A Bilevel Optimization Algorithm to Identify Enzymatic Capacity Constraints in Metabolic Networks—Development and Application

by

Laurence Yang

A thesis submitted in conformity with the requirements for the degree of Master’s of Applied Science
Graduate Department of Chemical Engineering and Applied Chemistry
University of Toronto

Copyright © 2008 by Laurence Yang
Abstract

A Bilevel Optimization Algorithm to Identify Enzymatic Capacity Constraints in Metabolic Networks—Development and Application

Laurence Yang

Master’s of Applied Science
Graduate Department of Chemical Engineering and Applied Chemistry
University of Toronto
2008

Constraint-based models of metabolism seldom incorporate capacity constraints on intracellular fluxes due to the lack of experimental data. This can sometimes lead to inaccurate growth phenotype predictions. Meanwhile, other forms of data such as fitness profiling data from growth competition experiments have been demonstrated to contain valuable information for elucidating key aspects of the underlying metabolic network. Hence, the optimal capacity constraint identification (OCCI) algorithm is developed to reconcile constraint-based models of metabolism with fitness profiling data by identifying a set of flux capacity constraints that optimally fits a wide array of strains. OCCI is able to identify capacity constraints with considerable accuracy by matching 1,155 in silico-generated growth rates using a simplified model of *Escherichia coli* central carbon metabolism. Capacity constraints identified using experimental fitness profiles with OCCI generated novel hypotheses, while integrating thermodynamics-based metabolic flux analysis allowed prediction of metabolite concentrations.
Acknowledgements

My thanks go out to the many individuals that supported me during my life as a graduate student. First and foremost, my supervisors, Professor Mahadevan and Professor Cluett, have provided me with a unique opportunity to fully embrace and pursue my intellectual inclinations. Furthermore, their kind guidance, acute insight, and encouraging presence helped me to always maintain clear view of my academic goals.

I have also been fortunate to be surrounded by many good friends and colleagues, beginning with Nikolaos, an irreplaceable part of graduate life by now, having shared the lab for almost two years now. I am unlikely to forget the lively discussions I had with Kai on topics ranging from non-equilibrium thermodynamics to martial arts. Karthik regularly intrigued me with detailed explanations of experimental techniques, a side of research that I otherwise would have been completely ignorant of, while Ahsan routinely engaged me in fascinating discussions of his latest research and computational systems biology methods. Likewise, I would like to thank all the members of the Laboratory for Metabolic Systems Engineering for their support and helpful knowledge. Also, Francis in our Process Control Group not only taught me about computer hardware and operating systems alike, but was always ready to lend a helping hand in any issues concerning the lab—to him I owe much thanks.

I extend my thanks to friends and family outside the lab, including Yaser, who was instrumental in helping me to improve upon my intellectual outlook by recommending deeply influential publications and through our discussions over expensive dinners. To my parents I owe great thanks, for their unshakeable trust in all my endeavours and support, even despite their being overseas.

Finally, I gratefully acknowledge financial support from the Natural Sciences and Engineering Research Council of Canada and from the University of Toronto [1].
Contents

Abstract ii

Acknowledgements iii

Table of Contents iv

List of Tables vi

List of Figures vii

1 Introduction 1

1.1 Literature Review ................................. 1

1.2 Statement of Objectives ............................. 4

2 The Optimal Capacity Constraint Identification (OCCI) Algorithm 7

2.1 Bilevel optimization for identifying enzymatic capacity constraints .... 7

2.2 Reformulation into a mathematical program with complementarity con-

straints (MPCC) ........................................ 10

2.3 Reformulation to improve convergence ......................................... 12

2.4 Bound Variability Analysis ............................. 12

3 Methods for Investigating Capabilities of the OCCI Algorithm 14

iv
3.1 Evaluating the capability of the OCCI algorithm for identifying capacity constraints ........................................ 14
3.2 Improving OCCI predictions with multiple fitness measurements ........................................ 17
3.3 Investigating the variability of capacity constrained fluxes ........................................ 19
3.4 Variability of the bounds on capacity constrained fluxes ........................................ 19
3.5 Effect of flux coupling characteristics on OCCI predictions ........................................ 20

4 Predictive Capabilities of the OCCI Algorithm ........................................ 23

4.1 Identification of a capacity constraint based on a single growth rate measurement ........................................ 23
4.2 OCCI predictions improve with multiple fitness measurements ........................................ 26
4.3 Variability of capacity constrained fluxes ........................................ 28
4.4 Decreasing bound variability with increasing number of strains ........................................ 30
4.5 Incorporation of flux coupling characteristics improves OCCI predictions ........................................ 32

5 OCCI, Thermodynamics, and Experiments ........................................ 35

5.1 Model reconciliation with experimental data ........................................ 36
5.2 Reverse PntAB flux may improve fitness under certain genetic backgrounds ........................................ 43
5.3 Flexible function of the membrane-bound transhydrogenase ........................................ 44
5.4 Metabolite Activities Predicted from Thermodynamic Analysis ........................................ 47

6 Conclusions and Recommendations ........................................ 51

7 Appendix ........................................ 56

7.1 Reformulating TMFA as a fractional program ........................................ 56
7.2 The metabolic reaction network model ........................................ 58

Bibliography ........................................ 66
List of Tables

5.1 Enzymatic Capacity Constraints ........................................... 39
5.2 Thermodynamically feasible NAD(P)(H) concentration ratios ........ 49
7.1 List of metabolic reactions used in this work ............................ 58
7.2 List of metabolite definitions used in this work .......................... 63
List of Figures

1.1 (A) Conceptual framework of the OCCI algorithm. (B) Simplified network model of *Escherichia coli* central carbon metabolism [14] where A to H, ATP, O2, NADH, Dext, Eext are metabolites, and italicized labels (*Tc*1, *Growth*, etc.) correspond to reactions. 5

3.1 Coupling characteristics of the simplified network model of *E. coli* central carbon metabolism where (A) shows the four fully coupled flux sets (*FC*1, *FC*2, *FC*3, *FC*4) and (B) shows the partially coupled flux set (*PC*1). In both cases, thick lines represent fluxes in a coupled set. 21

4.1 Flux distributions, shadow prices, and flux bounds for (A) FBA with no artificial capacity constraints, (B) FBA with an artificial capacity constraint at flux index 9, and (C) OCCI predictions made using the growth rate from (B). In all three figures, flux 20 corresponds to growth rate, white bars indicating shadow prices on lower bounds are to the left of their corresponding flux indices, while black bars indicating upper bound shadow prices are to the right. 24

4.2 FVA of the simplified network model of *E. coli* central carbon metabolism for a fixed growth rate (flux 20) equal to that of the constrained strain. Maximum fluxes of unbounded value are not shown. 26
4.3 Predictive capability of OCCI versus the number of growth phenotype measurements available. Scoring metrics are defined in Eqs. (3.1), (3.2), (3.3). ......................................................... 27

4.4 Percentage of artificially capacity constrained fluxes with fixed value (i.e., invariable), subject to a fixed growth rate for one to 14 strains in a set. A larger percentage of the capacity constrained fluxes are invariable as the number of strains increases. ............................................. 29

4.5 Bound variability analysis performed on the simplified network model for one to 14 strains. Bound variability is defined as the difference between maximum and minimum values a bound may have, subject to a specified growth rate. The variability of upper bounds on the eleven investigated fluxes decreases as the number of growth phenotypes that must be reconciled simultaneously increases. Ordinate limits are chosen to include the largest variability values that are less than 900. ............................................. 31

4.6 Improved predictive capability of OCCI versus the number of growth phenotype measurements available when coupling characteristics of the metabolic network are known. Scoring metrics are defined in Eqs. (3.1), (3.2), (3.3). ......................................................... 33

5.1 Relative growth rates \( \frac{\mu_{\text{ICD-NAD}}}{\mu_{\text{ICD-NADP}}} \) measured by [74] (black lines indicating experimental error) and predicted by FBA without (a) and with (b) enzymatic capacity constraints (ECCs). Bars above and below ”x” in (b) indicate upper and lower prediction uncertainties, respectively. ............... 38

5.2 Relative fitnesses of acetate-grown \( E. \ coli \) under additional genetic backgrounds predicted with FBA constrained by the ECCs identified by OCCI. Relative fitness = \( \frac{\mu_{\text{ICD-NAD}}}{\mu_{\text{ICD-NADP}}} \), where \( \mu \) denotes growth rate. Predictions span a range due to computational uncertainty in identifying the global optimum. 41
5.3 FBA predictions of intracellular flux of $icd^{NADP}$ strains, constrained by OCCI-predicted ECCs under (a) wild-type and (b) $\Delta zwf$ genetic backgrounds. .................................................... 42

5.4 Growth rates and fluxes of $icd^{NADP}$ E. coli strains under a wild-type genetic background, simulated using FBA. $\mu$ denotes growth rate. ......... 44

5.5 Sensitivity of predicted relative fitnesses when enzymatic capacities of forward and reverse fluxes of PntAB are relaxed. Of the genetic backgrounds studied in [74], those with $pntAB$ intact are investigated here. ........... 45

5.6 Cartoon schematic depicting PntAB enzyme structure, based on [4, 33]. 46

6.1 Schematic of the OCCI algorithm as part of an iterative process of model reconciliation and experimental validation, which allows for focused investigation of the key metabolic enzymes and pathways. ................. 55
Chapter 1

Introduction

1.1 Literature Review

Networked systems of high complexity are observed at many levels of biological systems, from gene interactions, protein signalling and transcriptional regulation, metabolic reaction networks, to cell-cell interactions. With the advent of systems biology, high-throughput methods are being used to generate global maps of system processes at many of these levels, such as genetic interactions [67], protein-protein interactions [29], and metabolic reaction networks [2]. In concert with the advances in experimental techniques, increasing effort is being made to develop computational models to interpret these global maps and to possibly extract insight into underlying simplicity and elegance from the complexity [58]. The field of metabolic engineering is one that has seen rapid growth in the past decade by embracing the integration of system-wide experimentation with efficient computational methods. We are now able to reconstruct whole-cell models of metabolic reaction networks consisting of hundreds to more than a thousand variables, as complete genome sequences of an increasing number of organisms are being determined (for an insightful review of current achievements and future directions of genome annotation efforts, see [54]).
These metabolic network models involve more variables than equations due to the high degree of connectivity of the nodes, or metabolites. To obtain a flux distribution within the reaction network, flux balance analysis (FBA) [68] is typically used to obtain a solution via linear programming (LP) with maximization of growth rate as the assumed cellular objective. In the case of *Escherichia coli*, this assumption of maximized growth has been found to be valid under certain conditions [18, 30]. The FBA approach is explained in more detail in Section 2.1.

FBA is an attractive method in that detailed reaction kinetics are not needed to conduct system-level analyses (for a review of FBA, see [19]). In certain instances, however, classical FBA predictions fail to agree with experimental observations. This has motivated the development of various computational frameworks for reconciling constraint-based models with experimental data [27, 55]. A key advantage of constraint-based modeling lies in its ability to narrow down the range of possible solutions by incorporating meaningful constraints. Accordingly, a great deal of successful effort has been made to incorporate various constraints (environmental, physicochemical, regulation) to make experimentally validated model predictions (for a comprehensive review, see [15]).

In particular, flux through each metabolic reaction is limited by the enzyme’s catalytic efficiency, which may vary over many orders of magnitude across different enzymes but is, nonetheless, finite [69]. Consequently, in addition to constraining substrate uptake rates with readily available measurements, all intracellular reactions should also be constrained based on flux measurements to obtain the most biologically consistent model predictions. In fact, previous studies have constrained certain intracellular fluxes based on measurements to fit model predictions with experimental data [42, 65]. Furthermore, Wiback et al. [72] have shown that incorporating intracellular flux constraints dramatically shrinks the solution space, which is crucial for predicting metabolic flux distributions that resemble real physiology. Despite their importance, implementation of finite flux capacity constraints on sets of intracellular reactions, however, has seen relatively little investiga-
tion. This can be attributed to the difficulty of obtaining quantitative measurements of flux bounds or any kinetic information for all intracellular fluxes in the network over a wide variety of conditions.

While direct measurement of such inherent flux capacity constraints may be difficult, other forms of data may be used to indirectly infer such capacity constraints. To this end, we require quantitative measurements of high precision that can provide insight into intracellular processes. In fact, we are witnessing an abundance of such data in recent years, enabled by large-scale phenotyping methods. For example, Zhu et al. [74] obtained highly precise, relative fitness measurements from 16 strains of *E. coli* under two different growth conditions. Their specific choice of deletion strains was hypothesis-driven. These studies formed a physiological basis for the group to identify the cause of an ancient adaptation in a metabolic enzyme of prokaryotes that first occurred about 3.5 billion years ago [74].

Tong et al. [66] developed a technique in which a specific mutant of *Saccharomyces cerevisiae* was crossed with ~4,700 deletion mutants, to identify functional relationships between genes. By extending this approach, they were able to construct a map of genetic interactions consisting of ~1000 genes and ~4000 interactions for *Saccharomyces cerevisiae* [67]. Others have developed strategies involving highly parallel measurements of fitness in growth competition cultures to screen ~700 strains and to identify functionally relevant mutants, such as those with growth defects [40]. Lee et al. [37] utilized a high-throughput method to identify genes that confer resistance to a DNA-damaging agent using genomics. Their method involved the use of barcode microarrays to obtain precise, quantitative measurements of relative fitness for ~4700 deletion strains in *S. cerevisiae*. Yet another group was able to produce quantitative measures of genetic interactions by measuring fitness of 650 double-deletion strains [63]. Clearly, these types of sophisticated, high-throughput methods are capable of generating quantitative data that are rich in information for analyzing processes within intracellular networks. Such recent advances in
Chapter 1. Introduction

4

genome-wide experimental techniques have been a major motivation for the current study.

