USING COMPUTATIONS TO ANALYZE
AND REDESIGN METABOLISM

A Dissertation in
Integrative Biosciences
by
Sridhar Ranganathan

© 2012 Sridhar Ranganathan

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2012
The dissertation of Sridhar Ranganathan was reviewed and approved* by the following:

Costas D. Maranas  
Donald B. Broughton Professor of Chemical engineering  
Dissertation Advisor & Chair of Committee

Reka Albert  
Professor of Physics and Biology

Eric Harvill  
Associate Professor in Microbiology and Infectious Diseases

Cooduvalli S. Shashikant  
Associate Professor of Molecular and Developmental Biology

Howard Salis  
Assistant Professor in Agricultural and Biological Engineering

Peter Hudson  
William Professor of Biology  
Head of the Integrative Sciences Graduate Program

*Signatures are on file in the Graduate School
Abstract

With the availability genome-wide datasets for various organisms, research in biology has moved towards a systems-level analysis that portrays a comprehensive picture of cellular physiology. In the recent past, complete inventories of all the known genetic capabilities of microorganisms have been built into computational models. In this regard, we present studies aimed at making use of computational models of metabolism and developing computational algorithms that help in analyzing and redesigning the metabolism of microorganisms. We use procedures developed in this thesis as a predictive tool for exploring genetic manipulations that lead to the overproduction of value-added biochemicals. The key aims addressed in this work are - **Aim 1**: to assemble novel biosynthesis routes from a known substrate or starting metabolite to biofuel molecules, **Aim 2**: to develop computational / mathematical procedures that suggest genetic interventions in microorganisms that lead to the overproduction of value-added biochemicals. Throughout this work, we rely on systems biology, graph-theory and novel optimization approaches (i.e. bilevel optimization) to address problems related to Aims 1 and 2 in metabolic engineering.

The potential of genetically engineering “user-friendly” microbes with non-native biosynthetic gene cascades has been identified as a promising method to produce biofuels during fermentation. To address this problem in Aim 1, we introduce a graph-based pathway prospecting approach that can help uncover all possible metabolic pathways to biofuel molecules such as 1-butanol. We derive computational predictions for these pathways by culling information from databases such as BRENDA, and KEGG.
Subsequently under Aim 2, we introduce a novel computational strain redesign procedure called OptForce custom-made to predict genetic interventions (i.e. up-/down-regulations, knockouts, knock-ins) that guarantees a pre-specified yield for the target biofuel molecule. In addition to its capability of predicting multiple genetic interventions at a time, OptForce is also primed to incorporate experimental data (i.e. metabolic flux analysis data) within the procedure before starting the procedure for redesign. In this work, we demonstrate the validity of OptForce by comparing computational predictions with metabolic engineering experiments for overproducing various biochemicals such as succinate, 1-butanol (from the pathways identified under Aim 1), flavanones, fatty and amino acids in *Escherichia coli*. 
# Table of Contents

List of Figures vii  
List of Tables ix  
Acknowledgements x  

1. Introduction and Literature Study 1  
  1.1. Background and Aims 1  
  1.2. Systems Biology and Metabolic Modeling 6  
  1.3. Computational Metabolic Engineering 9  
  1.4. Thesis Overview 12  

2. Biosynthetic Pathway Prospecting 20  
  2.1. Background 20  
  2.2. Computational Protocol 23  
  2.3. Results for Pathways to 3-hydroxypropional 26  
  2.4. Results for Pathways to 1-butanol 27  
  2.5. Concluding Remarks 29  

  3.1. Background 35  
  3.2. OptForce Methodology 38  
  3.3. OptForce for 1-butanol synthesis in *E. coli* 42  
  3.4. OptForce for succinate overproduction in *E. coli* 46  
  3.5. Discussion and Conclusions 53
4. Validation of OptForce: Improving the yield of malonyl-CoA and 
overproducing flavanones in *Escherichia coli* 69

4.1. Background 69

4.2. Materials and Methods 72

4.3. Results 76

4.4. Discussion and Conclusions 82

5. Validation of OptForce: Comprehensive Identification of Genetic 
Interventions to Understand Fatty Acid Biosynthesis 87

5.1. Background and Introduction 87

5.2. Flux changes that MUST happen in the network 89

5.3. Genetic interventions for overproducing hexanoate 92

5.4. Genetic interventions for overproducing octanoate and palmitate 93

5.5. Discussion and Conclusions 95

Synopsis 105

Appendix A: Computing maximum flux variability 108

Appendix B: Bilevel formulation for identification of MUST sets 111

Appendix C: Bilevel formulation for identification of FORCE sets 114

Bibliography 117
# List of Figures

1.1 General aims of thesis 18

1.2 Steps involved in reconstruction of metabolic networks 19

2.1 Graph-based procedure based Yen’s $k$- shortest path algorithm to mine novel pathways from reaction databases 32

2.2 Pathways from pyruvate to 3-hydroxypropional 33

2.3 Pathways from pyruvate to 1-butanol 34

3.1 Identification of fluxes that MUST change in the metabolic network for overproduction 60

3.2 An outline of the OptForce procedure to identify the minimal set of genetic interventions that guarantee the imposed yield 61

3.3 MUST$^U$ and MUST$^L$ set of reactions for 1-butanol synthesis in *E. coli* using the thiobutanoate pathway 62

3.4 MUST$^{UU}$, MUST$^{UL}$ and MUST$^{LL}$ set of reactions for overproducing 1-butanol in *E. coli* using the thiobutanoate pathway 63

3.5 Reactions that needs to be actively manipulated (FORCE sets) for overproducing 1-butanol 64

3.6 MUST$^U$ and MUST$^L$ set of reactions identified by OptForce for overproducing succinate in *E. coli* 65

3.7 MUST set of reactions considered two-at-a-time (pairs) for overproducing succinate in *E. coli* 66

3.8 MUST$^{UU}$, MUST$^{UUL}$, MUST$^{ULL}$, and MUST$^{LLL}$ set of reactions for overproducing succinate in *E. coli* 67
3.9 FORCE set of reactions for succinate overproduction in *E. coli* without and after addition of heterologous pyruvate carboxylase reaction 68

4.1 Metabolic interventions predicted using OptForce procedure for overproducing naringenin in *E. coli* 84

4.2 Plot between naringenin yield and the number of genetic interventions \((k)\) that were predicted by OptForce procedure 85

4.3 Maximum detected levels of flavanone naringenin produced from *E. coli* 86

5.1 Fatty acid biosynthesis pathways in *E. coli* 98

5.2 Venn diagram representing fluxes that MUST sets 99

5.3 FORCE set of reactions predicted by OptForce for the overproduction of hexanoic acid in *E. coli* 100

5.4 FORCE set of reactions predicted by OptForce for the overproduction of octanoic acid in *E. coli* 101

5.5 Impact of each genetic intervention predicted by OptForce on the yield of octanoate 102

5.6 Impact of each genetic intervention predicted by OptForce on the yield of palmitic acid 103

5.7 Venn diagram representing the shared genetic interventions predicted by OptForce for fatty acids of chain length \(C_6\) to \(C_{16}\) 104
List of Tables

1.1  List of genome-scale metabolic reconstructions for bacterial, archaeal and eukaryotic organisms 17

2.1  Maximum flux values for 1-butanol using the identified pathways from pyruvate 31

3.1  Comparison of the minimum guaranteed fluxes from OptKnock, OptReg and OptForce procedures for succinate production in E. coli 59
Acknowledgements

First and foremost, I want to thank my advisor Dr. Costas D. Maranas. It has been an honor to work with Prof. Maranas. I am ever so grateful to him for letting me join his group, providing me the independence and opportunities that helped me grow in this field and shaping my career. I have learnt more from him than any other person I have seen all my life. He has not only been a constant source of guidance over the years, but I also consider him my role model and a mentor. He has always had my best interest at heart and in order to live up to it I have to work harder to start deserving the honor of being his student.

I would like to thank Dr. Mattheos Koffas and his student Peng Xu from Rensselaer Polytechnic Institute (RPI) for their support. I also thank Prof. Jacqueline Shanks from Iowa State University, students and post-docs (Jesse Welsh, Ting Wei and Dr. Jong Moon Yoon) from her group who extended a very warm welcome for research in this thesis as an integral part of CBiRC projects. I would also like to thank Dr. Howard Salis for his enthusiasm in being another collaborator in this research. I must thank Dr. Reka Albert, Dr. Eric Harvill, Dr. Shashikant Cooduvalli and Dr. Howard Salis for serving on my doctoral committee and providing valuable insights to my research. I express my gratitude to my lab members, especially Alireza Zomorrodi, Dr. Vinay S. Kumar and Dr. Patrick Suthers, who have been my constant source of help and encouragement. I owe a lot to Dr. Anthony P. Burgard for letting me disturb him so many times in his busy days when I needed advice and help. Tony not only helped me learn metabolic engineering but also encouraged me towards doing meaningful research. Never in these years did he deny my request to clear my doubts or answer my questions in research.

Finally, I would like to dedicate this thesis to my parents, Ranganathan and Lakshmi. They have unconditionally dedicated their lives for me even under stressful times. Be it the best or worst periods of our lives, their love and support has been unwavering.
Chapter 1

Introduction and Literature Study

1.1. Background and Aims

Over the past few decades, we have witnessed an unprecedented progress in DNA sequencing and annotation projects (Liolios et al., 2006) across a wide range of species. At one end of the spectrum, we are at the forefront of pioneering research in the fields of sequence comparison (Altschul et al., 1990; Altschul et al., 1997), functional genomics (Giardine et al., 2005; Schwartz et al., 2000), human (Schuster et al., 2010) and microbial genomics (Eppinger et al., 2006; Linz and Schuster 2007), molecular evolution (Leebens-Mack et al., 2005), protein studies (Chothia and Lesk 1986), developmental and evolutionary biology (Hanson et al., 1999), plant biology (Boddu et al., 2006; Grotewold et al., 2006) and large-scale network biology (Albert 2005; Brown et al., 2004). Using data from these studies systems biologists and metabolic engineers have been successful in putting together models that depict the overall genetic map of organisms (Feist et al., 2009). By taking a complete inventory of all known genetic/metabolic capabilities of an organism, the hope in this area is to proactively convert genomic research data into models that replicate the physiological behavior of organisms. These “genome-scale models” accounts for gene to protein to reaction (GPR) associations, transcriptional regulation and other information that are compartmentalized according to organelle and
tissue-specific localizations. This mathematical formalism of metabolism have enabled the use of computational procedures to predict genetic interventions that help achieving biotechnological objectives that range from the overproduction of value-added chemicals from microbes (Alper et al., 2005b; Atsumi et al., 2008a; Atsumi and Liao 2008a; Bond and Lovley 2003; Burgard et al., 2003; Misawa et al., 1991; Nakamura and Whited 2003; Oliveira et al., 2005; Pharkya and Maranas 2006; Sauer et al., 2008; Scott et al., 2007) to understanding disease metabolism (Bosma 2003; Danpure 2006; Zelezniak et al., 2010) and pinpointing drug targets (Jamshidi and Palsson 2007; Lee et al., 2005a).

With over 35 genome-scale models available (see Table 1) for eukaryotic, prokaryotic and archaeal species, building computational algorithms that operate on these genome-scale models have become commonplace in several bioengineering projects. They are being used to identify novel biochemical routes to value-added products and to optimize carbon flow in an organism’s metabolism so as to maximize synthesis rates. As one of the first efforts, Costas Maranas group introduced computational frameworks that identified gene knockouts (OptKnock (Burgard et al., 2003; Pharkya et al., 2003)), reaction modulations (OptReg (Pharkya and Maranas 2006)) and identification of heterologous pathways (OptStrain (Pharkya et al., 2004)) that may lead to overproduction of desired biochemicals from microbial production hosts.

