a final
A.1. Experimental Analysis of the Diauxic Shift in *S. cerevisiae*

centrution of 1.5 M urea and sequencing grade porcine trypsin (Promega) was added to a final enzyme:substrate ratio of 1:100. The digestion was stopped by acidification with formic acid to a final pH <3. Peptide mixtures were cleaned on Sep-Pak tC18 cartridges (Waters, Milford, MA, USA) eluted with 60% acetonitrile. Peptides were evaporated on a vacuum centrifuge to dryness, resolubilized in 0.1% formic acid and immediately analyzed.

Protein abundances were determined by a multiple reaction monitoring (MRM) based mass spectrometry approach as follows. For developing and validating MRM assays, samples were analyzed on a hybrid triple quadrupole/ion trap mass spectrometer (4000QTrap, ABI/MDS-Sciex, Toronto) equipped with a nanoelectrospray ion source. Chromatographic separations of peptides were performed on a Tempo nano LC system (Applied Biosystems) coupled to a 15 cm fused silica emitter, 75 µm diameter, packed with a Magic C18 AQ 5 mm resin (Michrom BioResources, Auburn, CA, USA). Peptides were loaded on the column from a cooled (4°C) Tempo autosampler and separated with a linear gradient of acetonitrile/water, containing 0.1% formic acid, at a flow rate of 300 nl/min. A gradient from 5 to 30% acetonitrile in 30 or 60 min was used. The mass spectrometer was operated in MRM mode, triggering acquisition of a full MS2 spectrum upon detection of an MRM trace (threshold 300 ion counts). MRM acquisition was performed with Q1 and Q3 operated at unit resolution (0.7 m/z half maximum peak width) with an average of 100 transitions (dwelltime 20 ms/transition) per run. MS2 spectra were acquired in the trap mode (enhanced product ion) for the two highest MRM transitions, using 100 ms fixed fill time, Q0 trapping enabled, Q1 resolution low, scan speed 4000 amu/s, m/z range 300-1300, 2 scans summed. Quantitative analysis in MRM-mode was performed with Q1 and Q3 operated in unit resolution (0.7 m/z half maximum peak width). Collision energies for both MRM and MS2 analyses were calculated according to the formulas: $CE = 0.044 \cdot m/z + 5.5$ and $CE = 0.051 \cdot m/z + 0.5$ ($CE$: collision energy, $m/z$: mass-to-charge ratio of the precursor ion) for doubly and triply charged precursor ions, respectively.

For the quantitative analysis, a $^{15}$N-labeled yeast digest was derived from a yeast
A. Appendix

batch culture that displayed diauxic growth on minimal medium with 20 g/l glucose and $^{15}$N-labeled ammonium as nitrogen source. To gain high coverage of metabolic proteins, aliquots from the different phases (growth on glucose, transient phase and growth on ethanol) of this experiment were mixed. The such heavy labeled protein mixture was used as an internal standard and was spiked into each sample at a 1:1 ratio before digestion. The three most intense transitions from each validated peptide and for the corresponding $^{15}$N-labelled analogue were chosen for quantitative analysis in MRM mode. Quantitation was performed in scheduled-MRM mode, using a prototype acquisition software which features the acquisition of transitions restricted to retention time windows. Peak height was determined with MultiQuant 1.0.0.1 software (Applied Biosystems/MDS Sciex) after confirming for each peptide the co-elution of all 6 transitions. Peak height ratios of the sample transition and internal standard transition was corrected for spray efficiency and ionization differences between runs.
References


References


References


135


References


References


References


References


References


List of Publications

Peer-reviewed Articles

Zamboni N., Kümmel A., Heinemann M.
\textit{anNET: a tool for network-embedded thermodynamic analysis of quantitative metabolome data}

Kümmel A., Panke S., Heinemann M.
\textit{Systematic assignment of thermodynamic constraints in metabolic network models}

Kümmel A., Panke S., Heinemann M.
\textit{Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data}
Molecular Systems Biology 2, 2006, doi:10.1038/msb4100074
(\textit{featured as 'News & Views' in Molecular Systems Biology by S. van Dien and C. H. Schilling, doi:10.1038/msb4100078})

Heinemann M., Kümmel A., Ruinatscha R., Panke S.
\textit{In silico genome-scale reconstruction and validation of the Staphylococcus aureus metabolic network}

Heinemann M., Kümmel A., Giesen R., Ansorge-Schumacher M. B., Büchs J.
\textit{Experimental and theoretical analysis of phase equilibria in a two-phase system used for biocatalyzed esterifications}

Oral Presentations at Conferences

Kümmel A., Heinemann M.
\textit{Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data}
1st YSBN Workshop, Vienna, Austria, 16.-18.11.2006.
References

Poster Presentations at Conferences

Kümmel A., Ewald J., Zamboni N., Heinemann M.
Elucidating compartmentalized flux direction patterns in yeast
8th International Conference on Systems Biology, Long Beach, 03.-05.10.2007.