1.2 Statement of Objectives

We present the development and in silico testing of a computational algorithm that integrates experimental fitness measurements with constraint-based metabolic models to identify flux capacity constraints. We assume that these capacity constraints are inherent within the metabolic network of an organism over a wide variety of genetic backgrounds. This optimal capacity constraint identification (OCCI) algorithm is formulated as a bilevel-optimization problem, as depicted in Fig. 1.2A. The outer problem minimizes discrepancy between experimental data and model predictions of growth rates by identifying capacity constraints, subject to the inner optimization, which is the FBA problem. Such bilevel optimization methods have been formulated to solve problems in the context of metabolic engineering for a large variety of applications [9, 46, 7, 22, 23, 53]. Development of the OCCI algorithm is further motivated in part by the fact that these studies have demonstrated the flexibility with which bilevel optimization problems can be formulated to successfully solve problems of importance to metabolic engineering and systems biology, in general.

In this context, the main contribution of the OCCI algorithm is that it provides a computational framework to systematically generate hypotheses regarding key enzymatic capacities within metabolic networks. Additionally, our mathematical approach to solving the OCCI algorithm is different from the aforementioned bilevel optimization frameworks—our final nonlinear optimization problem is most effectively formulated as a mathematical program with capacity constraints (MPCC) [50] (see Section 2.2), rather than a mixed integer linear program (MILP). Hence, our solution method for the OCCI algorithm also adds to the existing literature [53, 52] for efficient methods to solve MPCC problems.
Chapter 1. Introduction

Figure 1.1: (A) Conceptual framework of the OCCI algorithm. (B) Simplified network model of *Escherichia coli* central carbon metabolism [14] where A to H, ATP, O2, NADH, Dext, Eext are metabolites, and italicized labels (*Tc*, *Growth*, etc.) correspond to reactions.

relevant to constraint-based modeling of metabolic networks.

The overarching hypothesis of this work is that a computational algorithm can be developed to identify and quantitatively define bottlenecks, or the rate-limiting steps, in cellular metabolic reaction networks, based on growth, or fitness, profiles. The objective of this work then, is to develop such computational algorithms, validate their predictive capabilities, and apply them to identify metabolic bottlenecks using actual experimental data.

This study will be presented as follows: the algorithms developed in this work are pre-
sented in Chapter 2. The methodology by which we assess the capabilities of the OCCI algorithm are described in Chapter 3. We present the results of the assessment in Chapter 4. Following this investigation, in Chapter 5 we investigate some intriguing hypotheses from the work of [74], by analyzing their fitness measurements with OCCI. Finally, we provide a conclusive summary of the major findings and significance of this study, along with recommendations for future work in Chapter 6.
Chapter 2

The Optimal Capacity Constraint Identification (OCCI) Algorithm

2.1 Bilevel optimization for identifying enzymatic capacity constraints

Before discussing development of the optimal capacity constraint identification (OCCI) algorithm, the flux balance analysis (FBA) modeling approach is briefly outlined, as it forms the basis of part of the OCCI algorithm.

FBA ([19, 35, 49]) modeling involves predicting flux through reactions, throughout the metabolic network of a cell. To do this, the metabolic network is represented as a matrix containing the stoichiometric coefficients of every reaction. Consequently, the metabolites are ordered along the rows of the matrix, while reactions are ordered along the columns. Each reaction consumes or produces a metabolite, which we assume does not accumulate nor deplete over time. This pseudo-steady state assumption separates the different time scales of cellular processes like metabolism (fast dynamics) and growth (slower dynamics). The pseudo-steady state assumption specifically allows us to treat these faster processes as constant, even if the overall system may have yet to reach a steady state over longer
time periods. Hence, the flux through each reaction is constrained such that a metabolite being produced by reactions is balanced by its production by other reactions. Due to the high connectivity of many metabolites, these matrices mostly involve more reactions than metabolites, or more variables (the fluxes) than constraints (mass balancing the metabolites). The fluxes, then, of this underdetermined system are solved for by a linear program in FBA, by assuming an appropriate objective function that reflects cellular behavior. While various objective functions can be used [61], the objective of maximizing growth rate has been shown to accurately describe cellular physiology [18, 30] and is used exclusively in this work.

The OCCI algorithm can be formulated as the following bilevel optimization problem for a single fitness measurement from a strain, $j$:

$$\min_{\mathbf{v}^L, \mathbf{v}^U} \left| c_j^T \mathbf{v} - \mu_j^{\exp} \right|_1$$

subject to

$$\begin{align*}
\max_{\mathbf{v}} & \quad c_j^T \mathbf{v} \\
\text{s.t.} & \quad S \mathbf{v} = 0 \\
& \quad \mathbf{v}^L \leq \mathbf{v} \leq \mathbf{v}^U
\end{align*}$$

FBA (2.1b)

where each $c_j^T$ is the vector of objective coefficients corresponding to strain $j$, $\mu_j^{\exp}$ is the growth rate measurement for strain $j$, $c_j^T$ is the vector of objective coefficients for all strains, and $\mathbf{v}$ is the decision variable vector of the inner optimization problem representing metabolic fluxes. $\mathbf{v}^L$ and $\mathbf{v}^U$ are the decision variables of the outer optimization problem representing lower and upper flux bounds, respectively, on the flux, $\mathbf{v}$. These bounds represent enzymatic capacity constraints on reactions in the metabolic network.

Eq. (2.1a) is the OCCI objective function that minimizes the discrepancy between predicted and measured growth rates, while the inner optimization (Eq. (2.1b)) consists of the FBA problem. The objective function (Eq. (2.1a)) may also be formulated as an $L_1$-norm. In this work, the $L_1$ norm is chosen over a quadratic objective function because, upon reformulation of the objective function, superior convergence of the problem
is observed (comparison not shown). Also, the nonlinear $L_1$-norm objective function of
the outer optimization (2.1a) can be reformulated as the following linear optimization
problem, which may reconcile $j$ fitness measurements by minimizing the sum of their
individual discrepancies, $r_j$:

$$
\min_{\tilde{v}_L, \tilde{v}_U} \sum_{j=1}^{n_{\text{exp}}} r_j \tag{2.2a}
$$

subject to:

$$
c_j^T \tilde{v} - r_j \leq \mu_j^{\text{exp}} \tag{2.2b}
$$

$$
c_j^T \tilde{v} + r_j \geq \mu_j^{\text{exp}} \tag{2.2c}
$$

Each $r_j$ is the absolute difference between the predicted and measured growth rate of each
strain, $j$. The number of strains used for fitness measurements is $n_{\text{exp}}$. The reformulated
objective function is the sum of elements of the new vector, $\tilde{r}$, which is subject to the
new constraints. For simplicity, the inner FBA problem (Eq. (2.1b)) can be reformulated
into its canonical form as follows:

$$
\min \quad \tilde{C} \tilde{x} \\
\text{s.t.} \quad \tilde{A} \tilde{x} \geq \tilde{B} \\
\quad \tilde{x} \geq 0
\tag{2.3}
$$

where the decision variable, $x = \begin{bmatrix} \tilde{v}_f \\ \tilde{v}_r \end{bmatrix}$, has been split up into non-negative forward
and reverse fluxes, $\tilde{v}_f$ and $\tilde{v}_r$, respectively. In Eq. (2.3), the following relations hold:

$$
A = \begin{bmatrix}
S & -S \\
-S & S \\
-I & I \\
I & -I
\end{bmatrix}, \quad B = \begin{bmatrix}
0 \\
0 \\
-v^U \\
v^L
\end{bmatrix}, \quad \tilde{C} = \begin{bmatrix}
-\varepsilon & \varepsilon
\end{bmatrix}
$$

The dual problem to Eq. (2.3) is the following:

$$
\max \quad w^T B \\
\text{s.t.} \quad wA \leq \tilde{C} \\
\quad w \geq 0
\tag{2.4}
$$
where $\mathbf{w}$ is the vector containing decision variables of the dual problem. These dual variables have established interpretations for certain problems, such as of economy [10]. Likewise, dual variables are useful in the OCCI algorithm as they indicate which bounds actively constrain the growth rate. More specifically, these variables describe the change in objective value due to local changes in available resources. Due to the reformulation of the primal problem as in Eq. (2.3), the vector of dual variables can be partitioned according to physical interpretations as follows:

$$
\mathbf{w} = \begin{bmatrix}
\mathbf{w}_{ss}^{1} \\
\mathbf{w}_{ss}^{2} \\
\mathbf{w}_{vU}^{U} \\
\mathbf{w}_{vL}^{L}
\end{bmatrix}
$$

The dual variables $\mathbf{w}_{ss}^{1}$ and $\mathbf{w}_{ss}^{2}$ correspond to metabolite mass balances, while $\mathbf{w}_{vU}^{U}$ and $\mathbf{w}_{vL}^{L}$ correspond to upper and lower flux bounds, respectively. The latter two dual variables depict the change in growth rate due to small variations in the upper and lower bounds. If a dual variable corresponding to the upper bound of a certain flux is positive, then the growth rate will increase as the upper bound increases. Similarly, a positive dual variable corresponding to a lower bound indicates that the growth rate increases as the lower bound becomes more negative. As such, the active upper and lower flux capacity constraints can be identified simply by finding the bounds associated with non-zero dual variables.

2.2 Reformulation into a mathematical program with complementarity constraints (MPCC)

The bilevel optimization problem of Eqs. (2.1a)–(2.1b) can be reformulated into a single-level nonlinear program (NLP) by expressing the inner optimization with its Karush-
Kuhn-Tucker (KKT) conditions as follows:

\[
\min_{x, \psi, \mu, \eta, v_L, v_U} \sum_{j=1}^{n_{\text{exp}}} r_j \\
\text{s.t.} \quad c_j^T \cdot p_x^v \cdot x - r_j \leq \mu_j^{\exp} \quad (2.5a) \\
\quad c_j^T \cdot p_x^v \cdot x + r_j \geq \mu_j^{\exp} \quad (2.5b) \\
\quad Ax \geq B \quad (2.5c) \\
\quad \psi A \leq C \quad (2.5d) \\
\quad \eta x + \psi \mu = 0 \quad (2.5e) \\
\quad v_L \leq p_x^v \cdot x \leq v_U \quad (2.5f)
\]

where \( \mu \) and \( \eta \) are slack variables on the primal (Eq. (2.3)) and dual (Eq. (2.4)) inequality constraints, respectively, and \( p_x^v = [I - I] \). The resulting NLP above can be shown to belong to the general class of mathematical programs with complementarity constraints (MPCCs) [51] and mathematical programs with equilibrium constraints (MPECs) [50]. These classes of problems occur frequently in biological and chemical engineering processes [52].

Another method for reformulating bilevel optimizations into single-level optimizations makes use of the strong duality theorem of an LP [10], which states that at optimality, the primal and dual objective values are equal. This method has been used in the context of constraint-based metabolic models [9, 46, 7]; however, for the OCCI problem, the primal inequality constraint of the inner optimization in Eq. (2.3) includes the decision variables of the outer optimization in the right-hand side bounds \( B \). If the strong duality theorem is used, a bilinear constraint results in the formulation, which offers little advantage over the reformulation with KKT conditions. Complementarity constraints often pose difficulties for conventional NLP solvers and this has motivated development of various algorithms geared towards solving problems with such constraints [53, 51]. The solution approach taken here is to reformulate the complementarity constraint as a penalty function [17] and to solve the resulting problem using robust, off-the-shelf NLP
Chapter 2. The Optimal Capacity Constraint Identification (OCCI) Algorithm

solvers.

2.3 Reformulation to improve convergence

To improve convergence of the NLP solution, the complementarity constraint (Eq. (2.5f)) is replaced by a penalty function that is minimized as part of the following new objective function that replaces Eq. (2.5a):

\[
\min_{x, w, \mu, \eta, \nu^l, \nu^u, r} \quad K_r \sum_{j=1}^{n_{\text{exp}}} r_j + K_{\eta x} (\eta x)^2 + K_{\mu \mu} (w \mu)^2
\]

(2.6)

where \(K_r, K_{\eta x}, \) and \(K_{w\mu}\) are positive penalty parameters \([17]\) for the OCCI objective and each complementarity constraint, respectively. The reformulated MPCC has the above objective and is constrained by Eqs. (2.5b)-(2.5g). For all computations in this work, \(K_r = 1\), while \(K_{\eta x}\) and \(K_{w\mu}\) are large (> \(1 \times 10^3\)) such that the growth rate discrepancy and complementarity constraints in Eq. (2.6) each converge to within \(1\text{E}^{-5}\) of zero.