Although several computational algorithms were proposed following these studies (OptGene (Patil et al., 2005), OptORF (Kim and Reed 2010), RobustKnock (Tepper and Shlomi 2009), TRMR (Warner et al., 2010), CiED (Chemler et al., 2009)), a number of
bottlenecks and deficiencies still plague existing approaches. Some of the key limitations are: (i) existing procedures rely heavily on surrogate biological fitness functions such as maximizing growth rate (Burgard et al., 2003; Pharkya et al., 2003; Pharkya et al., 2004) and minimizing the number of adjustments in the microbial metabolism (Segre et al., 2002); (ii) in addition, these procedures are computationally limited in that they uncover only one genetic/metabolic intervention at a time from among a plethora of combinations available; (iii) the use of experimental data (i.e. metabolic flux, microarray data) that characterizes the metabolic confines of an organism are seldom not used in these procedures; (iv) incorporating non-native functionalities (in the form of heterologous genes/enzymes) are overlooked because of computational challenges in identifying them.

Hence, in this work, we directly address these challenges (see Figure 1.1) by building on research milestones already achieved in the recent past. The framework developed in this thesis is designed to make use of efficient computational techniques to address the existing shortcomings and enable improved capability in identifying genetic interventions that can redirect metabolic flows not only in microbial strains but also will be reusable to analyze other complex organisms in the future. We address two specific aims in this thesis:

Aim 1: Identification of novel biosynthetic pathways to biofuels: Microorganisms from diverse environments naturally synthesize ethanol that has been used as biofuel. Recently, long-chain alcohols (C3-C6) have emerged as alternate biofuels because of their ease of storage and higher energy density (Atsumi and Liao 2008b). Except a few strains of Clostridia (Janssen 2004; Jones and Woods 1986), none of the known microbes
produce 1-butanol and other longer chain alcohols. Hence, the potential of engineering “user-friendly” microbes with non-native biosynthetic pathways has been identified as a promising method. An important goal of this research is requires extending the metabolic confines of microbial hosts by recruiting non-native biosynthetic pathways. Here, we anticipate the need to quickly and efficiently generate biosynthetic routes from various sets of databases available to the scientific community. As the first of our aims, we propose to computationally derive predictions for assembling pathways to biofuel candidate molecules by culling information from BRENDA (Barthelmes et al., 2007; Chang et al., 2009) and KEGG (Kanehisa et al., 2008; Kanehisa et al., 2010; Kanehisa et al., 2006) databases (spanning about ~ 250,000 metabolites and ~60,000 enzymatic transformations). The combinatorial nature of this problem poses a significant challenge to existing procedures such as OptStrain (Pharkya et al., 2004) that rely on optimization algorithms that are computationally expensive. In addition, several techniques such as PathMiner (McShan et al., 2003), PathComp (Kanehisa et al., 2006), Pathway Tools (Karp et al., 2002; Karp et al., 2009), MetaRoute (Blum and Kohlbacher 2008), PathFinder (Goesmann et al., 2002) and UM-BBD Pathway Prediction System (Ellis et al., 2006) that currently exist are mainly used for metabolic pathway reconstructions by matching putative enzymes with reference pathways. With their contribution to strain optimization being limited, we propose to develop a graph-based min-path procedure that can exhaustively identify all possible biochemical routes from a source to a target metabolite (i.e. biofuel candidate). The procedure will be designed to remain tractable even when the database entries reach hundreds of thousand reactions. In this work, we
present a few test cases of our implementation by comparing existing pathways to long-chain alcohols from biochemicals naturally synthesized in *E. coli* and yeast.

**Aim 2: Computational Strain Design:** Recently, a variety of optimization-based tools have been made available for identifying gene knockouts (i.e. OptKnock (Burgard et al., 2003; Pharkya et al., 2003)), knock-ins (i.e. OptStrain (Pharkya et al., 2004)) and up/down-regulations (i.e. OptReg (Pharkya and Maranas 2006)) that lead to the overproduction of target biochemicals from a microbial strain. The challenge with these methods is that they generate engineering strategies out of myriads of possibilities only one-at-a-time thus limiting the array of choices presented to the biotechnologists. In addition, metabolic flux data (e.g. through metabolic flux analysis (MFA) experiments) obtained for a wild-type strain are not directly integrated with the strain optimization process. To remedy these limitations, we propose to develop the next generation of strain optimization procedure, which we call as OptForce (Ranganathan et al., 2010a). The key concept here is that instead of looking at engineering strategies one-at-a-time, we classify all the fluxes in a metabolic model depending on whether or not they must increase, decrease or knocked down in the face of an overproduction objective. The classification scheme is not only limited to individual reactions but also for pairs, triplets etc. of reactions while actively making use of experimental flux data available for the wild-type strain before engineering. Subsequently, we use a bilevel optimization procedure that can identify the minimal set of engineering interventions from the reactions classified that must be actively engineered to guarantee an overproduction. Under this aim, we present
results from OptForce procedure for the overproduction of succinate, 1-butanol (from the pathways identified from Aim 1) and flavanones from *E. coli*.

1.2. Systems Biology and Metabolic Modeling

The aims of this thesis discussed in the previous sections require a holistic view rather than a reductionist approach to analyzing biological phenomenon. Given the complexity of metabolic reactions, its analysis and subsequent suggestions to redesign its pattern can be made only when all or most of the known components are taken into consideration. The task of assembling whole-cell biochemical data that can closely resemble the actual physiological state of an organism depends on the scope and usage of models. This pursuit to generate detailed lists of biological components through experimental analysis, determine their interactions, and generate genome-wide data sets has led to the emergence of systems biology (Ideker et al., 2001). Although there are many individual definitions, “*systems biology involves an interactive interplay between high-throughput ‘wet’ experiments, theory and computational modeling*” (2006). In essence, it involves the treatment of biological organisms as black (or grey) boxes whose inner structure and functioning can be modeled and analyzed by varying an internal or external condition (Kell et al., 2005). The result is a better understanding of the inner make-up and workings of the system as a whole. In biology, this translates into the study of what a cell secretes under given set of conditions that may be defined as one or more perturbations to its genetic setup. In essence, systems biology has promised a holistic approach to understanding biology as opposed to the conventional reductionist approach. Using
systems biology approaches, researchers in the past have successfully reconstructed models that depict protein-protein interactions, gene regulatory networks, cell-signaling pathways and metabolism, which forms the core of this thesis.

Metabolic models are now becoming a valuable tool for studying the systems biology of metabolism. In short, a metabolic model is a collection of all possible biochemical pathways that are known to exist in a given physiological state for an organism. They represent structured knowledge bases that abstract pertinent information on the biochemical transformations taking place within specific target organisms. The conversion of a reconstruction into a mathematical format facilitates a myriad of computational biological studies, including evaluation of network content, hypothesis testing and generation, analysis of phenotypic characteristics and metabolic engineering. Data from myriads of sources are gleaned during the reconstruction of a metabolic network of an organism. These include biochemical data (KEGG (Kanehisa et al., 2010), BRENDA (Barthelmes et al., 2007; Chang et al., 2009; Schomburg et al., 2002) etc.), protein localization information (PSORT (Gardy et al., 2005)), organism-specific data (e.g. EcoCyc, PyloriGene) and genomic information (GOLD (Liolios et al., 2006), ModelSEED (Overbeek et al., 2005) etc.).

In general, the reconstruction process (see Figure 1.2) follows a multi-stage iterative procedure (Thiele and Palsson 2009). Creating a draft metabolic model for the organism forms the primary stage of the metabolism reconstruction process. Using the genome annotation of the target organism and biochemical databases, a collection of genome-
encoded metabolic functions is documented. In this step, we march through the whole genome of the organism using keyword or gene ontology (GO) searches. For those genes that have a metabolic fingerprint, we connect a reaction associated with the specific gene. Note that the initial list of biochemical data may falsely include some reactions due to missing, wrong or incomplete annotations. In the second stage of the reconstruction process, we refine the model by scrutinizing the existence of each and every reaction in the draft metabolism by evaluating them against literature and experiments. This manual step is very important because of two reasons. Firstly, not all annotations have high confidence scores and secondly, biochemical databases list enzyme activities found in various organisms, which may or may not be present in the target organism.

Subsequently, we screen the draft model for substrate and cofactors involved in every reaction. As a rule of thumb, we assume that enzymes that are associated with only one reaction in KEGG or BREnda do not require refinement. However, for enzymes that are associated with multiple reactions, we carry additional searches in organism-specific databases and biochemical experimental studies for the particular organism. In certain cases, data for other organisms that are phylogenetically closer to the target organism are used. The stoichiometry of the reactions and the standard Gibbs free energy of formation ($\Delta G$) helps in further refining the draft model. Algorithms such as PSORT (Gardy et al., 2005) and PA-SUB (Lu et al., 2004) can be used to predict the localization of proteins in a multi-compartment organism based on amino acid sequences. One of the most important steps in model construction is defining the constituents of biomass for the organism. In every metabolic model, a biomass reaction is included that accounts for all
known ingredients and their fractional contributions towards the formation of overall biomass. This is normally evaluated by measuring the amounts of various biomass precursors such as phospholipids, amino acids and fatty acids when the cells grow in the log phase. Finally, we test the metabolic model by evaluating its ability to produce each and every precursor that is required for producing biomass. This generally leads to the identification of missing reactions and/or metabolites. We fill these gaps by either adding or removing reactions or metabolites from the model. Various computational procedures have been developed in order to identify and fill these gaps (Kumar and Maranas 2009; Satish Kumar et al., 2007). Thus, the procedure of reconstructing a metabolic model is a highly iterative and semi-automatic process that relies on the scope and usage of the model that defines its “completeness”.

1.3. Computational Metabolic Engineering

Metabolic engineering is about the analysis and modification of metabolic pathways. The field emerged during the past decade and powered by techniques from applied molecular biology and reaction engineering, it is becoming a focal point of research activity in biological and biochemical engineering. With the availability of organism-specific genome-scale metabolic models, we have been able to obtain wider views of metabolism. In addition, these models have allowed researchers to treat metabolic pathways quantitatively because of the emphasis it places on the metabolic fluxes and their control. Metabolic flux, here, refers to the rate at which each and every biotransformation takes place in a cell. Thus, in a metabolic network, the flux is the fundamental determinant of
cellular physiology and the most critical parameter of a metabolic pathway. Metabolic engineering is an analytical bioengineering method to quantify fluxes and their control to suggest genetic manipulations. As metabolic fluxes and pathways are the core of metabolic engineering, it is important to elaborate their definitions. We define a metabolic pathway to be any sequence of feasible and observable biochemical reaction steps connecting a specified set of metabolites. The pathway is then defined by the rate at which input metabolites are processed to form output metabolites. Since, there exists a balance of fluxes under steady-state conditions in the metabolic network, the analysis of metabolism through this method is often referred to as flux balance analysis (FBA).