Kümmel A., Heinemann H., Zamboni N.
anNET: a tool for quality check and integrated analysis of metabolome datasets

Kümmel A., Panke S., Heinemann M.
Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data
6th D-BIOL Symposium ETH, Davos, 24.-26.05.2006

Kümmel A., Panke S., Heinemann M.
Model-based integration of metabolome and fluxome data by network-constraint thermodynamic analysis

Kümmel A., Schümperli M., Heinemann M., Panke S.
Design of a system of biotransformations by means of stoichiometric network analysis
12th European Congress on Biotechnology, Copenhagen, Denmark, 21.-24.08.2005.

Kümmel A., Heinemann M., Büchs J., Meinberg H., Göke V.
Model-based experimental analysis of the diffusive mass transfer of butanol in hydrogel beads
Workshop SFB540/SPP1105, Aachen, 11. - 12.06.2003

Non-reviewed Articles

Panke S., Kümmel A., Schümperli M., Heinemann M.
Industrial multi-step biotransformations
Curriculum Vitae

Anne Maria Rita Kümmel

Date/place of birth: 07/18/1977/Höxter, Germany
Citizenship: German
Sex: Female

Education

since 03/08 Post-Doctoral Researcher; Novartis Pharma AG, Basel; Multivariate Data Analysis for High-content Screening

03/04 - 03/08 PhD studies, Institute of Process Engineering (Bioprocess Laboratory, 03/04 - 03/06) and the Institute of Molecular Systems Biology (since 04/06 - 03/08), ETH Zurich, Model-based and Experimental Approaches for Understanding Microbial Metabolism at Systems Level supervised by Dr. M. Heinemann

10/97 - 10/03 Diplom-Ingenieur (M.Sc. in Mechanical Engineering) RWTH Aachen (majoring biochemical engineering) Diploma thesis: Model-based Experimental Analysis of Diffusion in Hydrogels Based on Raman Spectroscopy supervised by M. Heinemann Grade: 1.3 (very good)

• 1st study thesis at the Department of Biochemical Engineering: Development of a Fluorescence Measuring System to Determine the Distribution of Fluid in Shaking Flasks

• 2nd study thesis at the Helmholtz-Institute for Biomedical Technologies: Blood Trauma by Time Variable Fluid Tension: Modular Charge Concept and Systematic Experimental Design

08/00 - 05/01 Abroad study at the KTH, Stockholm; Erasmus scholarship

08/88 - 06/97 Secondary school: Gymnasium Brede, Brakel; Graduation: Abitur (university entrance qualification): Grade 1.3 (very good)
References

Work Experiences

<table>
<thead>
<tr>
<th>Internships</th>
<th>BASF AG, Ludwigshafen: unit operations in bioseparation (10/02 - 12/02)</th>
</tr>
</thead>
</table>
| Student research assistance | Department of Biochemical Engineering: modelling of phase equilibria (10/01 - 06/02) and measurement of oxygen transfer rates (06/99 - 12/99)  
Institute for Machine Elements and Machine Design: tutor for course exercises (05/00 - 07/00) |

Skills

<table>
<thead>
<tr>
<th>Experimental techniques</th>
<th>Fermentation technique; Culturing of yeast; Sampling for proteome and metabolome analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computational techniques</td>
<td>Computational flux analysis; Optimization and parameter estimation; Dynamic modelling</td>
</tr>
<tr>
<td>Languages</td>
<td>German (mother tongue); English (fluent); Swedish (good)</td>
</tr>
<tr>
<td>EDV</td>
<td>Software: Matlab, Microsoft Office, gPROMS, Origin, Corel Draw, LaTeX; OS: Microsoft Windows, Linux</td>
</tr>
</tbody>
</table>

Additional Experiences and Interests

| Organization and leading of scout youth groups |
| Outdoor activities (Climbing, Hiking, Cycling) |