2.4 Bound Variability Analysis

Although OCCI predicts capacity constraints and the resulting flux distribution, the uniqueness of these solutions needs to be investigated. While methods have been developed for investigating the uniqueness of flux distributions predicted by FBA \([39]\), no algorithms are currently available for determining the range of values that capacity constraints may have for a given optimal objective value (minimal discrepancy between predicted and measured growth rates). Consequently, we develop the following bound variability analysis (BVA) and iteratively determine the alternate optimal values of all
upper and lower bounds on the capacity constrained reactions:

\[
\begin{align*}
\min_{\tilde{x}, \tilde{w}, \tilde{\mu}, \tilde{\eta}, \tilde{v}_i^L, \tilde{v}_i^U} & \quad \text{or} \quad \max_{\tilde{x}, \tilde{w}, \tilde{\mu}, \tilde{\eta}, \tilde{v}_i^L, \tilde{v}_i^U} \quad \tilde{v}_i^L \quad \text{or} \quad \tilde{v}_i^U \\
\text{s.t.} & \quad \sum_{j=1}^{n_{\text{exp}}} r_j = 0 \\
& \quad \tilde{c}_j^T \cdot \tilde{p}_x \cdot \tilde{x} - r_j \leq \mu_j^{\text{exp}} \\
& \quad \tilde{c}_j^T \cdot \tilde{p}_x \cdot \tilde{x} + r_j \geq \mu_j^{\text{exp}} \\
& \quad \Delta \tilde{x} \geq \tilde{B} \\
& \quad \tilde{w} \tilde{A} \leq \tilde{C} \\
& \quad \tilde{\eta} \tilde{x} + \tilde{w} \tilde{\mu} = 0 \\
& \quad \tilde{v}_i^L \leq \tilde{p}_x \cdot x \leq \tilde{v}_i^U
\end{align*}
\]

As in Eq. (2.6), the BVA problem can be reformulated to improve convergence using a penalty function to replace the complementarity constraint (Eq. (2.7g)). Hence, constraints Eqs. (2.7b)-(2.7h) and the following objective describe the BVA problem:

\[
\begin{align*}
\min_{\tilde{x}, \tilde{w}, \tilde{\mu}, \tilde{\eta}, \tilde{v}_i^L, \tilde{v}_i^U} & \quad \tilde{v}_i^L \quad \text{(or} \quad \tilde{v}_i^U, -\tilde{v}_i^L, -\tilde{v}_i^U) + \nonumber \\
& + K_r \sum_{j=1}^{n_{\text{exp}}} r_j + K_{\tilde{\eta} \tilde{x}} (\tilde{\eta} \tilde{x})^2 + K_{\tilde{w} \tilde{\mu}} (\tilde{w} \tilde{\mu})^2
\end{align*}
\]

where minimizing $-\tilde{v}_i^L$ or $-\tilde{v}_i^U$ is equivalent to maximizing $\tilde{v}_i^L$ or $\tilde{v}_i^U$, respectively. $K_{\tilde{\eta} \tilde{x}}$ and $K_{\tilde{w} \tilde{\mu}}$ are sufficiently larger than $K_r$ such that complementarity constraints are satisfied with higher priority than fitting the training set (i.e., minimizing sum of $r_j$). Solving the BVA problem results in maximum and minimum ranges for the upper and lower bound of the flux with index $i$. BVA is used to investigate alternate optimal solutions to the OCCI problem in Section 4.4.
Chapter 3

Methods for Investigating Capabilities of the OCCI Algorithm

3.1 Evaluating the capability of the OCCI algorithm for identifying capacity constraints

To verify the capability of OCCI to identify a capacity constraint, the following methodology is implemented: first, FBA is used to obtain the maximum growth rate of an *in silico* strain using the simplified network model of *E. coli* central carbon metabolism [14] depicted in Fig. 1.2B. Assuming conservation of mass at pseudo-steady state, the model consists of a stoichiometric matrix that is used to find fluxes, $v_i$, of 20 irreversible reactions involving 11 metabolites. For a detailed description of FBA modeling and the solution of underdetermined systems, see [35, 45, 49]. In this simulation of maximum growth rate, upper and lower bounds on the fluxes are all unconstrained ($0 \leq v_i \leq \infty$) except for those corresponding to exchange fluxes and knocked out reactions bound to zero flux. We will refer to this strain with unconstrained fluxes as the wild-type and its flux distribution as the wild-type flux distribution.

An artificial capacity constraint is then applied to one of the fluxes such that a growth rate less than that of the wild-type results. For illustration, flux 9 is chosen and its upper
bound is set to half of the wild-type flux. We will refer to this strain as the constrained strain and its growth rate as the fitness measurement. We use the term “fitness measurement” rather than growth rate, since experimental data may be in the form of relative fitness scores between mutants and wild-type strains, which OCCI is also intended to utilize. Finally, the growth rate and substrate uptake rate of the constrained strain are provided as inputs to the OCCI algorithm. Using the same network model as well as the upper and lower bounds from the wild-type strain, OCCI attempts to match the given suboptimal growth rate by finding a set of capacity constraints on the reactions of the metabolic network. Since the upper and lower bounds originally given to OCCI do not include the artificial capacity constraint applied to the constrained strain, this procedure evaluates the capability of OCCI to identify a capacity constraint given only the growth rate and substrate uptake rate.

We note that since the objective function and mass balance constraints are unchanged, an FBA prediction without flux capacity constraints would overpredict the fitness measurement. Hence, if the growth rate of a real biological system is actively constrained by enzymatic capacities, then the only way to identify this capacity constraint would be to either measure all fluxes and identify the reaction with largest discrepancy between prediction and measurement, or by \textit{a priori} knowledge of capacity constraints. Clearly, a systematic method to rapidly infer the most important enzymatic capacities, such as the OCCI algorithm, would be invaluable in such a scenario.

Three aspects of the OCCI algorithm are evaluated: a) the degree to which OCCI is able to match given fitness measurements, b) the capability of OCCI to accurately identify capacity constraints, and c) the extent to which OCCI is able to make accurate quantitative flux predictions of the capacity constrained reactions. The following metrics are defined for each evaluation:
Chapter 3. Methods for Investigating Capabilities of the OCCI Algorithm

Reconciliation Accuracy (RA) = \frac{\text{Number of correctly predicted fitness measurements}}{\text{Total number of fitness measurements}} \quad (3.1)

Identification Accuracy (IA) = \frac{\text{Number of predicted capacity constraint sets that include the artificial capacity constraint}}{\text{Total number of artificial capacity constraints}} \quad (3.2)

Quantitative Accuracy (QA) = \frac{\text{Number of correctly predicted fluxes of artificially capacity constrained reactions}}{\text{Total number of artificial capacity constraints}} \quad (3.3)

For example, if OCCI correctly matches all given growth rates in a series of tests where different artificial capacity constraints are enforced, RA = 100%. For each incorrect growth rate prediction, a lower score for RA is assigned, according to Eq. (3.1). For each test, OCCI also identifies a set of capacity constraints. If the artificial capacity constraint is included in the set identified by OCCI for all test cases, IA = 100%, whereas, if only one artificial capacity constraint is included in the identified sets for two test cases, IA = 50%. For each test, OCCI also makes quantitative predictions of the flux value corresponding to the artificially capacity constrained reaction. Again, if this prediction is correct for all test cases, QA = 100%, while a lower score is assigned according to Eq. (3.3) when incorrect predictions are made. The fitness predictions in Eq. (3.1) and quantitative flux predictions in Eq. (3.3) are considered correct if their discrepancies with measurements are \leq 1 \times 10^{-5}. 
3.2 Improving OCCI predictions with multiple fitness measurements

Based on the metrics defined in Eqs. (3.1), (3.2), (3.3), we expect limitations in the capability of OCCI to make accurate predictions when a large degree of alternate optimal solutions is present; however, the accuracy of predictions made by the OCCI algorithm can be improved significantly by increasing the number of fitness measurements under diverse genetic backgrounds. To illustrate how increasing the availability of fitness measurements will improve OCCI predictions, the following procedure is followed: first, a capacity constraint is artificially enforced on a chosen reaction in a single mutant strain. The resulting growth rate is provided to OCCI. If OCCI is able to match the given growth rate, a score is assigned towards the Reconciliation Accuracy as in Eq. (3.1). OCCI also identifies a set of capacity constraints. If the artificially imposed capacity constraint is included in the set identified by OCCI, a score is assigned towards the Identification Accuracy (Eq. (3.2)). Subsequently, a capacity constraint is enforced on a different flux in the network and the above procedure is repeated for all appropriate fluxes in the network. Each time, the enforced capacity constraint is defined to be an upper bound equal to half of the flux value predicted by an FBA simulation on the unconstrained mutant strain network. Consequently, reactions that carry no flux to begin with are excluded from the procedure. Also, note that all fluxes in the simplified metabolic model used for simulations are irreversible; therefore, all capacity constraints correspond to upper bounds, whereas, lower bounds would also need to be considered if reversible fluxes are included in the metabolic model.

The above procedure is then applied to simultaneous reconciliation of more than one fitness measurement. For example, two different strains are generated by knocking out different reactions in the network. An artificial capacity constraint is enforced on the same reaction in both strains. The value of the capacity constraint is again equal to half
of the flux value of the unconstrained FBA simulation. In cases where the same reaction in the two strains have initially different flux values, the capacity constraint is defined as an upper bound equal to half of the largest flux value. This single, maximum upper bound is enforced as the capacity constraint across all strains. This is consistent with the assumption of the OCCI algorithm that the identified capacity constraints are the maximum and minimum upper and lower bounds, respectively, of all strains investigated. The growth rates resulting from the two strains are then provided as inputs to OCCI. OCCI then identifies a set of capacity constraints that must be common to both strains while matching the fitness measurements given. If the artificially enforced capacity constraint is included in the identified set a score is assigned towards the Identification Accuracy (Eq.(3.2)). Again, this procedure is repeated by enforcing capacity constraints on all non-zero fluxes.

For each artificial capacity constraint, the capability of OCCI to make quantitative predictions of the flux value of capacity constrained reactions is also investigated. Hence, the flux value of the artificially capacity constrained reaction from the FBA simulation is directly compared with that predicted by OCCI. If the flux values are identical, a score is assigned towards the Quantitative Accuracy (Eq.(3.3)). Finally, the entire procedure is iterated for one to fourteen simultaneous strains, which are all grown solely on “Carbon1” (Fig. 1.2B). These fourteen strains correspond to one “wild-type” strain and thirteen “mutant” strains that each have a different reaction “knocked-out” (constrained to zero flux) in the metabolic network. These in silico strains are analogous to gene deletion strains for experimental fitness assays. The thirteen “knocked-out” reactions were chosen out of 20 reactions available in the network model by excluding six reactions that include “lethal” deletions (zero growth flux if reaction is constrained to zero) and blocked reactions [8]. For each set of strains, we iterate over eleven conditions for a total of 1,155 in silico experiments. The eleven reactions to be artificially capacity constrained are chosen by excluding those with zero flux in all strains.
3.3 Investigating the variability of capacity constrained fluxes

In FBA simulations of cellular metabolism, the objective of maximal growth rate can sometimes be satisfied by different flux distributions. Each of these flux distributions is an alternate optimal solution to the optimization problem of maximal growth rate. A method called flux variability analysis (FVA) has been developed by [39] for analyzing the effect of alternate optimal flux distributions of metabolic network models. This method determines the maximum and minimum fluxes of all reactions in a metabolic network that result in an equivalent, maximal growth rate. FVA is used in this work to determine the flux variability of capacity constrained fluxes subject to a single growth rate in Section 4.1. Additionally, we investigate the change in flux variability of capacity constrained reactions as an increasing number of growth rates must be satisfied simultaneously in Section 4.3. Specifically, we identify the number of capacity constrained reactions that must have a fixed flux (invariable capacity constrained flux) to satisfy a given set of growth rates across multiple strains.

3.4 Variability of the bounds on capacity constrained fluxes

Just as the existence of alternate optimal flux distributions is an important consideration for constraint-based model simulations using the FBA framework, the investigation of alternate optimal capacity constraints may provide additional insight when solving the OCCI problem. To quantify the variability of bounds subject to minimizing discrepancy between predicted and measured growth rates, we apply the the BVA algorithm described in Section 2.4. In Section 4.4, we calculate the variability of bounds on all eleven capacity constrained fluxes for one to fourteen strains.
Chapter 3. Methods for Investigating Capabilities of the OCCI Algorithm

3.5 Effect of flux coupling characteristics on OCCI predictions

In this section, we demonstrate how an existing framework for analyzing topological characteristics of metabolic networks is used to complement the OCCI algorithm and improve its predictive capability. Functional consequences of connectivity in complex metabolic reaction networks are often difficult to interpret intuitively, especially for large-scale models. Consequently, various methods have been developed to elucidate the connectivity of such networks [60, 62, 44]. In particular, flux coupling analysis of metabolic reaction networks has proven to be a useful tool for categorizing groups of fluxes based on their coupling characteristics for large-scale network models [8].

Flux coupling analysis identifies three types of topological flux characteristics. Namely, fully coupled, partially coupled, and directionally coupled fluxes [8]. In this work, the fully and partially coupled flux sets are most relevant. Any flux in a set of fully coupled fluxes uniquely determines the value of all other fluxes in the set. Partially coupled fluxes are similar, except that the flux determines a specific range of values that all other fluxes in a partially coupled flux set may have. Fully and partially coupled flux sets may have overlapping elements, and so they will collectively be referred to as coupled sets, hereafter.

The existence of coupled sets in a metabolic network holds important implications for the OCCI algorithm. OCCI identifies a set of capacity constraints, or bounds, to match given fitness measurements. If an identified capacity constraint is the bound on a flux that is a member of a coupled set, then constraining any of the fluxes in that coupled set would result in an equivalent growth rate due to the coupled nature of the network. The network model used in this study indeed contains coupled sets, as shown in Fig. 3.1. The existence of the coupled sets leads to significant challenges for the identification of correct capacity constraints. If OCCI is used to identify capacity constraints given
experimental fitness measurements, however, knowledge of coupled sets would aid OCCI in making more accurate predictions. Fluxes in a set of capacity constraints identified by OCCI would be investigated for coupling. Subsequently, any fluxes coupled to those originally identified would then be candidates for further investigation and experimental validation. Hence, the precision (defined as the capability of OCCI to pinpoint a capacity constrained flux) of OCCI predictions, may decrease, since the set of capacity constraints grows larger; however, the accuracy of OCCI predictions will increase, since equivalent capacity constraints will not be excluded from investigations.