Considerable interest in the redirection of metabolic fluxes towards industrial and pharmaceutical products has developed in recent years. In order to employ a systems metabolic engineering approach, we need to have a computational metabolic model for an organism of interest. A variety of experimental measurements for the metabolic fluxes, gene expression, metabolite concentration and enzyme activity are available to quantify the state of metabolism. Computational models and approaches can be used for integrating all such datasets to quantify metabolic fluxes and uncover avenues for novel biotechnological advancements. Once these metabolic states are known, computational metabolic engineering approaches can be used to identify which environmental or genetic perturbations would improve cellular phenotypes, such as the production of desired chemicals. In a typical metabolic engineering project, the above processes are looped and performed over several iterations depending upon the availability of experimental data and the targets for the product yield.
In the recent past, computational procedures that operate on metabolic models have been instrumental in suggesting novel genetic perturbations (up-regulations, knockouts, knock-ins etc.) for experimentalists. These procedures have been developed based on pathway-based and optimization-based approaches (Reed et al., 2010). In the pathway-based approach, identification of the relevant pathways of a metabolic network is essential for finding effective metabolic engineering strategies. These pathways can also help derive minimal media requirements for an organism and assess the robustness and redundancy of key metabolic pathways. Alternative to the pathway approach, an optimization-based approach can help identify mutations that would improve desired phenotypes. By using a stoichiometric relationship between various pathways, we can identify solutions that maximize or minimize a stated objective. One of the earliest efforts for computational strain design using stoichiometric models was the OptKnock procedure (Burgard et al., 2003; Pharkya et al., 2003) which identified gene knock-outs that force the coupling of the desired overproduction target to growth by the use of bilevel optimization problems. Later, OptReg (Pharkya and Maranas 2006) extended this procedure to allow for not only knock-outs but also up or down reaction regulation, and OptStrain (Pharkya et al., 2004) allowed for knock-ins of non-native functionalities from a comprehensive universal database of reactions to enable production of desired biochemicals. These sets of tools have been customized and modified by other groups using genetic algorithms (OptGene, (Patil et al., 2005)) and other evolutionary approaches (Chemler et al., 2009). Recently, a bilevel optimization procedure, OMNI, was developed to identify and correct differences in flux distributions between in silico predictions and in vivo data in the context of strain
design (Herrgard et al., 2006). Built on similar algorithmic protocols OptORF (Kim and Reed 2010), RobustKnock (Tepper and Shlomi 2009), and TRMR (Warner et al., 2010) are some of the more recent computational procedures that can identify genetic engineering strategies for targeted overproductions. So far, computational strain design procedures have been applied for a variety of metabolic engineering projects including the overproduction of lactic acid (Burgard et al., 2003; Fong et al., 2005), succinate (Cox et al., 2006; Lee et al., 2005b; Lee et al., 2002; Patil et al., 2005; Wang et al., 2006), 1,3-propanediol (Burgard et al., 2003), hydrogen (Pharkya et al., 2004), amino acids (Pharkya et al., 2003), L-lysine (Contador et al., 2009), L-valine (Park et al., 2007), threonine (Lee et al., 2007), lycopene (Alper et al., 2005b; Alper et al., 2005c), ethanol in E. coli (Anesiadis et al., 2008; Pharkya and Maranas 2006; Trinh et al., 2008) and Saccharomyces cerevisiae (Bro et al., 2006) and bioelectricity in Geobacter sulfurreducens (Mahadevan et al., 2006). In this thesis, we outline a more recent procedure called OptForce (Ranganathan et al., 2010a) that can help suggest genetic intervention strategies aimed at targeted overproductions.

1.4. Thesis Overview

The following chapters of this thesis are focused on development of optimization and graph-based computational procedures to analyze and redesign the metabolism of microorganisms to satisfy bioengineering objectives. In this thesis, we complement computational and mathematical analysis of metabolism with a network-based view of cellular physiology. Mathematical procedures presented in this thesis not only account for
accurate quantification of biochemical data, but also provides a fair approximation to parameters that cannot be measured experimentally. A graph-based view of the metabolism helps greatly in understanding the biological reasoning behind the mathematics involved. Chapters in this thesis have been arranged in a specific order such that one may appreciate, at first, the hidden network of genes, proteins, reactions and metabolites inside an organism. Subsequently, towards the later chapters, we introduce novel computational procedures that help in redesigning the genetic setup of organisms with the aim of achieving specific bioengineering targets.

Chapter 2 introduces a graph-based procedure for identifying novel biochemical synthesis routes to candidate biofuel molecules. The potential of engineering microorganisms with non-native pathways for the synthesis of long-chain alcohols has been identified as a promising route to biofuels. In this chapter, we describe computationally derived predictions for assembling pathways for the production of biofuel candidate molecules. A graph-based algorithm is illustrated that, by culling information from BRENDA and KEGG databases, identifies all possible pathways that link the target product with metabolites present in the production host. We demonstrate this procedure by suggesting new pathways and genetic interventions (in Chapter 3) for the overproduction of 1-butanol using the metabolic model for E. coli. The graph-based search method recapitulated all recent discoveries based on the 2-ketovaline intermediate and hydroxybutyryl-CoA but also pinpointed one novel pathway through thiobutanoate intermediate that to the best of our knowledge has not been explored before. Specifically, the search technique employed in this framework requires the assembly of biochemical
data from various databases such as BRENDA, KEGG etc. We deploy the algorithm to search for non-native functionalities that can empower industry-friendly microorganisms such as *E. coli* to synthesize biofuels. This chapter highlights key results published in:


In addition to identifying novel synthesis routes to biochemicals, predicting metabolic interventions leading to the overproduction of these biochemicals in microbial strains is considered the next step in metabolic engineering. In Chapter 3, we explore numerous computational procedures that are in use to predict metabolic engineering interventions for targeted overproductions. In spite of their wide usage, these methods rely on surrogate biological objectives (e.g. maximize growth rate or minimize metabolic adjustments) and do not make use of flux measurements often available for the wild-type strain. To remedy this, we introduce the OptForce procedure that identifies all possible engineering interventions by classifying reactions in the metabolic model depending upon whether their flux values must increase, decrease or become equal to zero to meet a pre-specified overproduction target. We hierarchically apply this classification rule for pairs, triples, quadruples, etc. of reactions. This leads to the identification of a sufficient and non-redundant set of fluxes that *must* change (i.e., MUST set) to meet a pre-specified overproduction target. Starting with this set we subsequently extract a minimal set of fluxes that must actively be *forced* through genetic manipulations (i.e., FORCE set) to
ensure that all fluxes in the network are consistent with the overproduction objective. We demonstrate our OptForce framework for succinate production in *Escherichia coli* using the most recent *in silico* *E. coli* model, iAF1260. The method not only recapitulates existing engineering strategies but also reveals non-intuitive ones that boost succinate production by performing coordinated changes on pathways distant from the last steps of succinate synthesis. Data included in this chapter were published in:


Computational procedures operating on metabolic models for microorganisms differ in their efficacies due to the rigor involved in performing mathematical computations and the closeness of their predictions to the real-time physiology. We circumvent this limitation faced by other procedures by building OptForce to incorporate experimental measurements available at any given quantitative and qualitative format. In Chapter 4, we validate the genetic engineering interventions predicted by OptForce framework by testing the computationally predicted genetic interventions for overproducing flavanones from *E. coli*. Flavanones (specifically, naringenin) are secondary metabolites mainly produced in plant species. Owing to their implications in diabetes, cancer, heart disorders and Parkinson’s disease, there is a need to develop better strains of industry-friendly microorganisms that can produce flavanones at high yields. In this important work, we derive perspectives about completeness of the metabolic models and bottlenecks involved
in translating computational predictions into experiments. In spite of this, the experimental yields reported in this chapter pinpoint a 4.5-fold increase in the yield of flavanones that results in the best naringenin-producing strain of *E. coli* in a lab-scale. A complete description of this work can be found in:


Biosynthesis of fatty acid-like chemicals from renewable carbon sources has attracted significant attention in recent years. Free fatty acids can be used as precursors for the production of fuels or chemicals. Usually, an acyl carrier protein (ACP) thioesterase gene results in the production free fatty acids yeast and some photosynthetic microorganisms. However, the metabolism of fatty acids is tightly regulated in the metabolism of most microbes because of its association with membrane biosynthesis. Moreover, the use of complex microorganisms for producing fatty acids is increasingly difficult because of their relatively unknown metabolic map. Hence, in Chapter 5, we present a comprehensive analysis of the fatty acid metabolism in *E. coli*. Specifically, we march through the fatty acid chain varying in carbon length from 6-14 and enlist all possible genetic manipulations that would force *E. coli* to synthesize free fatty acids even when it produces biomass. Finally, in Chapter 6, we summarize the key contributions of the preceding chapters and provide future directions to move forward.
Table 1.1: List of genome-scale metabolic reconstructions for bacterial, archaeal and eukaryotic organisms.

<table>
<thead>
<tr>
<th>Bacteria</th>
<th>Acinetobacter baylyi [35], Bacillus subtilis [36], Clostridium acetobutylicum [37], [38], Corynebacterium glutamicum [39], Escherichia coli [40], [41], [42], Geobacter metallireducens [43], Geobacter sulfurreducens [44], Helicobacter pylori [45], [46], Lactobacillus plantarum [47], Lactococcus lactis [26], Mannheimia succiniciproducens [48], Mycobacterium tuberculosis [49], [32], Mycoplasma genitalium [50], Neisseria meningitides [51], Porphyromonas gingivalis [52], Pseudomonas aeruginosa [53], Pseudomonas putida [54], Rhizobium etli [55], Salmonella Typhimurium [56], Staphylococcus aureus [57], [58], Streptomyces coelicolor [59]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eukarya</td>
<td>Aspergillus nidulans [60], Aspergillus oryzae [61], Homo sapiens [62], Leishmania major [63], Mus musculus [64], [65], Saccharomyces cerevisiae [66], [67], [68], [69], [70]</td>
</tr>
<tr>
<td>Archaea</td>
<td>Halobacterium salinarum [71], Methanosarcina barkeri [72].</td>
</tr>
</tbody>
</table>
**Figure 1.1 - General aims of thesis:** Aim 1 - to build a computational framework to predict genetic interventions that maximizes the yield of a specific biochemical of interest; Aim 2 - to leverage the availability of databases of reactions by using graph-based procedures that can trace novel biosynthetic pathways between a starting and a target biochemical.
Figure 1.2: Steps involved in reconstruction of metabolic networks
Chapter 2

Biosynthetic pathway prospecting

2.1. Background

Increasing demands for renewable energy and environmental concerns have stimulated an interest towards the production of second generation biofuels from renewable sources (Stephanopoulos 2007). For the past few decades, bio-ethanol was considered as a substitute for transportation fuels. More recently, long-chain alcohols (C3-C5) have also emerged as biofuel alternatives because of their higher energy density and ease of storage (Atsumi and Liao 2008b). Microorganisms from diverse environments naturally produce ethanol during fermentation. However, the natural synthesis of higher alcohols is not as commonplace with the exception of certain 
Clostridia
 strains (Janssen 2004; Jones and Woods 1986). One possible production alternative for 1-butanol and 1-propanol is to use native pathways in 
Clostridium acetobutylicum
 (Formanek et al., 1997; Lee et al., 2009; Lin and Blaschek 1983; Nair and Papoutsakis 1994; Sillers et al., 2008). An alternative approach is to integrate non-native pathways into standard microbial production hosts (i.e., 
Escherichia coli
 or yeast) by exploiting the conversion of key intermediary amino acids into long-chain alcohols (Hanai et al., 2007; Yan and Liao 2009). In this regard,

numerous efforts have been made in the recent past to clone and express Clostridia genes (butyryl-CoA dehydrogenase, bcd) responsible for the production of 1-butanol in E. coli (Boynton et al., 1996; Fontaine et al., 2002; Wiesenborn et al., 1988). Homologs and isoenzymes of bcd from Megasphaera elsdenii (Becker et al., 1993; O'Neill et al., 1998) and crotonoyl-CoA reductase (ccr) from Streptomyces coelicolor (Wallace et al., 1995) have been tested. Recently, enzymes catalyzing the final steps of the Ehrlich pathway (Sentheshanuganathan 1960) in yeast were recruited in E. coli to convert 2-ketoacids into 1-butanol and isobutanol (Atsumi et al., 2008b). The global aim to converting biomass to energy has led to an increased interest in transferring non-native metabolic pathways and enzymes into industrial production hosts such as E. coli (Atsumi et al., 2008a; Shen and Liao 2008) or Saccharomyces cerevisiae (Steen et al., 2008).