Figure 3.1: Coupling characteristics of the simplified network model of *E. coli* central carbon metabolism where (A) shows the four fully coupled flux sets \((FC_1, FC_2, FC_3, FC_4)\) and (B) shows the partially coupled flux set \((PC_1)\). In both cases, thick lines represent fluxes in a coupled set.

For the results presented in the following sections, all NLPs are solved using the feasible path, large-scale NLP solver, CONOPT (Arki Consulting and Development A/S), while LPs are solved using CPLEX 10.0 (ILOG). Both solvers are accessed through the TOMLAB (Tomlab Optimization Inc.) interface in MATLAB (Mathworks). All
simulations are performed on an Intel Xeon 3.2 GHz processor with 4 GB of RAM. CPU time for the OCCI problem is on the order of minutes for a batch of simulations involving up to 3,990 variables and 2,068 constraints.
4.1 Identification of a capacity constraint based on a single growth rate measurement

In this section, we present a detailed analysis of the predictions made by the OCCI algorithm, as well as their implications for identifying capacity constraints using fitness measurements. To best illustrate important details of the algorithm, we use a simplified model of *E. coli* central carbon metabolism in the form of different “strains,” as described in Section 3.1. The flux distributions for the wild-type and constrained strains, as well as that predicted by OCCI are shown in Fig. 4.1. Additionally, dual variables corresponding to the upper and lower bounds on each flux are shown for all cases. As shown in Fig. 4.1A, the active bounds in the wild-type strain include those corresponding to an exchange flux (flux 17) and knocked out reactions (fluxes 15 and 19). Fluxes 6 and 7 are actively bounded by their lower bounds of zero, indicating that growth rate can increase if those reactions are reversible. In the constrained strain (Fig. 4.1B), flux 9 is actively bounded by its upper bound, indicated by the positive upper bound dual variable. This is expected
Figure 4.1: Flux distributions, shadow prices, and flux bounds for (A) FBA with no artificial capacity constraints, (B) FBA with an artificial capacity constraint at flux index 9, and (C) OCCI predictions made using the growth rate from (B). In all three figures, flux 20 corresponds to growth rate, white bars indicating shadow prices on lower bounds are to the left of their corresponding flux indices, while black bars indicating upper bound shadow prices are to the right.
Chapter 4. Predictive Capabilities of the OCCI Algorithm

since flux 9 is artificially constrained to half of its wild-type value by the upper bound. Finally, the flux distribution and bounds predicted by OCCI are shown in Fig. 4.1C. Firstly, the growth rate (flux 20) predicted by OCCI matches that of the constrained strain. Secondly, OCCI identifies flux 9 to be an active bound, as shown by the positive dual variable corresponding to the upper bound of flux 9 (Fig. 4.1C). In addition, OCCI identifies three other active upper bounds (fluxes 12, 15, and 16) as indicated by their non-zero dual variables. Two of these bounds (fluxes 12 and 16) are not active in the FBA simulation of the constrained strain (Fig. 4.1B) as each bound has an associated dual variable of zero. Furthermore, one of the active upper bounds found in the constrained strain (flux 19) is predicted to be inactive by OCCI. Despite these differences in the active set, OCCI is able to match the growth rate of the constrained strain. These results suggest the presence of alternate optima that lead to multiple flux distributions and active bound sets that have the same growth rate. To investigate this further, flux variability analysis (FVA) is performed with the growth rate fixed to that of the constrained strain (Fig. 4.2). Results of FVA show that the growth rate of the constrained strain can indeed be achieved by alternate optimal flux distributions. This partially explains why the set of active bounds identified by OCCI is different from that originally found in the FBA simulation of the constrained strain. Further investigation into the matter involving coupled reaction sets is provided in Section 4.5.

Clearly, the presence of alternate optima limits the capability of OCCI to identify the actual capacity constraint since this means that the active bound set corresponding to a given growth rate is not unique. In the next section, we show how this limitation can largely be overcome when fitness measurements of a large array of strains is available.
Figure 4.2: FVA of the simplified network model of *E. coli* central carbon metabolism for a fixed growth rate (flux 20) equal to that of the constrained strain. Maximum fluxes of unbounded value are not shown.

### 4.2 OCCI predictions improve with multiple fitness measurements

We now demonstrate how the accuracy of OCCI predictions improves when an increasing number of fitness measurements is available to the algorithm. This feature provides strong incentive for using OCCI with high-throughput fitness profiling data, where methods for systematically analyzing large quantities of data are needed. As shown in Section 4.1, the presence of alternate optimal solutions in the network can negatively affect the accuracy of OCCI predictions. Fig. 4.3 illustrates how the increased availability of fitness measurements will improve OCCI predictions. Specifically, the score assigned to the percentage of active bounds identified (IA) increases significantly as the number of parallel strains provided to OCCI increases. Another important point to note is that OCCI is able to match the given fitness measurements in all cases. This result is significant because as the number of strains increases, the size of the nonlinear problem that OCCI must solve becomes larger, and thus more difficult to solve. This result also
suggests that the OCCI algorithm may utilize larger metabolic network models when performing more detailed analyses of experimental fitness measurements.

Figure 4.3: Predictive capability of OCCI versus the number of growth phenotype measurements available. Scoring metrics are defined in Eqs. (3.1), (3.2), (3.3).

The ability of OCCI to make quantitative predictions of the flux value of capacity constrained reactions follows an interesting trend with increasing number of strains. The results shown in Fig. 4.3 suggest that although capacity constraints are identified with increasing accuracy as more fitness measurements are available, quantitative flux value predictions do not show a drastic improvement. To explain why this is the case, we first address the fact that fitness measurements are matched with perfect accuracy in all cases studied. Hence, in the context of the OCCI algorithm, the optimization problem to minimize discrepancy between predicted and measured growth rates has reached its global optimum. Hence, the fact that different flux distributions can result in the same
growth phenotype suggests the presence of alternate optima even for a large set of strains. We also note the non-monotonic and apparently stochastic nature of the three scores as a function of the number of strains. This trend arises due to two reasons: (a) the OCCI optimization problem is nonlinear and nonconvex and we search for a global solution with random initial conditions, and (b) alternate optima allow a global optimum to be reached even if the capacity constraint identified is inaccurate. These reasons also explain similar trends in later figures (Fig. 4.6) and why the scores may drop before increasing. Nonetheless, an overall improvement in the capability of OCCI to identify capacity constraints is clearly seen.

### 4.3 Variability of capacity constrained fluxes

To investigate whether or not the effect of alternate optima changes with an increasing number of strains, FVA is performed for a set of one to fourteen strains. One should realize that a capacity constrained flux is part of an alternate optimal set of solutions, where constraints on other reactions can lead to the same growth rate for different values of the flux. Hence, in some strains, a capacity constrained flux may be variable for the same growth rate and OCCI cannot be expected to match the value of such fluxes. On the other hand, if a capacity constrained flux (or any flux, for that matter) is invariable for a given growth rate, then a unique, or fixed, flux value satisfies the given growth rate. In the case of such fluxes, OCCI must necessarily match that flux value if the resulting growth rate is also matched. We can expect that as the number of strains increases, the occurrence of invariable capacity constrained fluxes may increase. Hence, the degree to which quantitative flux predictions by OCCI improves depends on the number of these invariable capacity constrained fluxes. The actual percentage of invariable capacity constrained fluxes versus the number of strains is shown in Fig. 4.4. For each number of strains, the percentage of invariable capacity constrained fluxes is calculated as the total
number of fluxes with fixed flux value based on FVA for eleven capacity constraints in each strain.

![Graph showing the percentage of invariable capacity constrained fluxes with fixed value (i.e., invariable), subject to a fixed growth rate for one to 14 strains in a set.](image)

Figure 4.4: Percentage of artificially capacity constrained fluxes with fixed value (i.e., invariable), subject to a fixed growth rate for one to 14 strains in a set. A larger percentage of the capacity constrained fluxes are invariable as the number of strains increases.

Hence, for one strain, the score is based on eleven fluxes, while for two strains, there are $11 \times 2 = 22$ fluxes to consider, and so on, for a maximum of 14 strains (with $11 \times 14 = 154$ fluxes). The results show that the percentage of invariable capacity constrained fluxes increases significantly as the number of strains increases. The scores shown in Fig. 4.4 also represent the minimum that OCCI is expected to achieve with regards to quantitative flux value predictions. Interestingly, the quantitative flux value predictions made by OCCI (Fig. 4.3) consistently maintain a significantly better score than these theoretical minima of scores. In the context of computations, this is not a surprising result. This is due to the fact that while variable fluxes may operate within a range of values, a single value is actually found using an LP solver for the FBA problem, and an NLP solver for the OCCI problem. More often than not, the flux values predicted tend to stay at the lower range of possible flux values, and hence, the two flux value predictions may coincide despite the presence of alternate optima. This partially explains why the Quantitative
Accuracy score in Fig. 4.3 does not improve drastically with increasing number of strains—the score is already much higher than the largest of the minimum attainable scores for one strain, and we cannot expect an improvement over that score, regardless of the number of strains. By minimizing fluxes of certain reactions with appropriately small weightings in the objective function, we may facilitate quantitative flux value predictions in the presence of alternate optima. Such an approach is not taken here, since a relatively conservative evaluation of the capability of the OCCI algorithm is being demonstrated.

4.4 Decreasing bound variability with increasing number of strains

As shown in Fig. 4.3, OCCI predictions of the capacity constraints improves significantly as the number of strains increases. We now investigate why this trend is expected based on the underlying OCCI algorithm. Just as the presence of alternate optimal flux distributions limits the ability of OCCI to make quantitative flux value predictions, alternate optimal capacity constraints also exist. Specifically, alternate optimal sets of upper and lower bound values may satisfy a given set of fitness measurements. This is evident as the Identification Accuracy score in Fig. 4.3 does not reach 100%, although the growth rate is matched, indicating optimality of the NLP.

To investigate why an increased number of strains should improve OCCI predictions, we perform BVA, formulated in Section 2.4, for all capacity constraints for each set of strains. The results of BVA on all eleven capacity constraints for one to fourteen strains are shown in Fig. 4.5.

For a small number of strains, many of the capacity constraints have considerable variability in the upper and lower bounds. With five strains, several upper bounds are forced to have fixed values, subject to the fitness measurements. Also, a number of lower bounds show a significant decrease in variability. With six strains, another upper bound becomes
Figure 4.5: Bound variability analysis performed on the simplified network model for one to 14 strains. Bound variability is defined as the difference between maximum and minimum values a bound may have, subject to a specified growth rate. The variability of upper bounds on the eleven investigated fluxes decreases as the number of growth phenotypes that must be reconciled simultaneously increases. Ordinate limits are chosen to include the largest variability values that are less than 900.

Invariable and this trend mostly continues as the number of strains increases. Overall, we see the expected trends of decreasing bound variability with increasing numbers of fitness measurements; therefore, BVA illustrates how OCCI predictions are expected to improve with an increasing number of fitness measurements. While a number of capacity constraints are invariable for a sufficiently large number of strains, certain capacity constraints are still variable, even for the largest number of strains investigated here. These capacity constraints, then, are very difficult to predict by OCCI, since their bounds can be variable for the given set of fitness measurements. This suggests a method by which gene deletion studies may be screened, based on how much bound variability a candidate capacity constraint will have in the deletion strain. That is, a deletion strain that
does not decrease bound variability of any candidate capacity constraints would not contribute to the identification of capacity constraints and should be given lower priority (investigated later in the algorithm) for incorporation into the OCCI algorithm. Deletion strains that decrease bound variability for at least one strain should be included in the algorithm, while strains that decrease bound variability for many strains should be given highest priority (investigated early in the algorithm).

### 4.5 Incorporation of flux coupling characteristics improves OCCI predictions

As shown in the previous sections, OCCI is able to make increasingly accurate predictions as the number of strains increases; however, OCCI is not able to identify all of the capacity constraints even for the maximum number of strains in Fig. 4.3. We now investigate how predictions can be improved further by incorporating knowledge of flux coupling characteristics into the OCCI algorithm. As shown in Fig. 3.1, the metabolic network model used for computations includes four fully coupled sets consisting of two fluxes each, and one partially coupled set consisting of four fluxes. Reinforcing the OCCI algorithm with knowledge of flux coupling characteristics significantly improves predictions (Fig. 4.6). Even when only six strains are investigated, OCCI is able to identify the capacity constraints with 100% accuracy. This shows that a considerable number of capacity constraints are members of coupled sets. We note that some of the active sets identified have increased in size due to the inclusion of coupled fluxes that were not directly predicted by OCCI. This implies that while prediction accuracy has increased, precision has decreased, where precision is defined by the ability to pinpoint a capacity constrained flux; however, this tradeoff in predictive capability is expected to become less apparent as the model size increases. This is based on the observation that as metabolic network size increases, the number of reactions that each metabolite is involved in gen-
erally also increases [41]. Hence, large coupled sets would be decoupled into a number of smaller coupled sets [8], thereby possibly reducing the impact of flux coupling on the precision of OCCI predictions. Additionally, coupling characteristics vary with growth conditions and genetic modifications, as demonstrated by [8]. These observations suggest how one may design fitness competition assays by determining gene deletions and culture conditions that minimize the formation of large coupled sets.