An important goal of this research requires extending the metabolic confines of microbial hosts by recruiting non-native biosynthetic pathways. So far, studies concerning the incorporation of heterologous pathways relied largely on human intuition and literature reports followed by experimentation (Bode and Muller 2005; Wenzel and Muller 2005). Nowadays, rapidly expanding compilations of biotransformations such as KEGG (Kanehisa et al., 2008) and BRENDA (Chang et al., 2009) are increasingly being prospected to identify biosynthetic routes to long-chain alcohols. With a combined size that accounts for over 60,000 enzymatic reactions and 250,000 metabolites, these databases include reactant and product designation, stoichiometric coefficients, organism assignment, and occasional thermodynamic information for pathways (Bader et al., 2006). Several optimization and graph-based methods have been employed to
computationally assemble novel biochemical routes from these sources. Given a set of reactions (i.e., Universal database) the OptStrain (Pharkya et al., 2004) procedure uses a mixed-integer linear optimization representation to identify the minimal number of reactions to be added (i.e. knock-ins) into a genome-scale metabolic model to enable the production of the new molecule. However, the developed universal database, at the time, was limited to only approximately 4,000 reaction entries. The combinatorial nature of the problem poses a significant challenge to the OptStrain methodology as the number of reaction database entries increase from a few to tens of thousands.

At the expense of not enforcing stoichiometric balances graph-based algorithms have inherently better-scaling properties for exhaustively identifying all min-path reaction entries that link a source with a target metabolite. Hatzimanikatis et. al. (Hatzimanikatis et al., 2005) introduced a graph-based heuristic approach to identify all possible biosynthetic routes from a given substrate to a target chemical by hypothesized enzymatic reaction rules. Recently, a new scoring algorithm (Cho et al., 2010) was introduced to evaluate and compare novel pathways generated using enzyme-reaction rules. The identified pathways may involve conversions for which no enzymatic activity has been isolated for before. While this could shed light to truly novel production avenues, it may be more time-consuming to implement. In addition, several techniques such as PathMiner (McShan et al., 2003), PathComp (Kanehisa et al., 2006), Pathway Tools (Karp et al., 2002; Karp et al., 2009), MetaRoute (Blum and Kohlbacher 2008), PathFinder (Goesmann et al., 2002) and UM-BBD Pathway Prediction System (Ellis et al., 2006) are in use to search for bioconversion routes in reaction databases. Most of these methods, so
far, have been employed to aid metabolic pathway reconstructions by matching putative enzymes with reference pathways while their contribution towards strain optimization has so far been limited.

In this chapter, we introduce a min-path graph procedure for in overcoming the complexity associated with exhaustively identifying all possible ways of linking a source with a target metabolite. The procedure is designed to remain tractable even when reaction database entries reach hundred of thousands. The first step, in this effort involved the incorporation of reaction and metabolite entries from both KEGG (Kanehisa et al., 2008) and BRENDA (Chang et al., 2009) databases into a single repository. A customized min-path algorithm (Yen 1971) is then employed to compute all possible pathways that enable the bio-production of a target alcohol molecule. We further scrutinize the identified pathways by first incorporating them into the genome-scale metabolic model of the production host microorganism and subsequently examining their maximum theoretical yields, number of enzymatic steps needed and cofactor availability. We demonstrate our integrated framework by exploring pathways from pyruvate (produced in E. coli) to 1-butanol.

2.2. Computational Protocol

The graph-based procedure discussed here is aimed at elucidating all possible biochemical routes from compounds found in the metabolic network of a desirable production host to a target molecule of interest. Alternatively, the procedure can also be
used to track native routes that may increase productivity over known synthesis pathways by restricting the reaction entries to the ones present in the metabolic model of the production host. To provide the search procedure with known metabolic routes, we downloaded the most up-to-date version of the KEGG database (Kanehisa et al., 2008) and extracted approximately 9,000 reactions and 16,000 metabolites. Unfortunately, the KEGG database does not contain complete production pathways of long-chain alcohols. The KEGG database offers a comprehensive overview of metabolic reactions including key parameters such as stoichiometry, reversibility of reactions and metabolites. However, it encompasses only about ~10,000 reactions and metabolites. Even when the KEGG database is growing in its size by the day, it does not include connecting pathways to long-chain alcohols of interest. For instance, reactions that involve isobutanol as a product are disconnected from the central metabolism due to lack of adjoining reactions.

On the other hand, the BRENDA database included about ~55,000 enzymatic reactions known to occur and importantly, includes pathways that are relevant to biofuel production. The difficulty in using BRENDA database is that it includes generic reactions (e.g. acetyl-CoA + a primary alcohol = CoA + an acetyl ester) that hamper pathway prospecting and metabolic modeling efforts. In order to combine the benefits of these two databases, we merge reaction and metabolite entities from the two databases into a single repertoire. A major challenge here is to deal with incompatibilities of representation, duplications (i.e. metabolites with multiple names across the two databases and in some cases within the BRENDA database) and errors. We have found from our initial studies that about 3% of all metabolites and 8% of all reactions have the aforementioned problems in their nomenclature. We carefully eliminated and/or corrected reactions with
incorrect stoichiometry and generic descriptors to the best of our abilities. The integrated
database as published (Ranganathan and Maranas 2010) contained about 9,921 reactions
and 17,013 metabolites. This effort has served as a stepping stone for the MetRxn
database (Kumar et al., 2011) that our group has developed very recently. We therefore,
added a few hundred reaction entries from the BRENDA database (Chang et al., 2009)
that are relevant to biofuels production to restore the metabolic connectivity to long-chain
alcohols. It is important to note that we did not globally reconcile the entire KEGG
database with BRENDA database (containing ~250,000 metabolites and 67,191
reactions). Instead, for all reactions in BRENDA associated with the synthesis of the
target alcohol, we manually recorded identifiers for all the reactants, products and
stoichiometric coefficients and integrated them with the KEGG entries into a single
database.

Sorting out the naming inconsistencies for compounds was the most time consuming
step. To accomplish this, we made use of available synonym data from PubChem (Wang
et al., 2009) to arrive at unique metabolite identifiers. Reactions with generic (e.g.
metabolites named as “alcohol”, “aldehyde” etc.) descriptions for reactant/product
compounds, unknown stoichiometry and the ones that involve macromolecules (e.g.
RNAP) were excluded. The integrated database used in this work spans 9,921 reactions
and 17,013 metabolites from both BRENDA and KEGG.

We used the min-path procedure as depicted in Figure 2.1 to trace all possible paths
between a source and a target metabolite. We first computationally transformed the
information contained within the stoichiometric coefficients \( (S_{ij}) \) that track participation of metabolites in reactions into a directed metabolite-to-metabolite graph \( (N_{ij}) \) where nodes represent metabolites. A directed arc with a weight of one exists between two nodes if one or more reactions in the database allow the direct bioconversion from one metabolite to the other. If no such reaction exists then a very large cost value is assigned to signify that their direct interconversion is disallowed. Small molecules (e.g. water, carbon dioxide) and cofactors (e.g. NADP, ATP) are involved in a large number of reactions and thus can link reaction steps that do not share any additional metabolites. We therefore exclude all such associated directed arcs before employing the shortest path algorithm. We next compute all \( k \)-shortest “loopless” pathways (Yen 1971) between a source and a target alcohol molecule. We start from the shortest path \( (k = 1) \) and exhaustively sample the combinatorial space of alternative pathways by subsequently eliminating arcs, one at a time, belonging to the shortest pathway. We recompute the shortest path until we record all “\( k - 1 \)” shortest possible metabolic linkages to the target molecule.

### 2.3. Results for pathways to 3-hydroxypropional

As a simple test case, we implemented the algorithm to find novel biochemical routes to 3-hydroxypropionate (3HP). 3HP is a biodegradable polymer compound used as precursor for the synthesis of acrylic polymers and 1,3-propanediol that is used as biofuel. It is listed in the top 20 value-added chemicals by the US Department of Energy (DOE) (Werpy et al., 2004). The algorithm not only pinpointed the existing chemical
synthesis route (direct decarboxylation of 3-hydroxypropanaldehyde (Haas et al., 2000)) but also identified the fermentation pathway that involves 3-hydroxypropanoyl-CoA as an intermediate (see Figure 2.2) that has been patented by Cargill (Gokarn et al., 2001; Liao et al., 2007).

2.4. Results for pathways to 1-butanol

In this section, we demonstrate our min-path procedure by identifying all synthesis routes using KEGG and BRENDA database entries for producing 1-butanol from pyruvate. Traditionally, two distinct synthesis routes have been employed in E. coli for the production of 1-butanol. The first pathway involves a fermentative transformation of pyruvate and acetyl-CoA to 1-butanol by the action enzymes from C. acetobutylicum (Atsumi et al., 2008a). The second pathway takes advantage of enzymes with broad-range substrate specificity to convert natural amino acids in E. coli into ketoacid precursors (Atsumi et al., 2008b; Shen and Liao 2008) and eventually 1-butanol. In both pathways, pyruvate acts as an important precursor and a branching metabolite for butanol synthesis (Clomburg and Gonzalez). The fate of pyruvate at the end of glycolysis depends on the engineering strategies imparted to the production host. Therefore, here we selected pyruvate as a source metabolite in exploring pathways to 1-butanol (sink metabolite).

Figure 2.3 and Table 2.1 illustrates all identified pathways from pyruvate to 1-butanol using the integrated reaction database. With the exception of the thiobutanoate pathway
(present in the BRENDA database), all other pathways involved butanoyl-CoA and 1-butanal as shared intermediates that are converted to 1-butanol using secondary alcohol dehydrogenase (adhE) from *C. acetobutylicum*. The min-path procedure recapitulated both the fermentative and ketoacid pathways for 1-butanol synthesis (shown in dotted lines). In addition, the algorithm uncovered a number of possible transformations to butanoyl-CoA involving intermediate metabolites that are produced in *E. coli*. For example, pyruvate can be converted into acetyl-CoA using pyruvate dehydrogenase natively present in *E. coli*. However, the conversion from acetyl-CoA to butanoyl-CoA is not favored because 1-butanal produced along the pathway is used up as a co-reactant along other reactions in the same pathway. This severely reduces the flux of the 1-butanol to less than 10 mmol/g.DW.hr which is about ten times less than the yields from existing pathways (Atsumi et al., 2008b; Atsumi and Liao 2008b). Similarly, pathways involving methylmalate and methylbutanoate as intermediates require cofactors, which in turn, adversely reduce the yield of 1-butanol. Upon integrating these reactions in the metabolic model of *E. coli*, we estimated that the maximum theoretical yield of 1-butanol synthesis was only around 32 mmol/g.DW.hr.