Figure 4.6: Improved predictive capability of OCCI versus the number of growth phenotype measurements available when coupling characteristics of the metabolic network are known. Scoring metrics are defined in Eqs. (3.1), (3.2), (3.3).

In this section, we have demonstrated the integration of an existing method for topological analysis of metabolic networks (FCF, [8]) with the OCCI algorithm to improve prediction accuracy. As demonstrated by [38], metabolic networks exhibit unique prop-
erties that are most effectively understood by the combined analysis of both network structure (topology) and pathway function. This insight encompasses the major mode of analysis in several recent studies that utilize a combined FBA and pathway analysis approach [71, 72, 59, 48, 24]. We may envision the integration of the OCCI algorithm with such pathway-based methods, wherein the global network effect of an identified capacity constraint can be assessed by investigating the capacity constraint as part of an extreme pathway [60], or an elementary flux mode [62]. For example, in the case of gene deletion studies, analyzing the removed reactions as part of extreme pathways has shown to directly attribute the gene deletion to loss of pathway function that results in suboptimal growth rates [59]. In this context, OCCI may be able to pinpoint a small subset of reactions based on fitness measurements, which may further be analyzed by focusing on the extreme pathways that involve this subset of reactions. In this way, we may complement various existing pathway-based techniques with OCCI to further expand the tools available for conducting combined structural and functional network analyses.
Chapter 5

OCCI, Thermodynamics, and Experiments

In the previous chapters, the OCCI algorithm was developed and validated in a simplified model of central metabolism against a set of test data. Now, OCCI is utilized to begin answering some intriguing questions arising from the work of [74], which includes fitness measurements of several *E. coli* strains. This chapter is presented as follows: in Section 5.1, the relative fitness profiles of [74] are used as a training set for OCCI and we subsequently investigate the identified enzymatic capacity constraints, predict relative fitnesses under additional genetic backgrounds, and compare predicted intracellular flux distributions with measurements in the literature. In Sections 5.2 and 5.3, we ensue a detailed investigation of capacity constraints on the energy-linked, pyridine nucleotide transhydrogenase, PntAB, extending into a thermodynamic analysis in Section 5.4 of the hypotheses we generate in the preceding sections.

As discussed in previous chapters, the OCCI optimization problem involves multiple locally optimal solutions; therefore, the predictions are mostly presented as a range of values, spanning the minimum and maximum values obtained from multiple runs of the optimization problem from random initial conditions. Such computational efforts, how-
ever, still do not imply that our predictions reflect the actual biological phenomena. Hence, our predictions must be validated and the model iteratively refined by experimental verification. While experimentation is not performed in this work, the in silico investigations have been designed to accelerate future experimental validation. For this, we take an eclectic, or multi-pronged, validation approach, i.e., we facilitate experimental validation of our hypotheses by approaching them from multiple computational angles in the following specific forms: (1) relative fitnesses under additional genetic backgrounds are predicted with the ECCs (Section 5.1), which are testable in the future, (2) intracellular fluxes are predicted under various genetic backgrounds using the ECCs and compared to the metabolic flux analysis (MFA) literature, (3) testable physiological states, such as metabolite concentration ratios, are predicted, which can be validated by future experimentation. Where literature is not available, our model predictions suggest a focused set of experiments for validating our hypotheses.

5.1 Model reconciliation with experimental data

We use the experimental fitness profiles from Zhu et al. [74] to identify bottlenecks in E. coli metabolism. In [74], a novel strain of E. coli is engineered—one expressing the isocitrate dehydrogenase (IDH) enzyme with NAD cofactor specificity. The mutant expressing this novel enzyme, IDH-NAD, differs from the wild-type, which expresses an NADP-specific IDH (IDH-NADP). In the wild-type, IDH-NADP is an important source of NADPH, the reduced form of NADP, which is required for the cell to grow. In fact, when grown on acetate as the sole carbon source, E. coli depends on IDH-NADP for about 90% of NADPH production. Consequently, the mutant, which generates NADH, rather than NADPH, loses a major source of NADPH and relies on its production from other reactions. By sequentially removing these additional NADPH-generating reactions
by deleting the corresponding genes, the relative contributions of NADPH from these reactions can be inferred. Due to complex interactions and uncertain causal relations amongst the metabolic reactions, however, a systematic approach such as OCCI is useful, if not necessary to do so. An especially useful feature of the experimental design of Zhu et al. [74], is the fact that they grew the wild-type and mutant *E. coli* strains in competition for the sole carbon source in a chemostat culture. This leads to a balance between cell growth and dilution flow into the chemostat, and growth proceeds at steady state for many cell generations. The consequence of this is extremely low experimental uncertainty in the relative growth rates of the competing strains of *E. coli* (0.3% ∼ 12%, [74]). We used these relative growth rates as the training set to OCCI, to identify the bottlenecks from 113 metabolic reactions, balancing 67 metabolites, which constituted the stoichiometric matrix (see Tables 7.1 and 7.2 for a description of the reactions and metabolites). As seen in Figure 5.1(b), OCCI is able to reconcile all relative growth rates, except one, to within experimental error. The single relative growth rate may not have been matched as the relative magnitudes of experimental error were not accounted for in OCCI. OCCI identified ECCs that are inherent in all strains under all genetic backgrounds studied. As is evident in Figure 5.1(a), in the absence of ECCs, FBA predictions of relative fitness fail to resemble the experiments under most genetic backgrounds.

The ECCs identified by OCCI (Table 5.1) are represented as a range of values spanning multiple sets of ECCs. Each of the sets are obtained by solving the OCCI problem from a random initial condition, in order to explore the solution space characterized by multiple local optimal solutions.

The measurements, of [74], themselves show an interesting trend with varying genetic background. As originally hypothesized by [74], the genetic backgrounds other than wild-type progressively remove sources of NADPH, which is required for growth. *icd*$_{NADP}$ strains retain the major source of NADPH, IDH-NADP, even as additional sources of NADPH are removed. *icd*$_{NAD}$ strains, however, rely heavily on the auxiliary sources of
Figure 5.1: Relative growth rates ($\frac{\text{icd}}{\text{icd}^\text{NADP}}$) measured by [74] (black lines indicating experimental error) and predicted by FBA without (a) and with (b) enzymatic capacity constraints (ECCs). Bars above and below "x" in (b) indicate upper and lower prediction uncertainties, respectively.

NADPH, such as PntAB, MaeB, and pentose-phosphate pathway reactions, since its IDH enzyme, IDH-NAD, generates NADH, rather than NADPH. Consequently, under most genetic backgrounds, $\text{icd}^\text{NADP}$ strains grow faster than $\text{icd}^\text{NAD}$ strains.

Interestingly, under $\Delta\text{udhA}$ and $\Delta\text{udhA}$, $\Delta\text{pntAB}$ genetic backgrounds, $\text{icd}^\text{NAD}$ strains have the fitness advantage. As hypothesized by [74], the bottleneck to growth in these cases is not NADPH, but NADH. This is because UdhA, if proceeding in the forward direction, is a significant source of NADH for acetate-grown E. coli. By removing UdhA, then, $\text{icd}^\text{NADP}$ strains must now rely on other sources of NADH, while $\text{icd}^\text{NAD}$ strains possess a major source of NADH, the IDH-NAD-catalyzed reaction.

A question arises, however, when comparing the strains under $\Delta\text{udhA}$ and $\Delta\text{udhA}$, $\Delta\text{pntAB}$ genetic backgrounds. Why does removing PntAB further deter fitness of
Table 5.1: Enzymatic Capacity Constraints

<table>
<thead>
<tr>
<th></th>
<th>Lower Bounds</th>
<th>Upper Bounds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>min</td>
<td>max</td>
</tr>
<tr>
<td>PntAB</td>
<td>-0.63</td>
<td>-0.02</td>
</tr>
<tr>
<td>UdhA</td>
<td>0.87</td>
<td>3.64</td>
</tr>
<tr>
<td>MDH</td>
<td>-67</td>
<td>-24.9</td>
</tr>
<tr>
<td>MaeB</td>
<td>0.89</td>
<td>3.85</td>
</tr>
<tr>
<td>IDH-NAD</td>
<td>-100</td>
<td>1.2</td>
</tr>
<tr>
<td>IDH-NADP</td>
<td>-100</td>
<td>0.03</td>
</tr>
</tbody>
</table>

*icd* \(^{NADP}\) strains if PntAB is just another producer of NADPH? Indeed, initial attempts to reconcile all measurements with OCCI with the assumption that PntAB activity proceeds only in the forward direction, thus producing NADPH, failed to capture the difference in these two genetic backgrounds. The basis for assuming strictly forward PntAB activity came from both observations reported in the literature [57], as well as thermodynamics-based assumptions. Both of these factors will be discussed in more detail in Sections 5.2 and 5.3 as well as how a more critical assessment of these assumptions led to generation of new hypotheses.

Another intriguing capacity constraint identified by OCCI, is the lower bound on UdhA (Table 5.1). The UdhA-catalyzed reaction is as follows:

\[
\text{NAD}^+ + \text{NADPH} \leftrightarrow \text{NADH} + \text{NADP}^+
\]

A positive, non-zero, lower bound indicates that unless the corresponding gene is deleted, the reaction always proceeds forward. This way, UdhA acts as a continuous drain of NADPH and generator of NADH (see list of reactions in Table 7.1). Considering that [74] originally hypothesized that insufficient NADPH limits growth for most of the investigated genetic backgrounds, this capacity constraint actually deters growth. From the multitudes of simulations performed to obtain a globally optimal solution, this capacity
constraint was consistently found. Also, it appears that it is necessary to match the observed difference between wild-type and $\Delta udhA$ relative fitnesses. That is, by deleting $udhA$, the $icd^{NAD}$ strains have a growth advantage over the $icd^{NADP}$ strains. The UdhA ECC can be interpreted from a thermodynamic perspective, as well. The UdhA-catalyzed reaction operates near equilibrium at standard conditions \[20\]. Hence, metabolite concentration gradients involving low NADP to NADPH or high NAD to NADH ratios may force flux through this reaction. Considering that concentration of metabolites as ubiquitous to the metabolic network as NAD(P)(H) would have far-reaching effects throughout the metabolic network, a quantitative analysis appears difficult. Nonetheless, computational algorithms exist to deal with exactly this sort of complexity, which will be discussed further in Section 5.4.

As discussed in the beginning of this chapter, our model predictions take on an eclectic approach to facilitate future experimental verification. One method involves predicting relative fitnesses that were not included in the training set (i.e., data of \[74\]). We extended relative fitness predictions to four additional genetic backgrounds (Figure 5.2). $gnd$ encodes the 6-phosphogluconate dehydrogenase (GND) enzyme, which catalyzes the following reaction (see Table 7.2 for definition of metabolites in the reaction):

\[
D6PGC + NADP^+ \rightarrow RL5P + NADPH + CO2
\]

GND, therefore, is another source of NADPH, and the effect of removing it is expected to be non-trivial under genetic backgrounds where NADPH is limiting. Consequently, the four genetic backgrounds all involve $\Delta gnd$ and combined deletion of $udhA$, $pntAB$, and $maeB$, which also affect NADPH availability. Interestingly, the $\Delta gnd$, $\Delta udhA$ genetic background may or may not have a similar effect on relative fitness as the $\Delta udhA$ genetic background, depending on the ECC set used. Experiments verifying these predictions, therefore, will be invaluable in curating the sets of identified enzymatic capacity constraints.

Another approach to validating our predictions involves comparing predicted and ob-
Figure 5.2: Relative fitnesses of acetate-grown *E. coli* under additional genetic backgrounds predicted with FBA constrained by the ECCs identified by OCCI. Relative fitness $= \frac{\mu_{icdNAD}}{\mu_{icdNADP}}$, where $\mu$ denotes growth rate. Predictions span a range due to computational uncertainty in identifying the global optimum.

erved intracellular flux distributions. In addition to the wild-type genetic background originally included in the training set for OCCI, another genetic background, $\Delta zwf$ is also simulated (Figure 5.3). These simulations involve using FBA with the ECCs identified by OCCI. At this point, we should recall that the optimization problem solved by OCCI involves optimally matching the provided relative fitnesses (training set), while no information is provided regarding intracellular flux distributions. Hence, OCCI matches the training set by distributing flux throughout the metabolic network such that growth rates are most similar to the observed ones. As discussed in the previous chapters, widening the training set improves predictive capabilities of OCCI, and in this investigation, up
Figure 5.3: FBA predictions of intracellular flux of icq$^{NADP}$ strains, constrained by OCCI-predicted ECCs under (a) wild-type and (b) ∆zwf genetic backgrounds.

to seven relative fitnesses, involving fourteen growth rate s, were used as inputs to OCCI. As a result, OCCI is able to improve the agreement of predicted and observed intracellular flux distributions. Most notably, [73] observe that glucose- and acetate-grown E. coli under a ∆zwf genetic background operate the non-oxidative pentose phosphate pathway in the reverse direction (see Figure 5.3(b)), a characteristic that was only captured when ECCs were used in the FBA simulation. In this way, model predictions are verified with observations and this approach can be extended to investigate additional genetic backgrounds. Metabolic flux analysis [3, 64] is an established method to experimentally infer intracellular flux distributions and may be used to verify our predictions.
5.2 Reverse PntAB flux may improve fitness under certain genetic backgrounds

In this section, we consider the consequences of enzymatic capacity constraints on PntAB activity to cell growth. A critical assessment of the validity of reverse PntAB activity is discussed in Section 5.3.