The thiobutanoate pathway recruits a decarboxylase and a reductase enzyme and defines a novel synthesis route distinct from the two existing pathways. Instead of using dehydrogenases to convert butyraldehyde into 1-butanol, the new pathway proceeds with the transamination of methionine into 2-oxomethylthiobutanoate and eventually into 1-butanol. Notably, a native transaminase (E.C. 2.6.1.42) enzyme in *E. coli* is known to catalyze the conversion of L-methionine to L-glutamate with 2-ketoglutarate as a co-
reactant (Kagamiyama and Hayashi 2000). The intermediate product, 2-methylthiobutanoate, is subsequently decarboxylated (E.C. 4.1.1.72) to 3-methylthiopropanal. This conversion is native in *Lactococcus lactis* (Smit et al., 2005). Subsequently, 3-methylthiopropanal is reduced (E.C. 1.1.1.265) to 1-butanol by a reductase present in yeast (Perpete and Collin 1999). It is important to note that the decarboxylase reaction removes a considerable amount carbon in the form of carbon dioxide, reducing the yield of 1-butanol by ~22% in comparison to the ketoacid pathway.

### 2.5. Concluding Remarks

We have presented a graph-based min-path procedure that combines metabolic information from online databases (KEGG and BRENDA) to identify all possible biochemical synthesis routes to target biofuel candidates. The results for 1-butanol pathways reveal several new heterologous synthesis routes that can be computationally evaluated for overexpression and cloning experiments. Our algorithm was able to identify existing pathways (ketoacid and fermentative pathways) used for 1-butanol production. Interestingly, we the results also suggested several native synthesis routes to precursors of 1-butanol in *E. coli*. For example, seven pathways from pyruvate to butanoyl-CoA involved intermediate metabolites produced by naturally occurring enzymes in *E. coli*. However, the yield of 1-butanol using these pathways was limited. In addition, the algorithm also uncovered a new alternative route to 1-butanol synthesis through the thiobutanoate pathway. Although, the decarboxylation of methylthiobutanoate reduced 1-butanol production, the computationally derived yield was comparable to the existing
strains (Atsumi et al., 2008a; Atsumi et al., 2008b; Atsumi and Liao 2008b). The genetic manipulations required in a microbe hosting these pathways determine the efficiency of these pathways. Although, pyruvate is the main precursor for both ketoacid and thiobutanoate pathways, the use of amino acid precursors make these pathways distinct in terms of yields. In the next chapter, we demonstrate how the thiobutanoate pathway can be efficiently used in *E. coli* for the synthesis of 1-butanol.

The graph-based procedure can be used to identify alternative synthesis routes found entirely within the production host by selectively exploring pathways that are native. Currently, the procedure uses all the biotransformations found in the KEGG database (Kanehisa 2002; Kanehisa et al., 2008) and a selected set of reactions from the BRENDA (Chang et al., 2009) database. The min-path search procedure remains tractable for much larger compilations of reactions/metabolites. It is to be noted that the interventions proposed by OptForce pertain to the reactions. A complete mapping between the reactions and the genes is required for projecting the results at the gene-level.
Table 2.1: Maximum flux values for 1-butanol using the identified pathways from pyruvate.

<table>
<thead>
<tr>
<th>List of identified pathways between pyruvate and 1-butanol and intermediate reactions</th>
<th>Maximum flux value for 1-butanol (in mmol / gDW hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Thiobutanoate Pathway:</strong></td>
<td></td>
</tr>
<tr>
<td>L-glutamate + Pyruvate $\rightleftharpoons$ 2-ketoglutarate + L-alanine</td>
<td></td>
</tr>
<tr>
<td>2-ketoglutarate + L-methionine $\rightarrow$ 2-keto methylthiobutyrate + L-glutamate</td>
<td></td>
</tr>
<tr>
<td>2-keto methylthiobutyrate $\rightleftharpoons$ 3-methylthiopropanal + CO$_2$</td>
<td></td>
</tr>
<tr>
<td>3-methylthiopropanal + NADPH $\rightarrow$ 1-butanol + NADP</td>
<td>74.8</td>
</tr>
<tr>
<td><strong>Fermentative Pathway:</strong></td>
<td>100</td>
</tr>
<tr>
<td>CoA + NAD + Pyruvate $\rightarrow$ Acetyl-CoA + CO$_2$ + NADH</td>
<td></td>
</tr>
<tr>
<td>2 Acetyl-CoA $\rightarrow$ Acetoacetyl-CoA + CoA</td>
<td></td>
</tr>
<tr>
<td>Acetoacetyl-CoA + NADPH + H$^+$ $\rightarrow$ 3-Hydroxybutanoyl-CoA + NADP</td>
<td></td>
</tr>
<tr>
<td>3-Hydroxybutanoyl-CoA $\rightarrow$ Crotonoyl-CoA + H$_2$O</td>
<td></td>
</tr>
<tr>
<td>Crotonoyl-CoA + NADH + H$^+$ $\rightarrow$ Butanoyl-CoA + NAD+</td>
<td></td>
</tr>
<tr>
<td>Butanoyl-CoA + NADPH + H$^+$ $\rightarrow$ 1-Butanal + CoA + NADP+</td>
<td></td>
</tr>
<tr>
<td>1-Butanal + NADH + H$^+$ $\rightarrow$ 1-Butanol + NAD+</td>
<td></td>
</tr>
<tr>
<td><strong>Ketoacid Pathway:</strong></td>
<td>100</td>
</tr>
<tr>
<td>Pyruvate $\rightarrow$ L-Threonine (Isoleucine, Serine and Threonine Metabolism)</td>
<td></td>
</tr>
<tr>
<td>L-Threonine $\rightleftharpoons$ Oxobutanoate + NH$_3$</td>
<td></td>
</tr>
<tr>
<td>Oxobutanoate + Acetyl-CoA + NAD $\rightarrow$ 2-ketovalerate + CO$_2$ + NADH</td>
<td></td>
</tr>
<tr>
<td>2-ketovalerate $\rightarrow$ 1-Butanal + CO$_2$</td>
<td></td>
</tr>
<tr>
<td>1-Butanal + NADH + H$^+$ $\rightarrow$ 1-Butanol + NAD+</td>
<td></td>
</tr>
<tr>
<td><strong>Methylmalate Pathway:</strong></td>
<td>32</td>
</tr>
<tr>
<td>Acetyl-CoA + Pyruvate + H$_2$O $\rightarrow$ 2-Methylmalate + CoA</td>
<td></td>
</tr>
<tr>
<td>Acetyl-CoA + 2-Methylmalate $\rightarrow$ Acetate + Butanoyl-CoA + CO$_2$ + H$_2$O</td>
<td></td>
</tr>
<tr>
<td>Butanoyl-CoA + NADPH + H$^+$ $\rightarrow$ 1-Butanal + CoA + NADP+</td>
<td></td>
</tr>
<tr>
<td>1-Butanal + NADH + H$^+$ $\rightarrow$ 1-Butanol + NAD+</td>
<td></td>
</tr>
<tr>
<td><strong>Other Pathways:</strong></td>
<td>9.3</td>
</tr>
<tr>
<td>Pyruvate $\rightarrow$ Acetaldehyde + CO$_2$</td>
<td></td>
</tr>
<tr>
<td>Acetaldehyde + CoA + NAD+ $\rightarrow$ Acetyl-CoA + NADH + H$^+$</td>
<td></td>
</tr>
<tr>
<td>ATP + Butanoic acid + CoA $\rightarrow$ AMP + Diphosphate + Butanoyl-CoA</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2.1: Graph-based procedure based Yen’s $k$-shortest path algorithm to mine novel pathways from reaction databases
Figure 2.2: Pathways from pyruvate to 3-hydroxypropanal
Figure 2.3: Pathways from pyruvate to 1-butanol. Pathway shown in widely spotted dots depicts the ketoacid pathway while the one shown in thick dotted lines represent the fermentative pathway that is currently in use for 1-butanol synthesis. The newly identified thiobutanoate pathway is shown in gray.
Chapter 3

OptForce: An optimization-based framework for identifying all possible genetic interventions for targeted overproductions†

3.1. Background

An overarching challenge for metabolic engineers is to optimize the conversion of biomass and other renewable resources into useful metabolic products through fermentation and other biological conversions (Stephanopoulos et al., 1998; Zeikus 1980). Metabolic reaction fluxes are a fundamental determinant of the cell physiology, primarily because they provide a degree of engagement of various pathways in metabolic processes (Mavrovouniotis and Stephanopoulos 1990). Earlier efforts addressed parts of metabolism with an emphasis on dynamics using kinetic approximations of reaction rates (Heinrich and Rapoport 1974; Kacser and Burns 1973; Torres and Voit 2002; Voit 1992). These approximations included the popular the S-system representation (Delgado and Liao 1997; Galazzo and Bailey 1989; Galazzo and Bailey 1990; Savageau 1976; Torres et al., 1996; Voit 1992) and Michaelis-Menten based descriptions (Hatzimanikatis et al.,

† This chapter has been published as: Ranganathan et. al. (2010) “An optimization-based framework for identifying all possible genetic interventions for targeted overproductions”. PLoS Comp Biol 6(4), (e1000744)
Despite many success stories, it is increasingly becoming accepted that strain optimization requires taking account of the totality of biotransformations present in a production strain. This global view of metabolism is needed to enable the complete elucidation of all carbon fluxes diverted away from the desired product, diagnose unbalanced cofactor requirements limiting the extent of reactions as well as remedy deficiencies in the production of all biomass components leading to growth arrest.

Flux balance analysis (FBA) has emerged as an important framework (Bonarius et al., 1997; Edwards et al., 2002; Lee et al., 2006; Schilling et al., 2000) to assess the metabolic potential of a microbial production system. By taking a complete inventory of all (known) metabolic capabilities of an organism, FBA can assess the maximum possible yield of a desired product for different substrates and growth levels (Burgard and Maranas 2001). Given the lack of a truly predictive nature, FBA results must be carefully interpreted as performance limits and supplemented with MFA data whenever possible. Shortly after the introduction of FBA, a number of computational tools emerged that identified strain engineering modifications leading to targeted overproductions. One of the earliest efforts was the OptKnock (Burgard et al., 2003) procedure that suggested gene knockouts leading to targeted overproductions. A bilevel optimization framework was postulated that computationally coupled the desired overproduction target to growth with unforeseen, at the time, implications for strain stability. Later, OptReg (Pharkya and Maranas 2006) extended OptKnock to consider not only knockouts but also overexpressions and down regulations of various reactions in the network. In addition,
OptStrain (Pharkya et al., 2004) allowed for knock-ins of non-native functionalities from a comprehensive universal database of reactions to enable production of desired biochemicals. Evolutionary search procedures for solving the resulting combinatorial optimization problems were explored in OptGene (Patil et al., 2005) and applied for the production of succinic acid, glycerol and vanillin in yeast. The Ensemble Modeling approach (Tran et al., 2008) circumvented the kinetic modeling approach by incorporating flux measurements from knockout and enzyme overexpression experiments. Recently, the GDLS algorithm (Lun et al., 2009) was used for reduced metabolic models employing GPR associations to predict gene knockouts for succinate and acetate production in E. coli. So far, computational strain design procedures have been applied for a variety of metabolic engineering projects including the overproduction of lactic acid (Burgard et al., 2003; Fong et al., 2005), succinate (Cox et al., 2006; Lee et al., 2005b; Lee et al., 2002; Patil et al., 2005; Wang et al., 2006), 1,3-propanediol (Burgard et al., 2003), hydrogen (Pharkya et al., 2004), amino acids (Pharkya et al., 2003), L-lysine (Contador et al., 2009), L-valine (Park et al., 2007), threonine (Lee et al., 2007), lycopene (Alper et al., 2005b; Alper et al., 2005c), ethanol in E. coli (Anesiadis et al., 2008; Pharkya and Maranas 2006; Trinh et al., 2008) and Saccharomyces cerevisiae (Bro et al., 2006) and bioelectricity in Geobacter sulfurreducens (Mahadevan et al., 2006).

The use of computational tools operating on metabolic reconstructions to identify strain modifications is becoming commonplace. Nevertheless, a number of shortcomings plague all existing approaches. All are sequential in nature generating a single
engineering strategy per run thus requiring multiple restarts to generate a set of candidate list of alternatives (i.e., typically less than ten) that is dwarfed by the myriads of engineering possibilities afforded by genome-scale models spanning thousands of reactions. Furthermore, in the absence of kinetic descriptions OptKnock and other methods rely on the maximization of surrogate biological fitness functions (e.g. maximization of biomass yield (Burgard et al., 2003) or minimization of metabolic adjustments MOMA (Segre et al., 2002)) to estimate flux redirection upon strain engineering. These estimates may or may not be an accurate representation of how metabolism responds to genetic or environmental perturbations with significant consequences in the quality of the suggested re-designs. Existing methods do not pro-actively make use of flux measurements for the wild-type and/or an engineered strain to identify which fluxes need to be actively engineered in response to a production target. To remedy these limitations, we introduce a new computational framework termed OptForce that identifies all possible engineering interventions for a wild-type strain characterized by specific metabolic flux data consistent with an imposed production target(s).

3.2. OptForce Methodology

3.2.1. Characterizing the wild-type strain using experimental data

The key concept of OptForce is to maximally resolve which fluxes (or combinations thereof) must depart away from the range of values allowed to span in the wild-type strain in response to an overproduction target. This maximal range of flux variability for
the wild-type strain can be elucidated by iteratively maximizing and minimizing each flux (Burgard and Maranas 2001; Mahadevan and Schilling 2003) subject to the stoichiometric constraints, uptake conditions and MFA flux data (either exact values or ranges) whenever available for the wild-type strain. This yields a set of lower and upper bounds for every flux in the metabolic network. Narrow ranges for the bounds are indicative of fluxes whose value is well bracketed given the information available for the wild-type strain whereas wide ranges indicate fluxes that are not significantly limited by the imposed (stoichiometric, MFA, etc.) constraints. Flux ranges can be used not only for characterizing the metabolic flux limits of the wild-type strain but also for identifying all flux combinations consistent with a single (i.e., \( v > v_{\text{target}} \)) or multiple desired overproduction targets (see Appendix A for optimization formulations). The flux ranges consistent with the overproduction target(s) can be derived as before by iteratively maximizing and minimizing every flux in the metabolic network subject to stoichiometric constraints, uptake conditions and overproduction targets.

### 3.2.2. Changes that MUST happen in the network for overproduction

Contrasting the flux ranges for the (wild-type) metabolic network against the ones consistent with the overproduction target(s) provides the cornerstone of OptForce. Figure 1 pictorially illustrates the proposed concept. By superimposing the flux ranges for a given reaction in the wild-type vs. the overproducing network a number of possible outcomes are revealed. If there is any degree of overlap between the two reaction flux ranges (Figure 3.1a) then it may be possible to achieve the overproduction target without changing the value of the corresponding reaction flux in the wild-type strain. In contrast,
if the flux ranges for a reaction in the wild-type metabolic network are completely to the left (Figure 3.1b) or to the right (Figure 3.1c) of the corresponding ranges for the overproducing metabolic network then the overproduction target cannot be achieved unless the reaction flux is directly or indirectly changed. The case depicted in Figure 3.1b calls for an increase whereas the one shown in Figure 3.1c requires a decrease in the reaction flux value. Note that if the reaction flux range collapses to zero then the corresponding reaction needs to be eliminated (e.g., through a gene knock-out). The gap between the two flux ranges quantifies the degree of required reaction flux modification.

We refer to reaction fluxes that must increase (see Figure 3.1b) in the face of the imposed overproduction requirements as $\text{MUST}^U$ whereas the ones that must decrease (see Figure 3.1c) as $\text{MUST}^L$. Fluxes of reactions with overlapping ranges (see Figure 3.1a) between the wild-type and overproducing network do not provide any imperatives on network modifications when considered one at a time. Therefore, we further scrutinize them by considering sums of two reaction fluxes at a time and subsequently calculating their ranges in the wild-type and overproducing metabolic networks. As was the case of single reaction fluxes, three outcomes are possible (see Figure 3.1d-f). Non-overlapping ranges imply that in the overproducing network \textit{either one or the other reaction flux} (but not necessarily both) must increase (Figure 3.1d) or decrease (Figure 3.1e) in value. These pairs of reactions form sets $\text{MUST}^{UU}$ and $\text{MUST}^{LL}$ respectively. One can extend this concept further by analyzing the range of not just the sum of two fluxes but also their difference for the wild-type and overproducing networks (see Figure 3.1f). As before, non-overlapping ranges imply that \textit{either the first reaction flux must increase or the}
second reaction flux must decrease. By extension, these pairs of reactions form the equivalent sets \( \text{MUST}^{UL} \) and \( \text{MUST}^{LU} \), respectively. One can systematically extend this analysis by considering sums and/or differences of three, four, etc. reactions at a time. Collectively, the derived sets (e.g., \( \text{MUST}^L \), \( \text{MUST}^U \), \( \text{MUST}^{UU} \), \( \text{MUST}^{LLL} \), \( \text{MUST}^{UULL} \), etc.) encompass all the necessary reaction flux changes that MUST take place in the wild-type metabolic network for the desired overproduction. Appendix B introduces a bilevel formulation for identifying all MUST sets.

3.2.3. Identifying the minimal set of genetic interventions from MUST sets

This reaction flux modification does not necessarily have to be realized by actively engineering the gene that codes for the enzyme catalyzing the reaction (e.g., through changed promoter, codon usage, or gene disruption/knock-out). It may come about indirectly by propagating through stoichiometry the effect of modifications occurring in other parts of metabolism (e.g., coupled reactions in series, cofactor coupling, etc.). The next step of OptForce is to identify how the collective set of changes (encoded within the MUST sets) can be imparted on the wild-type metabolic network with the minimal number of direct interventions (i.e., knock-up/down/outs). We make use of a max-min bilevel optimization problem (Figure 8) to identify the minimal set of engineering interventions that forces the yield of the product to the target value. As depicted in Figure 3.2, the OptForce procedure identifies metabolic interventions that guarantee the imposed yield even when the network fights against these interventions. Modeled as a worst-case optimization problem, we iteratively solve this problem by increasing the number of direct manipulations \( (k) \) in each step until the target yield is achieved. OptForce first
identifies the interventions that have the largest contribution towards meeting the overproduction target thus providing a way to prioritize the implementation of genetic interventions. We have found that when $k$ is increased, genetic interventions found before are largely conserved. In addition, at the optimal yield, the use of integer cuts allows for the identification of alternate optimal solutions that can serve as alternate genetic intervention choices.

### 3.3. OptForce for 1-butanol synthesis in E. coli

Here, we demonstrate our min-path procedure (outlined in Chapter 2) and OptForce by identifying all synthesis routes using KEGG and BRENDA database entries for producing 1-butanol from pyruvate. We first select promising pathways and subsequently integrate them with the genome-scale metabolic model of E. coli, iAF1260 (Feist et al., 2007). Using OptForce (Ranganathan et al., 2010b) we next pinpoint metabolic engineering strategies for overproduction. We have already listed all possible pathways between pyruvate and 1-butanol in the previous chapter.

In this section, we integrate these reactions in the iAF1260 metabolic model of E. coli and use OptForce (Ranganathan et al., 2010b) to identify metabolic interventions to meet an imposed overproduction target. The identified results are contrasted against the ones derived when the ketoacid pathway is integrated into the E. coli model. In both the case studies, the initial strain is first characterized by estimating the maximal range of flux variability using the intracellular flux measurements (Shimizu 2004) available for the
wild-type strain of *E. coli*, BW25113. The OptForce employs a bilevel optimization procedure to first identify the reaction fluxes that must increase or decrease (MUST sets) outside the wild-type flux ranges to meet the overproduction target. A minimal set of direct interventions (i.e. knock-up/down/outs) that guarantee a pre-specified yield for 1-butanol is next extracted from the MUST sets. All abbreviations for reactions and metabolites adhere to the iAF1260 metabolic model conventions.

### 3.3.2. 1-butanol Synthesis using Thiobutanoate Pathway

Figure 3.3 lists the identified MUST set of reactions considered one reaction at-a-time. The yield for 1-butanol was set at 95% of its theoretical maximum, while allowing the production of 5% biomass to support growth. The thiobutanoate pathway branches away from 2-ketoglutarate along the oxidative arm of the TCA cycle. In order to increase the pool of oxaloacetate available for the TCA cycle, the fluxes of reactions in the glycolytic pathway (PGI, PGM, PGK, PPC etc.) increase beyond their initial ranges. Many reactions in the pentose phosphate pathway (e.g. GND, TKT1/2, TALA etc.) were also classified in the MUST\(^U\) sets. The increase in the fluxes for these reactions replenishes the glycolytic intermediary metabolites. Since, methionine is required as an important precursor for 1-butanol pathway, reactions in methionine biosynthesis (e.g., CYSTL, METS, MTHFR2, CYSS) also members of the MUST\(^U\) set. The fluxes of reactions leading to competing by-products, pyruvate kinase (PYK) and pyruvate formate lyase (PFL) decrease below their initial ranges. Since biomass production is reduced to 5% of its theoretical maximum, reactions in amino acid biosynthesis that are directly coupled to growth appear in the MUST\(^L\) sets.
As expected, more complex flux changes are revealed in the network of MUST$^{UU}$, MUST$^{UL}$ and MUST$^{LL}$ sets shown in Figure 3.4. These results underscore the importance of increasing the flux through the oxidative arm of the TCA cycle (FUM etc.) or at the same time negating the drain towards by-products such as acetate and ethanol. Additionally, in the MUST$^{UU}$ set, the flux of propanoyl CoA:succinyl CoA transferase (PPCSCT) or the flux of succinyl CoA synthetase (SUCOAS) must increase. Both of these fluxes are in close proximity to 2-ketoglutarate, which is an important branching metabolite in the TCA cycle for the thiobutanoate pathway. We also carry out this hierarchical classification by considering three reactions at-a-time. The increase in fluxes for IPPMI, IMPC and AIRC3 further boosts the synthesis of precursors for methionine through amino acid biosynthetic pathways.

It is to be noted that the MUST set of reactions represent the changes that must take place in the metabolic network for overproduction that can be directly or indirectly imparted by means of metabolic interventions. OptForce identifies the minimal set of reaction interventions (culled from the MUST sets) that forces the target yield for 1-butanol. Figure 3.5 shows the FORCE set of reactions for overproducing 1-butanol in *E. coli* using the thiobutanoate pathway. Up regulating one of the two glycolytic fluxes, glucose-6-phosphate isomerase (PGI) or phosphoglycerate mutase (PGM), replenishes phosphoenolpyruvate available for the anaplerotic conversion to oxaloacetate. The up-regulation for phosphoenol pyruvate carboxylase (PPC) results in increasing the amount of oxaloacetate for the TCA cycle. Increase in fluxes of PPCSCT or SUCOAS ensure the availability of
2-ketoglutarate for transamination along the thiobutanoate pathway. In addition, the FORCE sets also include knockouts for pyruvate formate lyase (PFL) to reduce the drain towards by-products (acetate and ethanol) and methylenetetrahydrofolate dehydrogenase (MTHFD) to prevent the drain of L-methionine away from the thiobutanoate pathway. These coordinated set of interventions lead to a guaranteed yield for 1-butanol of 73 mmol / g.DW.hr.