Considering *E. coli* grown solely on acetate, if both PntAB and UdhA are reversible and have equal capacity in either direction, we would expect both catalyzed reactions to operate in reverse to improve fitness. This way, PntAB will generate a protonmotive force instead of dissipating it, while UdhA will supply NADPH by reoxidizing NADH produced by reverse PntAB activity. FBA simulations comparing growth rate with varied reversibility of the transhydrogenases show that when both PntAB and UdhA are reversible, *E. coli* may grow at twice the rate as when PntAB is irreversible (Figure 5.4). In contrast, literature observations indicate primarily forward activity of both enzymes [57]. Why is this so, and what physiological conditions are missing in the model to overpredict fitness? Using OCCI, we identified key metabolic bottlenecks in specific reactions (Table 5.1). From this, we hypothesize and analyze the physiological conditions that would enforce these constraints (Section 5.4). Finally, where evidence for or against our hypotheses is not available in the literature, we suggest experiments that would validate the model predictions generated by incorporating our hypothesized enzymatic capacity constraints in Section 6.

Additionally, all other ECCs being constant, the ECCs on forward or reverse PntAB activity can significantly affect relative fitness (Figure 5.5). Consequently, the quantitative flux capacities identified by OCCI (Table 5.1), are important for making quantitative predictions of metabolism and, once verified through experiments, represents a significant step forward from previous methods that determine reaction direction only. At the same time, the constraint-based modeling approach allows for efficient integration of different
constraint-based methods. We take advantage of this attribute in Section 5.4, where an existing algorithm to incorporate thermodynamic constraints is integrated with the enzymatic capacity constraints identified by OCCI.

5.3 Flexible function of the membrane-bound transhydrogenase

The membrane-bound, energy-dependent transhydrogenase (Figure 5.6), encoded by pntAB in E. coli, catalyzes the transhydrogenation between NAD(H) and NADP(H) coupled to translocation of protons [31] as follows:

\[ \text{NADH} + \text{NADP}^+ + x\text{H}^+_{\text{periplasm}} \leftrightarrow \text{NAD}^+ + \text{NADPH} + x\text{H}^+_{\text{cytosol}}. \]
Figure 5.5: Sensitivity of predicted relative fitnesses when enzymatic capacities of forward and reverse fluxes of PntAB are relaxed. Of the genetic backgrounds studied in [74], those with \textit{pntAB} intact are investigated here.

Ever since discovery of their activity in \textit{Pseudomonas fluorescens} by [11, 34], the physiological role of transhydrogenases in bacterial and mammalian cells has been inconclusive. [6] first suggested its role in generating NADPH for biosynthesis in \textit{E. coli} and this role continues to find supportive observations [57, 21]. The physiological role of membrane-bound transhydrogenases in other bacteria and mammalian cells, however, is still an open question [28, 56].

Even in the well-studied \textit{E. coli}, we are encountering ambiguity in PntAB function. Specifically, the different relative fitnesses observed between \(\Delta udhA\) and \(\Delta udhA\), \(\Delta pntAB\) genetic backgrounds cannot be reconciled with a rigid PntAB function of strictly producing NADPH.

As discussed in [74], and also found \textit{in silico}, NADH is limiting for \(icd^{NADP}\) strains under a \(\Delta udhA\) genetic background. Forward PntAB activity reoxidizes NADH to generate
Figure 5.6: Cartoon schematic depicting PntAB enzyme structure, based on [4, 33].

NADPH, which would deter fitness in conditions where NADH is limiting. Also, since loss of PntAB activity further reduces $icd^{\text{NADP}}$ fitness, PntAB activity must benefit the strain’s fitness. This can only be explained by reversed PntAB activity under the $\Delta udhA$ genetic background. Under most physiological conditions, PntAB primarily catalyzes the forward reaction but reverse activity under a high NADPH/NADP ratio has also been observed [16, 43].

From the ECCs identified using OCCI, we observe several intriguing aspects of PntAB activity:

1. forward PntAB flux capacity is 1.14~97-fold higher than that of the reverse

2. forward UdhA flux capacity is 1.4~420-fold higher than that of reverse PntAB (both catalyzing reduction of NADP to NADPH)

From 1, we hypothesize that while the primary function of PntAB is to produce NADPH, under certain conditions of excessive NADPH and no alternative activity to reoxidize NADPH, the reverse reaction can perform this function; however, from 2, PntAB cannot fully surrogate UdhA function.
Two potentially additive reasons may explain the slow reverse rate of PntAB: (1) a proton electrochemical gradient ($\Delta p$) stimulates the forward reaction 10-fold [5], shifting equilibrium towards NADPH formation to >400 [32], and (2) PntAB expression is down-regulated by increased NADPH, while UdhA is down-regulated with decreased NADPH [57]. Hence, reverse PntAB activity is achieved under substrate/product ratios that exceed the protonmotive force driving the forward reaction. As discussed in Section 5.4, we can computationally predict the NAD(P)H/NAD(P) ratio necessary for reverse PntAB activity to be thermodynamically feasible, while taking into account the network-wide effects that changing concentrations of these ubiquitous cofactors has on the metabolic network as a whole.

If verified, reverse PntAB activity, albeit low, demonstrates remarkable flexibility of *E. coli* central metabolism to cope with genetic and environmental perturbations to intracellular NAD(P)(H) balancing and, as such, promotes a non-static view of transhydrogenase function and possibly a means to control engineered bacterial strains with altered redox metabolism.

## 5.4 Metabolite Activities Predicted from Thermodynamic Analysis

As discussed by [36], metabolite concentrations thermodynamically determine reaction direction for reactions that are sometimes dispersed throughout the metabolic network, leading to complex, nonlinear causal relationships amongst metabolites and reactions. This motivated the development of several algorithms to systematically assess the thermodynamic forces relating metabolite concentrations and reaction direction [36, 26, 25, 20]. Here, we use the approach of thermodynamics-based metabolic flux analysis (TMFA) [26, 25, 20], with modifications explained later in this section, to determine the range of metabolite concentrations necessary for reverse PntAB activity. The contributions of
this investigation to the existing approach are mainly twofold: (1) we describe a modification to the algorithm to calculate metabolite concentration ratios, and perhaps more importantly, (2) we integratively apply the two approaches of OCCI and TMFA to gain new insight that was unavailable from either model in isolation. Indeed, the ease with which this integration can be performed is an attractive feature of the constraint-based modeling approach.

TMFA allows us to predict metabolite activities and reaction Gibbs free energy changes ($\Delta_r G'$) while accounting for stoichiometric and thermodynamic interactions within the metabolic network. Using this approach, we predicted ranges of $\Delta_r G'$ and activities necessary to match the fitness profiles of [74].

Reverse PntAB activity simply requires $0 < \Delta_r G'_{PntAB} < \infty$, where,

$$\Delta_r G'_{PntAB} = \Delta_r G_{PntAB}^{m} + RT \sum n \ln x$$

where $n$ is the stoichiometric coefficient for each metabolite of concentration, $x$, involved in the reaction, $T$ is temperature in Kelvin, $R$ is the universal gas constant, and $\Delta_r G_{PntAB}^{m}$ is a transformed standard Gibbs free energy change accounting for both specified pH (7.0 is used in the analysis) and mM activities, rather than the standard 1 M activities. Using the mM Gibbs free energy change calculated in [20], the following relation holds:

$$\Delta_r G'_{PntAB} = -6.4 \pm k \sigma_{PntAB} + RT \left\{ \ln \left( \frac{x_{NADPH}}{x_{NADP}} \right) + \ln \left( \frac{x_{NAD}}{x_{NADH}} \right) + 2 \ln \left( \frac{x_{H_{cytoplasm}}^{+}}{x_{H_{periplasm}}^{+}} \right) \right\}$$

where $\sigma$ is the standard error in calculating $\Delta_r G_{PntAB}^{m}$, reported in [20], which can be accounted for by $k$ multiples. Utility of TMFA is evident as NAD(P)(H) are ubiquitous cofactors used throughout the metabolic network whose activities are thus constrained by thermodynamic feasibility of multiple interacting reactions.

In order to determine the minimum and maximum concentration ratios of NAD(P)(H), TMFA (originally a mixed-integer linear program) was modified into the appropriate mixed-integer fractional programming problem, which was then reformulated into a
mixed-integer linear program, as in [8] (see Section 7.1 for the reformulation). This problem was then solved assuming no uncertainty in the mM Gibbs free energies calculated in [20], with bounds on metabolite activity, as well as with and without the ECCs as additional constraints on reaction flux. As seen in Table 5.2, the absence of ECCs allows a wide range of metabolite concentrations, while this is significantly narrowed by incorporating ECCs.

Comparing predicted concentration ranges for wild-type and $\Delta udhA$ genetic backgrounds, we see distinct differences. For reverse PntAB activity to be thermodynamically favorable in the $\Delta udhA$ genetic background, we see that the maximum NADP/NADPH ratio is much lower (0.000616 vs. 1510.7), while the minimum NAD/NADH ratio is much higher (285.33 vs. 10.0), than that of the wild-type genetic background, where PntAB activity is in the forward direction. We hypothesize that UdhA activity, driven by concentration gradients of NAD/NADH and NADP/NADPH, does not allow for reverse PntAB activity, as NADPH is constantly re-oxidized to NADP and NAD is reduced to NADH. This is reflected in the ECC of UdhA, where a positive lower bound on UdhA indicates that as long as a thermodynamic driving force is maintained, i.e., NAD/NADH and NADP/NADPH ratios, UdhA activity is also maintained. Hence, persistent UdhA activity dissipates the NAD(P)(H) gradient necessary to overcome the protonmotive force in PntAB and allow reverse activity. When UdhA is removed under the $\Delta udhA$ genetic
background, NADPH and NAD accumulate to the point that reverse PntAB activity is thermodynamically favorable, and its activity becomes an important source of NADH for cellular energy in place of UdhA. As discussed in Section 5.3, however, reverse PntAB activity cannot fully surrogate UdhA’s function of producing NADH for the hypothesized reasons discussed in the previous section.
In this work, the bilevel optimization-based algorithm, OCCI, has been developed for reconciling constraint-based models of metabolism with fitness measurements by identifying flux capacity constraints. The final formulation of the algorithm consists of an NLP that can be solved using robust, off-the-shelf NLP solvers for the problem sizes investigated (maximum problem size of 3,990 variables and 2,068 constraints in Chapters 2 to 4 and 22,750 variables and 15,586 constraints in Chapter 5).

An in silico investigation has been performed to evaluate the capability of OCCI to identify capacity constraints, using a simplified network model of E. coli central carbon metabolism for ease of illustration. Eleven different fluxes are artificially constrained in parallel across sets of one to fourteen different strains. The resulting growth rates, analogous to experimental fitness measurements, and substrate uptake rates are used by OCCI to identify the capacity constraint in each investigation. For a total of 1,155 in silico investigations, OCCI correctly predicts all growth rates, despite the difficult complementarity constraints in the underlying NLP. This suggests potential for expanding the application of OCCI for experimental fitness profiles using more detailed metabolic network models.

The capability of OCCI to identify capacity constraints improves significantly as the
number of fitness measurements increases. This improvement is shown to be largely due to the decreasing variability of upper and lower flux bounds for a given set of fitness measurements, as the size of this set increases. Reinforcing OCCI with knowledge of flux coupling characteristics results in 100% identification accuracy when as few as six strains are in a parallel set. This improvement in accuracy comes at the cost of reduced precision due to the presence of a large coupled flux set in the model used. This tradeoff in accuracy and precision due to the presence of large coupled sets is expected to decrease as model size increases, based on the observations that network connectivity increases [41] so that the number of large coupled sets decreases [8] for larger networks.

Quantitative prediction of capacity constrained fluxes does not rapidly improve with increasing number of fitness measurements. This is mainly due to the high levels of flux variability in the model, although this variability does decrease with an increased number of fitness measurements. Both flux and bound variability change with genetic modifications and growth conditions. This suggests how one may design fitness competition assays when using OCCI to identify capacity constraints with experimental fitness profiles.

In a review of metabolic modeling in the post-genomic era [13], the authors emphasize the importance of iterative model-building and experimentation as one of the ongoing efforts towards the construction of whole-cell models and systemic understanding of cellular processes. In this respect, the OCCI algorithm represents one of many recent developments for integrating experimental data with computational modeling of complex, intracellular processes. In general, such methodologies may be depicted as in Fig. 6.1. For example, the OCCI algorithm can be applied within an iterative process in which high-throughput experimental data is used to reconcile constraint-based models, which are then used for system-level analysis of intracellular function. Hypotheses generated in silico are verified by “wet” experiments, and this forms the basis for design of further experimentation to verify previous, and generate new, hypotheses. This process eventually
identifies the key metabolic enzymes and pathways on which focused investigations can be performed through both in silico simulation and experimentation. Such integration of high-throughput omics data with computational modeling has been shown to both improve accuracy of model predictions and also to aid in the elucidation of key mechanisms and interactions in the metabolic network that are responsible for the discrepancies between model predictions and experimental data [72, 12]. In Chapter 5, we began this iterative discovery process and demonstrated how we arrived at the novel hypothesis linking UdhA and reverse PntAB activity.

Following the paradigm of constraint-based modeling, OCCImay be used to complement other existing frameworks in order to further constrain the solution space of metabolic flux distributions, based on additional experimental measurements in the form of fitness profiling data. In Section 5.4, we integrated the capabilities of TMFA and OCCII to predict metabolite concentration ratios for reverse PntAB activity to be thermodynamically feasible.

The next step involves verifying our model predictions. Due to our eclectic modeling efforts, experiments can be conducted at multiple fronts. First of all, in Figure 5.2, we predict relative fitnesses under additional genetic backgrounds. Using the approach of [74], a chemostat culture with $icd^{NAD}$ and $icd^{NADP}$ strains competing for acetate as the sole carbon source can be used to verify the predictions. If experiments fall within the predicted ranges, we gain confidence in our in silico analyses, while a discrepancy would aid us to correct ECC values, or to identify what salient constraints were overlooked in this first round of investigations.

In Section 5.1, we predicted intracellular flux distributions under wild-type and $\Delta zwf$ genetic backgrounds, which qualitatively agreed with observations in the literature. This approach can be readily extended to simulations under additional genetic backgrounds, which can then be verified by experiments. We suggest choosing genetic backgrounds designed to investigate novel hypotheses, such as the function of an enzyme, or to aid
understanding of causal relations amongst the reactions, such as the effect a gene deletion has across the metabolic network at a system level.

Finally, in Section 5.4, we predicted metabolite concentration ratios that would enable reverse PntAB activity to be thermodynamically feasible under a $\Delta udhA$ genetic background. Various experimental methods may verify these predictions, including Nuclear Magnetic Resonance spectrometry-based methods [70], or fluorescence-based and more classical methods [47].

By performing the experiments suggested above, the first cycle of modeling, experimentation and validation is complete. What follows is an iterative process of refining the model by verifying new predictions and in silico analysis. In the process, we can expect to generate novel hypotheses, similar in fashion to our hypothesis regarding reverse PntAB activity, further enhance the model itself, and to expand this process by encompassing additional algorithms, like TMFA, and increasingly sophisticated experimental techniques, including metabolomics, fluxomics, and other high-throughput profiling methods.
Figure 6.1: Schematic of the OCCI algorithm as part of an iterative process of model reconciliation and experimental validation, which allows for focused investigation of the key metabolic enzymes and pathways.
Chapter 7

Appendix

7.1 Reformulating TMFA as a fractional program

Here, we formulate minimization and maximization of concentration ratios as a fractional program based on TMF A of [25].

\[
\begin{align*}
\min_{v, \ln(\varepsilon), \Delta_r G', \varepsilon} & \quad \frac{\ln(x_1)}{\ln(x_2)} \\
\text{s.t.} & \quad Sv = 0, \\
& \quad 0 \leq v_i \leq z_i v_{max}, \ \{i = 1, ..., n\}, \\
& \quad \Delta_r G'_i - K + K z_i < 0, \ \{i = 1, ..., n\}, \\
& \quad \Delta_r G''_i + E_i + RT \sum_{j=1}^{m} n_{i,j} \ln(x_j) = \Delta_r G'_i, \\
& \quad \{i = 1, ..., n\}, \\
& \quad \ln(x_i^L) \leq \ln(x_i) \leq \ln(x_i^U), \ \{i = 1, ..., m\}, \\
& \quad \Delta_r G''_i \leq \Delta_r G'_i \leq \Delta_r G''_i, \ \{i = 1, ..., n\},
\end{align*}
\]

where \(n\) and \(m\) are numbers of reactions and metabolites in the metabolic network, respectively, \(S\) is the stoichiometric matrix of dimension \(m\)-by-\(n\), \(v\) is the flux vector, \(x_i\) are metabolite concentrations, \(\Delta_r G'_i\) and \(\Delta_r G''_i\) are the reaction gibbs free energy and
mM reaction Gibbs free energy. $K$ is a large positive constant. As in [8], the above fractional program is reformulated into an equivalent linear program as follows:

\[
\begin{align*}
\text{min} \quad & \ln(\tilde{x}_1) \\
\text{or max} \quad & \ln(\tilde{x}_1) \\
\text{s.t.} \quad & \ln(\tilde{x}_2) = -1, \\
& S\tilde{v} = 0, \\
& 0 \leq v_i \leq z_i v_{\text{max}}, \{i = 1, ..., n\}, \\
& \Delta_r \tilde{G}_i' - K + K z_i < 0, \{i = 1, ..., n\}, \\
& \Delta_r \tilde{G}_i^o + \tilde{E}_i + RT \sum_{j=1}^{m} n_{i,j} \ln(x_j) = \Delta_r \tilde{G}_i', \\
& \{i = 1, ..., n\}, \\
& \ln(\tilde{x}_i^L) \leq \ln(\tilde{x}_i) \leq \ln(\tilde{x}_i^U), \{i = 1, ..., m\}, \\
& \Delta_r \tilde{G}_i'^L \leq \Delta_r \tilde{G}_i' \leq \Delta_r \tilde{G}_i'^U, \{i = 1, ..., n\}.
\end{align*}
\]  

The $\tilde{\cdot}$ notation denotes a transformed variable equal to a multiple of the original variable and the new variable, $t$. Note that in constraint 7.2b, $\ln(\tilde{x}_2) = -1$, rather than 1. This intentional variation from the original reformulation of [8] ensures that the new variable, $t = \frac{-1}{\ln(x_2)}$ is positive. This is because the bounds on the concentrations considered as $x_2$, which are NADH and NADPH, are such that $\ln(x_2)$ is always negative. Hence, constraint 7.2b ensures that $t$ is always positive, and the transformed variables and constraints are also valid. Furthermore, the binary variables, $z_i$ need not be transformed—a necessity in constraint 7.2d, where $z_i$ must be either 0 or 1. Instead, in constraint 7.2e, the large constant, $K$, is multiplied by $t$ and transformed, simply resulting in another large constant. For this reason, $z_i$, and also flux variables, $v_i$ are unchanged.
7.2 The metabolic reaction network model

The metabolic reaction network model consists of a matrix whose elements are stoichiometric coefficients. The reactions and metabolites represented in the model are listed in Tables 7.1 and 7.2.

Table 7.1: List of metabolic reactions used in this work

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLK1</td>
<td>-1<em>GLC -1</em>ATP +1<em>G6P +1</em>ADP</td>
</tr>
<tr>
<td>PGI1R</td>
<td>-1<em>G6P +1</em>F6P</td>
</tr>
<tr>
<td>PFKA</td>
<td>-1<em>ATP +1</em>ADP -1<em>F6P +1</em>FDP</td>
</tr>
<tr>
<td>FBP</td>
<td>+1<em>F6P -1</em>FDP +1*PI</td>
</tr>
<tr>
<td>FBAR</td>
<td>-1<em>FDP +1</em>T3P1 +1*T3P2</td>
</tr>
<tr>
<td>TPIAR</td>
<td>+1<em>T3P1 -1</em>T3P2</td>
</tr>
<tr>
<td>GAPAR</td>
<td>-1<em>PI -1</em>T3P1 -1<em>NAD +1</em>NADH +1*13PDG</td>
</tr>
<tr>
<td>PGKR</td>
<td>+1<em>ATP -1</em>ADP -1<em>13PDG +1</em>3PG</td>
</tr>
<tr>
<td>GPMAR</td>
<td>-1<em>3PG +1</em>2PG</td>
</tr>
<tr>
<td>ENOR</td>
<td>-1<em>2PG +1</em>PEP</td>
</tr>
<tr>
<td>PPSA</td>
<td>-1<em>ATP +1</em>PI +1<em>PEP -1</em>PYR +1*AMP</td>
</tr>
<tr>
<td>PYKF</td>
<td>+1<em>ATP -1</em>ADP -1<em>PEP +1</em>PYR</td>
</tr>
<tr>
<td>ACEE</td>
<td>-1<em>NAD +1</em>NADH -1<em>PYR -1</em>COA +1<em>CO2 +1</em>ACCOA</td>
</tr>
<tr>
<td>ZWFR</td>
<td>-1<em>G6P -1</em>NADP +1<em>D6PGL +1</em>NADPH</td>
</tr>
<tr>
<td>PGL</td>
<td>-1<em>D6PGL +1</em>D6PGC</td>
</tr>
<tr>
<td>GND</td>
<td>+1<em>CO2 -1</em>NADP +1<em>NADPH -1</em>D6PGC +1*RL5P</td>
</tr>
<tr>
<td>RPIAR</td>
<td>-1<em>RL5P +1</em>R5P</td>
</tr>
<tr>
<td>RPER</td>
<td>-1<em>RL5P +1</em>X5P</td>
</tr>
<tr>
<td>TKTA1R</td>
<td>+1<em>T3P1 -1</em>R5P -1<em>X5P +1</em>S7P</td>
</tr>
</tbody>
</table>

Continued on next page ...
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>TKTA2R</td>
<td>+1<em>F6P +1</em>T3P1 -1<em>X5P -1</em>E4P</td>
</tr>
<tr>
<td>TALAR</td>
<td>+1<em>F6P -1</em>T3P1 -1<em>S7P +1</em>E4P</td>
</tr>
<tr>
<td>GLTA</td>
<td>+1<em>COA -1</em>ACCOA -1<em>OA +1</em>CIT</td>
</tr>
<tr>
<td>ACNAR</td>
<td>-1<em>CIT +1</em>ICIT</td>
</tr>
<tr>
<td>icdNAD</td>
<td>-1<em>NAD +1</em>NADH +1<em>CO2 -1</em>ICIT +1*AKG</td>
</tr>
<tr>
<td>icdNADP</td>
<td>+1<em>CO2 -1</em>NADP +1<em>NADPH -1</em>ICIT +1*AKG</td>
</tr>
<tr>
<td>SUCA</td>
<td>-1<em>NAD +1</em>NADH -1<em>COA +1</em>CO2 -1<em>AKG +1</em>SUCCOA</td>
</tr>
<tr>
<td>SUCCR</td>
<td>+1<em>ATP -1</em>ADP -1<em>PI +1</em>COA -1<em>SUCCOA +1</em>SUCC</td>
</tr>
<tr>
<td>SDHA1</td>
<td>-1<em>SUCC -1</em>FAD +1<em>FADH +1</em>FUM</td>
</tr>
<tr>
<td>FRDA</td>
<td>+1<em>SUCC +1</em>FAD -1<em>FADH -1</em>FUM</td>
</tr>
<tr>
<td>FUMAR</td>
<td>-1<em>FUM +1</em>MAL</td>
</tr>
<tr>
<td>MDHR</td>
<td>-1<em>NAD +1</em>NADH +1<em>OA -1</em>MAL</td>
</tr>
<tr>
<td>DLD1R</td>
<td>+1<em>NAD -1</em>NADH -1<em>PYR +1</em>LAC</td>
</tr>
<tr>
<td>ADHE2R</td>
<td>+2<em>NAD -2</em>NADH +1<em>COA -1</em>ACCOA +1*ETH</td>
</tr>
<tr>
<td>PFLA</td>
<td>-1<em>PYR -1</em>COA +1<em>ACCOA +1</em>FOR</td>
</tr>
<tr>
<td>PTAR</td>
<td>-1<em>PI +1</em>COA -1<em>ACCOA +1</em>ACTP</td>
</tr>
<tr>
<td>ACKAR</td>
<td>+1<em>ATP -1</em>ADP -1<em>ACTP +1</em>AC</td>
</tr>
<tr>
<td>ACS</td>
<td>-1<em>ATP +1</em>AMP -1<em>COA +1</em>ACCOA -1<em>AC +1</em>PPI</td>
</tr>
<tr>
<td>PCKA</td>
<td>-1<em>ATP +1</em>ADP +1<em>PEP +1</em>CO2 -1*OA</td>
</tr>
<tr>
<td>PPC</td>
<td>+1<em>PI -1</em>PEP -1<em>CO2 +1</em>OA</td>
</tr>
<tr>
<td>MAEB</td>
<td>+1<em>PYR +1</em>CO2 -1<em>NADP +1</em>NADPH -1*MAL</td>
</tr>
<tr>
<td>SFCA</td>
<td>-1<em>NAD +1</em>NADH +1<em>PYR +1</em>CO2 -1*MAL</td>
</tr>
<tr>
<td>ACEA</td>
<td>-1<em>ICIT +1</em>SUCC +1*GLX</td>
</tr>
<tr>
<td>ACEB</td>
<td>+1<em>COA -1</em>ACCOA +1<em>MAL -1</em>GLX</td>
</tr>
<tr>
<td>PPA</td>
<td>+2<em>PI -1</em>PPI</td>
</tr>
</tbody>
</table>