### 3.3.3. 1-butanol synthesis using ketoacid pathway

Figure 3.5 contrasts the metabolic pathways and branching points for the ketoacid and thiobutanoate pathways on a metabolic map of *E. coli*, respectively. While the thiobutanoate pathway branches out from a TCA cycle intermediate, pyruvate serves as an important precursor for 1-butanol produced via the ketoacid pathway. We integrated the reactions along this pathway to iAF1260 metabolic model of *E. coli* and applied our OptForce procedure to predict the MUST sets and subsequently, the FORCE sets. Figure 3.5 (right) shows the FORCE set of eight engineering interventions for 1-butanol synthesis in *E. coli* using the ketoacid pathway. Herein, OptForce suggested the up-regulation in the fluxes of reactions that convert key amino acids to 1-butanol precursors (i.e., serine deaminase (SERD) and methylglyoxal synthase (MGSA)). Presumably due to the proximity of the ketoacid pathway to the synthesis routes for natural fermentation products (acetate, ethanol, formate, lactate etc.), the down-regulations for pyruvate formate lyase (PFL) and lactate dehydrogenase (LDH) are needed to reduce carbon drain. Additionally, down-regulation of TCA cycle reactions, fumarate reductase (FRD3) and
aconitase (ACONTa/b), also appear as essential network changes to ensure overproduction.

A notable difference between the two cases is the down-regulation of phosphogluconate dehydrogenase (GND) using the ketoacid pathway. While the flux of GND must increase for the thiobutanoate pathway (i.e., member of MUSTU set), OptForce suggests that its flux must be reduced to facilitate 1-butanol synthesis when using the ketoacid pathway. In addition, while PGI and PGM were identified as up-regulations for the thiobutanoate pathway no glycolytic reactions were up-regulated in the FORCE set for the ketoacid route. Since the ketoacid pathway branches out from precursors synthesized at the end of glycolytic pathway, OptForce indicates that the depletion of carbon can be minimized through a number of down-regulations for competing pathways without the need of overexpressing glycolytic enzymes. However, in the thiobutanoate case, the anaplerotic phosphoenol pyruvate carboxylase (PPC) is required to replenish oxaloacetate and to sustain an increased flux through the TCA cycle.

3.4. OptForce for succinate overproduction in *E. coli*

In this section, we benchmark the OptForce framework by identifying metabolic interventions that lead to the overproduction of succinate using the latest genome-scale metabolic model for *E. coli*, iAF1260 (Feist et al., 2007). There have been extensive efforts to re-engineer metabolic pathways in *E. coli* for improving succinate yield (Bunch et al., 1997; Chatterjee et al., 2001; Cox et al., 2006; Donnelly et al., 1998; Gokarn et al.,
2000; Hong and Lee 2001; Hong and Lee 2002; Lin et al., 2005a; Lin et al., 2005b; Lin et al., 2005c; Lin et al., 2004; Sanchez et al., 2005a; Sanchez et al., 2005b; Sanchez et al., 2006; Stols and Donnelly 1997; Stols et al., 1997; Vemuri et al., 2002b). We explored the production of succinate under anaerobic conditions to take advantage of the inherently high yield towards succinate (Lin et al., 2005c). Under anaerobic conditions, the synthesis route for succinate takes place along the reductive arm of the TCA cycle and involves the conversion of oxaloacetate (OAA) to malate, fumarate and eventually to succinate. The initial strain was characterized by estimating the maximal range of flux variability using intracellular flux measurements available for the wild-type strain of *E. coli*, MG1655 (Sanchez et al., 2006). The OptForce algorithm was used to explore engineering interventions under two different scenarios. First, we identified strain modifications that guarantee 100% theoretical yield for succinate. Not surprisingly, these engineering modifications come at the expense of completely negating biomass formation. In the second case study, we examined the effect of adding the activity of the heterologous pyruvate carboxylase (*pyc*) gene to the *iAF1260* model of *E. coli*. Note that the abbreviations and directionalities of reactions adhere to the *iAF1260* metabolic model definitions.

### 3.4.1. Succinate overproduction in *E. coli*

Figure 3.6 lists the identified MUST\textsuperscript{U} and MUST\textsuperscript{L} sets of reactions whose fluxes must depart the original ranges. Note that because all members of set MUST\textsuperscript{L} involve fluxes set to zero we re-designate them as MUST\textsuperscript{X} to signify that they all correspond to reaction
eliminations. Not surprisingly, the transport reaction directing succinate out of the cytosol (SUCt3rpp) was classified into the MUST\textsuperscript{U} whereas transport reactions for competing by-products such as ethanol (ETOHzrpp, ETOHtex), acetate (ACtex), formate (FORtex) and acetaldehyde (ACALtpp, ACALDtex) were completely blocked (i.e., members of the MUST\textsuperscript{X} set). In addition, a number of reactions from histidine (ATPPRT, HISTD, HISTP, HSTPT, IG3PS, IGPDH, PRAMPC, PRATPP and PRPPS) and methionine metabolism (AHCYSNS, DHPTDCs, HCYSMT and RHCE) were also set to zero. Note that these reactions are essential for amino acid biosynthesis and are fully coupled to growth. Therefore, the drain of carbon flux from the pentose phosphate pathway towards histidine and methionine synthesis is prevented thus halting the production of biomass.

While results for MUST\textsuperscript{U} and MUST\textsuperscript{L} involve primarily intuitive negations of by-products formation, sets MUST\textsuperscript{UU}, MUST\textsuperscript{UL} and MUST\textsuperscript{LL} allude to more complex flux re-allocations (see Figure 3.7). For example, in the MUST\textsuperscript{UU} set the increase in the flux for reaction phosphoenolpyruvate carboxylase (PPC) can only be compensated by the simultaneous increase in the flux of five TCA cycle reactions (i.e., MALS, CS, ACONTa, ACON Tb and ICL). This implies that at least one of two possible avenues for succinate production must be increased under anaerobic conditions (see Figure 3.7a). Specifically, either the flux along the traditional succinate synthesis route through the reductive pathway that converts oxaloacetate (oaa) to malate and fumarate or the flux through the glyoxylate shunt needs to increase. Interestingly, the higher succinate yield of the latter mechanism due to NADH availability has been implemented in \textit{E. coli} by deactivating
the iclR repressor (to activate the glyoxylate bypass) under anaerobic conditions by (Sanchez et al., 2005b).

Figure 3.7a reveals that a number of flux up-regulations (e.g., PPC, PGM, CS, ICL, ACONTa/b, PGM, ATPS4rpp, ALDD2x, ACALD) and down-regulations (e.g., PFL, TPI, RPI, ASPTA, PGK) appear frequently as choices in multiple pairs. These mutually compensatory flux changes can be more clearly discerned by fusing all interacting components from MUST\textsuperscript{UU}, MUST\textsuperscript{UL} and MUST\textsuperscript{LL} into a single graph (see Figure 3.7b) where fluxes that increase are shown in green and those that decrease are shown in red. The importance of PPC up-regulation is manifested by the fact that as many as ten separate reaction flux modifications would be needed to replace it. Similarly, the decrease in flux through PFL can only be compensated by up-regulating the flux of four reactions along the glyoxylate shunt while the down-regulation of the flux through ENO can only be replaced by the up-regulation of four reactions supplying flux to the TCA cycle. The compensatory interconnections in Figure 3.7b suggest that not all depicted flux modifications are simultaneously needed to reach the desired phenotype (i.e., 100% yield of succinate). Instead, all flux modifications implied by sets MUST\textsuperscript{LL}, MUST\textsuperscript{UU} and MUST\textsuperscript{UL} can be satisfied by up- or down-regulating a minimal set of reactions. We identified all such minimal reaction flux modification sets and depicted them in the form of a Boolean diagram in Figure 3c. As expected, up-regulation of the flux through PPC is a consensus choice while the up-regulation of only one out of ACONTa, ACONTb, CS and ATPS4rpp is needed. Interestingly, the down-regulation of PFL which diverts flux
towards organic acids such as formate, lactate, acetate, ethanol, etc. emerged as a required change despite its relatively low connectivity in the diagram of Figure 3.7b.

Figure 3.8 depicts the reaction flux changes that happen when considering three reaction fluxes at a time (one out of three). The reactions are denoted as ovals where green nodes represent the flux of the reaction that increases and red nodes indicate those that decrease. They span up-regulations (MUST\textsuperscript{UUU}), down-regulations (MUST\textsuperscript{LLL}) or combinations thereof (MUST\textsuperscript{UUL} and MUST\textsuperscript{ULL}). Figure 3.8a re-affirms the key role of up-regulating PPC but also reveals the importance of redirecting the flux of reactions from pyruvate metabolism (i.e. PFL, ACS, ACALD, ACKr, PTAr) towards acetyl-CoA. Furthermore, Figure 3.8a reveals that the decrease in the value of the flux for phosphotransacetylase (PTAr) and acetate kinase (ACKr) reduces the export of acetate and increases the amount of acetyl-CoA available for the glyoxylate pathway. These results are in agreement with the knockouts for \textit{ackA} and \textit{pta} in strain SBS990MG constructed for succinate synthesis (Sanchez et al., 2005b). The reaction modifications implied in MUST\textsuperscript{LLL}, MUST\textsuperscript{UUU}, MUST\textsuperscript{UUL} and MUST\textsuperscript{ULL} can also be distilled into a minimal set of modifications (see Figure 3.8b). Many of these modifications were present in Figure 3.7, however, a number of new imperatives such as reducing the flux of FUM emerge. One can methodically, continue to identify additional constraints that need to be satisfied to achieve the desired phenotype by looking into higher-order combinations of fluxes.

We next used the bilevel optimization formulation to identify the minimal set of reaction modifications (i.e., FORCE set) that guarantee the imposed yield (100% succinate yield).
Note that the identified MUST reaction flux modifications were added as constraints in the FORCE set formulation. The identified minimal set of forced modifications (see Figure 3.9a) is comprised of ten different interventions. The up-regulation of PPC and CS ensures that the pool of oxaloacetate is diverted towards the TCA cycle. The up-regulation for PGK and TPI increases the glycolytic activity providing precursor metabolites such as phosphoenol pyruvate, oxaloacetate etc. to succinate synthesis. The down-regulation of PFL, GLUDy and ASPTA prevents the formation of by-products such as formate, lactate, ethanol, glutamate, aspartate and 2-ketoglutarate. The up-regulation for ACALD converts any residual acetate back into acetyl-CoA, which in turn is converted to succinate. Notably, for two such interventions there exist two equivalent alternatives. The first one involves the up-regulation of either of ACONTa/b isozymes to ensure conversion of citrate into glyoxylate and succinate. The second one requires either the down-regulation of malate dehydrogenase (MDH) that converts malate into oxaloacetate or the down-regulation of ICDHy that diverts flux away from the glyoxylate shunt. Interestingly, none of the transport reaction regulations identified in the MUST^U and MUST^X sets are present in the FORCE sets. The optimization formulation for the FORCE set identified more economical upstream flux modifications that negated the formation of multiple by-products. A consequence of imposing 100% yield to succinate is that biomass formation is halted as histidine and methionine formation is seized.
3.4.1. Succinate overproduction in E. coli after the addition of heterologous pyruvate carboxylase (PYC) reaction

Pyruvate carboxylase (PYC) has been overexpressed in E. coli from Lactococcus lactis (Lin et al., 2004; Sanchez et al., 2005b) and Rhizobium etli (Gokarn et al., 2000). The addition of the new reaction to the metabolic network boosts the succinate yield by 15.3% above the original theoretical maximum (1.72 moles/mole of glucose). PYC using ATP directly converts pyruvate into oxaloacetate which serves as a precursor for the glyoxylate and the fermentative pathway. In this study, we allowed the production of biomass at 1% of theoretical yield and identified the flux changes when succinate was produced at 98% of theoretical maximum (1.7 moles / mole of glucose).