Continued on next page . . .
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLPK</td>
<td>-1<em>ATP +1</em>ADP -1<em>GL +1</em>GL3P</td>
</tr>
<tr>
<td>GPSA1R</td>
<td>+1<em>T3P2 -1</em>NADP +1<em>NADPH -1</em>GL3P</td>
</tr>
<tr>
<td>RBSK</td>
<td>-1<em>ATP +1</em>ADP +1<em>R5P -1</em>RIB</td>
</tr>
<tr>
<td>NUOA</td>
<td>+1<em>NAD -1</em>NADH -1<em>Q +1</em>QH2 +3.5*HEXT</td>
</tr>
<tr>
<td>FDOH</td>
<td>+1<em>CO2 -1</em>FOR -1<em>Q +1</em>QH2 +2*HEXT</td>
</tr>
<tr>
<td>GLPD</td>
<td>+1<em>T3P2 -1</em>GL3P -1<em>Q +1</em>QH2</td>
</tr>
<tr>
<td>CYOA</td>
<td>+1<em>Q -1</em>QH2 +2.5<em>HEXT -0.5</em>O2</td>
</tr>
<tr>
<td>SDHA2R</td>
<td>+1<em>FAD -1</em>FADH -1<em>Q +1</em>QH2</td>
</tr>
<tr>
<td>PNT1A</td>
<td>-1<em>NAD +1</em>NADH +1<em>NADP -1</em>NADPH</td>
</tr>
<tr>
<td>PNT2A</td>
<td>+1<em>NAD -1</em>NADH -1<em>NADP +1</em>NADPH -2*HEXT</td>
</tr>
<tr>
<td>ATPAR</td>
<td>+1<em>ATP -1</em>ADP -1<em>PI -4</em>HEXT</td>
</tr>
<tr>
<td>GLCUP</td>
<td>+1<em>GLC -1</em>HEXT -1*GLCxt</td>
</tr>
<tr>
<td>GLCPTS</td>
<td>+1<em>G6P -1</em>PEP +1<em>PYR -1</em>GLCxt</td>
</tr>
<tr>
<td>GLUPR</td>
<td>+1<em>GL -1</em>GLxt</td>
</tr>
<tr>
<td>RIBUP</td>
<td>-1<em>ATP +1</em>ADP +1<em>PI +1</em>RIB -1*RIBxt</td>
</tr>
<tr>
<td>ACUPR</td>
<td>+1<em>AC -1</em>HEXT -1*ACxt</td>
</tr>
<tr>
<td>LACUPR</td>
<td>+1<em>LAC -1</em>HEXT -1*LACxt</td>
</tr>
<tr>
<td>FORUPR</td>
<td>+1<em>FOR -1</em>FORxt</td>
</tr>
<tr>
<td>ETHUPR</td>
<td>+1<em>ETH -1</em>HEXT -1*ETHxt</td>
</tr>
<tr>
<td>SUCCUPR</td>
<td>+1<em>SUCC -1</em>HEXT -1*SUCCxt</td>
</tr>
<tr>
<td>PYRUPR</td>
<td>+1<em>PYR -1</em>HEXT -1*PYRxt</td>
</tr>
<tr>
<td>PIU PR</td>
<td>+1<em>PI -1</em>PIxt</td>
</tr>
<tr>
<td>O2TXR</td>
<td>+1<em>O2 -1</em>O2xt</td>
</tr>
<tr>
<td>CO2TXR</td>
<td>+1<em>CO2 -1</em>CO2xt</td>
</tr>
<tr>
<td>ATPM</td>
<td>-1<em>ATP +1</em>ADP +1*PI</td>
</tr>
</tbody>
</table>

Continued on next page . . .
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADK</td>
<td>-1<em>ATP +2</em>ADP -1*AMP</td>
</tr>
<tr>
<td>VGRO</td>
<td>Biomass synthesis reaction</td>
</tr>
<tr>
<td>GLCxtI</td>
<td>+1*GLCxt</td>
</tr>
<tr>
<td>GLCxtO</td>
<td>-1*GLCxt</td>
</tr>
<tr>
<td>GLxtI</td>
<td>+1*GLxt</td>
</tr>
<tr>
<td>GLxtO</td>
<td>-1*GLxt</td>
</tr>
<tr>
<td>RIBxtI</td>
<td>+1*RIBxt</td>
</tr>
<tr>
<td>RIBxtO</td>
<td>-1*RIBxt</td>
</tr>
<tr>
<td>ACxtI</td>
<td>+1*ACxt</td>
</tr>
<tr>
<td>ACxtO</td>
<td>-1*ACxt</td>
</tr>
<tr>
<td>LACxtI</td>
<td>+1*LACxt</td>
</tr>
<tr>
<td>LACxtO</td>
<td>-1*LACxt</td>
</tr>
<tr>
<td>FORxtI</td>
<td>+1*FORxt</td>
</tr>
<tr>
<td>FORxtO</td>
<td>-1*FORxt</td>
</tr>
<tr>
<td>ETHxtI</td>
<td>+1*ETHxt</td>
</tr>
<tr>
<td>ETHxtO</td>
<td>-1*ETHxt</td>
</tr>
<tr>
<td>PYRxtI</td>
<td>+1*PYRxt</td>
</tr>
<tr>
<td>PYRxtO</td>
<td>-1*PYRxt</td>
</tr>
<tr>
<td>PIxtI</td>
<td>+1*PIxt</td>
</tr>
<tr>
<td>PIxtO</td>
<td>-1*PIxt</td>
</tr>
<tr>
<td>SUCCxtI</td>
<td>+1*SUCCxt</td>
</tr>
<tr>
<td>SUCCxtO</td>
<td>-1*SUCCxt</td>
</tr>
<tr>
<td>O2xtI</td>
<td>+1*O2xt</td>
</tr>
<tr>
<td>O2xtO</td>
<td>-1*O2xt</td>
</tr>
<tr>
<td>CO2xtI</td>
<td>+1*CO2xt</td>
</tr>
</tbody>
</table>

Continued on next page ...
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO2xtO</td>
<td>-1*CO2xt</td>
</tr>
<tr>
<td>ATPMAX</td>
<td>-1<em>ATP +1</em>ADP +1*PI</td>
</tr>
<tr>
<td>NADHMAX</td>
<td>+1<em>NAD -1</em>NADH</td>
</tr>
<tr>
<td>NADPHMX</td>
<td>+1<em>NADP -1</em>NADPH</td>
</tr>
<tr>
<td>G6PMAX</td>
<td>-1*G6P</td>
</tr>
<tr>
<td>F6PMAX</td>
<td>-1*F6P</td>
</tr>
<tr>
<td>R5PMAX</td>
<td>-1*R5P</td>
</tr>
<tr>
<td>E4PMAX</td>
<td>-1*E4P</td>
</tr>
<tr>
<td>T3P1MAX</td>
<td>-1*T3P1</td>
</tr>
<tr>
<td>3PGMAX</td>
<td>-1*3PG</td>
</tr>
<tr>
<td>PEPMAX</td>
<td>-1*PEP</td>
</tr>
<tr>
<td>PYRMAX</td>
<td>-1*PYR</td>
</tr>
<tr>
<td>ACCOAMAX</td>
<td>+1<em>COA -1</em>ACCOA</td>
</tr>
<tr>
<td>OAMAX</td>
<td>-1*OA</td>
</tr>
<tr>
<td>SUCAMAX</td>
<td>+1<em>COA -1</em>SUCCOA</td>
</tr>
<tr>
<td>AKGMAX</td>
<td>-1*AKG</td>
</tr>
<tr>
<td>Growth</td>
<td>-23<em>ATP +23</em>ADP +23<em>PI -1</em>Biomass</td>
</tr>
<tr>
<td>EDD</td>
<td>-1<em>D6PGC +1</em>2DDG6P</td>
</tr>
<tr>
<td>EDA</td>
<td>+1<em>T3P1 +1</em>PYR -1*2DDG6P</td>
</tr>
</tbody>
</table>

VGRO, the biomass synthesis reaction is as follows: -41.25*ATP -0.2*G6P +41.25*ADP -0.07*F6P +41.25*PI -0.12*T3P1 -3.54*NAD +3.54*NADH -1.49*3PG -0.51*PEP -2.83*PYR +3.74*COA -3.74*ACCOA +18.22*NADP -18.22*NADPH -0.89*R5P -0.36*E4P -1.78*OA -1.07*AKG +1*Biomass
Table 7.2: List of metabolite definitions used in this work

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Metabolite</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLC</td>
<td>D-glucose</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>G6P</td>
<td>Glucose 6-phosphate</td>
</tr>
<tr>
<td>ADP</td>
<td>Adenosine diphosphate</td>
</tr>
<tr>
<td>F6P</td>
<td>Fructose 6-phosphate</td>
</tr>
<tr>
<td>FDP</td>
<td>Fructose 1,6-diphosphate</td>
</tr>
<tr>
<td>PI</td>
<td>Inorganic phosphate</td>
</tr>
<tr>
<td>T3P1</td>
<td>Glyceraldehyde 3-phosphate</td>
</tr>
<tr>
<td>T3P2</td>
<td>Dihydroxyacetone phosphate</td>
</tr>
<tr>
<td>NAD</td>
<td>Nicotinamide adenine dinucleotide</td>
</tr>
<tr>
<td>NADH</td>
<td>Nicotinamide adenine dinucleotide reduced</td>
</tr>
<tr>
<td>13PDG</td>
<td>1,3-bis-Phosphoglycerate</td>
</tr>
<tr>
<td>3PG</td>
<td>3-Phosphoglycerate</td>
</tr>
<tr>
<td>2PG</td>
<td>2-Phosphoglycerate</td>
</tr>
<tr>
<td>PEP</td>
<td>Phosphoenolpyruvate</td>
</tr>
<tr>
<td>PYR</td>
<td>Pyruvate</td>
</tr>
<tr>
<td>AMP</td>
<td>Adenosine monophosphate</td>
</tr>
<tr>
<td>COA</td>
<td>Coenzyme A</td>
</tr>
<tr>
<td>CO2</td>
<td>Carbon dioxide</td>
</tr>
<tr>
<td>ACCOA</td>
<td>Acetyl-CoA</td>
</tr>
<tr>
<td>NADP</td>
<td>Nicotinamide adenine dimucleotide phosphate</td>
</tr>
<tr>
<td>D6PGL</td>
<td>D-6-Phosphate-glucono-delta-lactone</td>
</tr>
</tbody>
</table>

Continued on next page...
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Metabolite</th>
</tr>
</thead>
<tbody>
<tr>
<td>NADPH</td>
<td>Nicotinamide adenine dinucleotide phosphate reduced</td>
</tr>
<tr>
<td>D6PGC</td>
<td>D-6-Phosphate-gluconate</td>
</tr>
<tr>
<td>RL5P</td>
<td>D-Ribulose 5-phosphate</td>
</tr>
<tr>
<td>R5P</td>
<td>Ribose 5-phosphate</td>
</tr>
<tr>
<td>X5P</td>
<td>D-Xylulose-5-phosphate</td>
</tr>
<tr>
<td>S7P</td>
<td>sedo-Heptulose 7-phosphate</td>
</tr>
<tr>
<td>E4P</td>
<td>Erythrose 4-phosphate</td>
</tr>
<tr>
<td>OA</td>
<td>Oxaloacetate</td>
</tr>
<tr>
<td>CIT</td>
<td>Citrate</td>
</tr>
<tr>
<td>ICIT</td>
<td>Isocitrate</td>
</tr>
<tr>
<td>AKG</td>
<td>alpha-Ketoglutarate</td>
</tr>
<tr>
<td>SUCCOA</td>
<td>Succinyl-CoA</td>
</tr>
<tr>
<td>SUCC</td>
<td>Succinate</td>
</tr>
<tr>
<td>FAD</td>
<td>Flavin adenine dinucleotide</td>
</tr>
<tr>
<td>FADH</td>
<td>Flavin adenine dinucleotide reduced</td>
</tr>
<tr>
<td>FUM</td>
<td>Fumarate</td>
</tr>
<tr>
<td>MAL</td>
<td>Malate</td>
</tr>
<tr>
<td>LAC</td>
<td>Lactate</td>
</tr>
<tr>
<td>ETH</td>
<td>Ethanol</td>
</tr>
<tr>
<td>FOR</td>
<td>Formate</td>
</tr>
<tr>
<td>ACTP</td>
<td>Acetyl-phosphate</td>
</tr>
<tr>
<td>AC</td>
<td>Acetate</td>
</tr>
<tr>
<td>PPI</td>
<td>Pyrophosphate</td>
</tr>
<tr>
<td>GLX</td>
<td>Glyoxylate</td>
</tr>
<tr>
<td>GL</td>
<td>Glycerol</td>
</tr>
</tbody>
</table>

Continued on next page . . .
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Metabolite</th>
</tr>
</thead>
<tbody>
<tr>
<td>GL3P</td>
<td>Glycerol 3-phosphate</td>
</tr>
<tr>
<td>RIB</td>
<td>Ribose</td>
</tr>
<tr>
<td>Q</td>
<td>Ubiquinone</td>
</tr>
<tr>
<td>QH2</td>
<td>Ubiquinol</td>
</tr>
<tr>
<td>HEXT</td>
<td>External hydrogen</td>
</tr>
<tr>
<td>O2</td>
<td>Oxygen</td>
</tr>
<tr>
<td>GLCxt</td>
<td>External Glucose</td>
</tr>
<tr>
<td>GLxt</td>
<td>External Glycerol</td>
</tr>
<tr>
<td>RIBxt</td>
<td>External Ribose</td>
</tr>
<tr>
<td>ACxt</td>
<td>External Acetate</td>
</tr>
<tr>
<td>LACxt</td>
<td>External Lactate</td>
</tr>
<tr>
<td>FORxt</td>
<td>External Formate</td>
</tr>
<tr>
<td>ETHxt</td>
<td>External Ethanol</td>
</tr>
<tr>
<td>SUCCxt</td>
<td>External Succinate</td>
</tr>
<tr>
<td>PYRxt</td>
<td>External Pyruvate</td>
</tr>
<tr>
<td>PIxt</td>
<td>External Inorganic phosphate</td>
</tr>
<tr>
<td>O2xt</td>
<td>External Oxygen</td>
</tr>
<tr>
<td>CO2xt</td>
<td>External Carbon dioxide</td>
</tr>
<tr>
<td>Biomass</td>
<td>Biomass</td>
</tr>
<tr>
<td>2DDG6P</td>
<td>2-dehydro-3-deoxy-6-phospho-D-gluconate</td>
</tr>
</tbody>
</table>
Bibliography


[65] B. Teusink, A. Wiersma, D. Molenaar, C. Francke, W. M. de Vos, R. J. Siezen, and E. J. Smid. Analysis of Growth of *Lactobacillus plantarum* WCFS1 on a