The FORCE set of engineering interventions for this scenario is contrasted against the earlier case study and is shown in Figure 3.9b. The addition of the PYC reaction significantly reduces the number of engineering interventions required to guarantee the target yield for succinate. The interventions required to reduce the drain of carbon away from the pyruvate metabolism are absent indicating that the pyruvate carboxylase enzyme can safeguard against the consumption of pyruvate towards side-products. However, the down regulation for ASPTA is again needed to reduce the secretion of aspartate and glutamate. Importantly, the up-regulation for PYC could be substituted by up-regulating PPC which suggest that the OAA pool can be replenished by either of these two reactions. The increase in activity for some reactions in the glycolytic pathways (TPI, PGK) and the TCA cycle (ACONTa, ACONTb and MDH) is required as before. In contrast with the previous case-study, the complete elimination of PFL and isocitrate
dehydrogenase (ICDHy), rather than just their down-regulation is needed. The elimination of PFL is imposed to completely prevent the conversion of pyruvate into by-products. The elimination of ICDHy blocks the flow of carbon flux through the TCA cycle into the glutamate pathway thus ensuring the complete conversion of isocitrate into glyoxylate and succinate.

3.5. Discussion and Conclusions

3.5.1. Overproduction of succinate

In this chapter, an optimization-based methodology called OptForce was introduced for predicting all possible metabolic modifications that could guarantee, subject to the model stoichiometry and conditions, a pre-specified overproduction level of a desired biochemical. The results for succinate overproduction in *E. coli* reveal that the needed interventions results remain the same upon requiring the production of a small amount of biomass but change significantly upon the addition of a key reaction to the model.

Many of the suggested interventions recapitulate existing strain redesign strategies for succinate synthesis. For example, experimental evidence suggests that the overexpression of PPC from *Sorghum vulgare* and *Actinobacillus succinogenes* in *E. coli* not only increases the yield of succinate but also reduces the secretion of acetate (Cox et al., 2006; Farmer and Liao 1997; Kim et al., 2004; Lin et al., 2004; Millard et al., 1996; Sanchez et al., 2005b). In addition, succinate production has been enhanced by the increased carboxylation of PEP and pyruvate (to increase the pool of OAA for TCA cycle) in the *E.
coli mutant NZN111 by decreasing the activity for pyruvate formate lyase (PFL) and lactate dehydrogenase (Hong and Lee 2001; Stols and Donnelly 1997). Furthermore, Vemuri et al. (Vemuri et al., 2002a; Vemuri et al., 2002b) made use of the glyoxylate pathway for succinate synthesis thus overcoming the limitation of NADH availability for the fermentation pathway. The up-regulations for the isozymes ACONTa/b and the down regulations for ICDHy, ASPTA and GLUDy predicted by OptForce allude to the same strategy of glyoxylate shunt utilization for succinate synthesis. Finally, multiple studies (Cox et al., 2006; Sanchez et al., 2005a; Sanchez et al., 2005b; Sanchez et al., 2006) have shown that the deletion of adhE and ackA-pta coding for acetaldehyde dehydrogenase (ACALD) reduces the formation of by-products ethanol, acetate and acetaldehyde as suggested by OptForce.

The up-regulation of citrate synthase (CS), aconitase (ACONTa/b) and reactions from the glycolytic pathway (PGK and TPI) are engineering strategies suggested by OptForce that to the best of our knowledge have not yet been implemented for succinate production. Heterologous overexpression of the citZ gene from Bacillus subtilis that encodes citrate synthase increased the activity through the TCA cycle towards isocitrate and 2-ketoglutаратate (Underwood et al., 2002). However, when this gene was overexpressed in E. coli strain SBS550MG, an increase in the yield of succinate was not observed (Sanchez et al., 2005b). The reason for this could be the absence of the down regulations for ICDHy and GLUDy that lead to the production of glutamate and other amino acids required for growth. The results predicted by OptForce suggest that by collectively incorporating the flux modulations for citrate synthase, isocitrate dehydrogenase and
glutamate dehydrogenase along with the existing strategies, the yield of succinate can be further enhanced from the current experimental yield (1.7 moles/mole of glucose) as observed for strains SBS550MG and SBS990MG (Sanchez et al., 2005b).

### 3.5.2. Overproduction of 1-butanol

The results suggested by our OptForce procedure (Ranganathan et al., 2010b) revealed the differing nature of metabolic interventions required to overproduce 1-butanol using the thiobutanoate and ketoacid pathway. Recruiting the thiobutanoate pathway for 1-butanol overproduction required up-regulations for glycolytic fluxes (PGI, PGM). On the other hand, the ketoacid precursors were made available to 1-butanol synthesis by knocking down competing pathways (PFL, ACONTa/b etc.). The flux changes observed in the MUST sets for the two cases also showcased contrasting patterns. For example, for the thiobutanoate pathway, the fluxes of the pentose phosphate pathway increased so that alternative routes for glutamate and other amino acids are maintained to support growth. Although, none of the reactions from pentose phosphate pathway appeared in the FORCE sets, on the contrary, the OptForce procedure indicated that the fluxes of phosphogluconate dehydrogenase (GND) must be down-regulated while using ketoacid pathway to synthesize 1-butanol.

Several interventions that were identified in the FORCE sets have been used in existing strains to produce 1-butanol. For example, recent strategies to delete host competing pathways encoded by the genes \textit{ldhA}, \textit{frdBC}, \textit{pta}, \textit{pfl} and \textit{adhE} (Atsumi et al., 2008a; Atsumi et al., 2008b; Atsumi and Liao 2008b) have resulted in a three-fold increase in the
yield of 1-butanol. In addition, enhancing glycolytic fluxes by overexpressing NADH-
regenerating enzymes were implemented in an E. coli strain (Nielsen et al., 2009) that
yielded 580 mg/L of 1-butanol. In addition to the existing interventions, the OptForce
procedure also uncovered new knockouts and up-regulations that coordinate an increased
synthesis of 1-butanol. For example, the up-regulation of glycolytic fluxes and
phoephoenolpyruvate carboxylase (PPC) increase the amount of oxaloacetate for the
TCA cycle. However, in order to effectively utilize the transamination pathway, OptForce suggested up-regulations for PPCSCT and SUCOAS that are in close proximity
to the branching thiobutanoate pathway.

3.5.2. Remarks about OptForce framework
The genetic interventions predicted by OptForce underscore the importance of up-
regulating key fluxes along the succinate pathway in addition to the knockouts for by-
products. Existing strain optimization procedures (e.g. OptKnock (Burgard et al., 2003)
and OptReg (Pharkya and Maranas 2006)) that couple the maximization of growth rate
and secretion of the product tend to prevent the yield of succinate from reaching the
theoretical maximum. Table 3.1 contrasts the yields predicted for succinate
overproduction by OptKnock (Burgard et al., 2003), OptReg (Pharkya and Maranas
2006) and OptForce. OptKnock and OptReg rely on biomass maximization to perform
flux allocation in the metabolic network whereas OptForce reports the most conservative
value for succinate production allowed by the stoichiometry and conditions. It is
noteworthy that for more than two interventions even the worst-case succinate yield
predictions by OptForce are far more superior to strategies predicted by OptKnock and
OptReg. Notably, OptForce suggested the down regulation but not the knockout of PFL and GLUDy (Sanchez et al., 2005b) along with a number of additional interventions missed by both OptKnock and OptReg due to their inconsistency with biomass maximization.

The OptForce procedure allows for the complete enumeration of engineering modifications consistent with an overproduction target(s). The incorporation of metabolic flux information about the wild-type network allows for a sharper elucidation of engineering interventions. The engineering interventions predicted by OptForce depend on the available flux measurements for the initial strain. OptForce can be modified to predict globally valid metabolic interventions by utilizing biological objectives (i.e. maximization of biomass) when sufficient metabolic flux data are not available. Furthermore, the procedure can hierarchically be applied at intermediate stages of a metabolic engineering project by re-calculating the set of engineering interventions as new flux data for (multiple) mutant strains become available. The restriction of minimality in the calculated FORCE set can be relaxed allowing for the exploration of less parsimonious engineering interventions. For example, we studied the case for identifying additional interventions after retaining the best eight out of the ten interventions originally identified by the OptForce method (for cases 1 and 2). However, we found that even after allowing seven additional interventions (i.e. K = 15), the resulting FORCE set was not sufficient to increase the yield to more than 80% of the theoretical maximum. In addition, reactions that cannot (e.g., diffusion limited transport, non-gene associated reactions, etc.) be directly manipulated can be excluded from
consideration during the derivation of the FORCE set. It is to be noted that the OptForce procedure provides targets for genetic manipulations at the metabolic flux level. The lack of a completely quantitative mapping between gene expression and flux levels implies that multiple rounds of experimental strain modifications may be needed to translate the FORCE set of reaction fluxes to the required gene expression levels.
**Table 3.1:** Comparison of the minimum guaranteed fluxes from OptKnock, OptReg and OptForce procedures for succinate production in *E. coli.* (*) The values within parentheses denote the maximum flux values for succinate from OptKnock and OptReg.

<table>
<thead>
<tr>
<th>Number of metabolic interventions (K)</th>
<th>Knockouts</th>
<th>Results from OptKnock</th>
<th>Results from OptReg</th>
<th>Results from OptForce</th>
<th>Minimum guaranteed flux for succinate (*) (mmol / gDW.hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>K = 2</strong></td>
<td>ALCD2x, GLUDy</td>
<td>5.5 (84.1)</td>
<td>PFL (×), PPC (†)</td>
<td>2.1 (79.4)</td>
<td>PPC (†), CS (†)</td>
</tr>
<tr>
<td></td>
<td>PFL, LDH</td>
<td>1.2 (76.8)</td>
<td>-</td>
<td>-</td>
<td>PPC (†), MDH (↓)</td>
</tr>
<tr>
<td><strong>K = 3</strong></td>
<td>ALCD2x, PFL, LDH</td>
<td>5.9 (85.7)</td>
<td>PFL (×), PPC (†), ALCD2x (↓)</td>
<td>2.8 (84.3)</td>
<td>PPC (†), CS (↑)</td>
</tr>
<tr>
<td></td>
<td>ALCD2x, ACKr, PTAr</td>
<td>1.1 (84.6)</td>
<td>-</td>
<td>-</td>
<td>PPC (↑), ACONT (↓)</td>
</tr>
<tr>
<td><strong>K = 4</strong></td>
<td>ALCD2x, ACKr, PTAr, PYK</td>
<td>4.9 (88.8)</td>
<td>PPC (↑), PDH (↓), ALCD2x (↑), CS(↑)</td>
<td>2.8 (88.4)</td>
<td>PPC (↑), CS (↑)</td>
</tr>
<tr>
<td></td>
<td>ALCD2x, ACKr, PTAr, TKT1</td>
<td>2.1 (87.4)</td>
<td>-</td>
<td>-</td>
<td>PPC (↑), ACONT (↑)</td>
</tr>
</tbody>
</table>
Figure 3.1: Identification of fluxes that MUST change in the metabolic network for overproduction
Figure 3.2: An outline of the OptForce procedure to identify the minimal set of genetic interventions that guarantee the imposed yield. The representation on left shows the bi-level optimization algorithm. The graph on the right outlines the plot between the yield of the product and the number of direct manipulations identified by OptForce.
**Figure 3.3:** \( \text{MUST}^U \) and \( \text{MUST}^L \) set of reactions for 1-butanol synthesis in *E. coli* using the thiobutanoate pathway.
Figure 3.4: MUST\textsuperscript{UU}, MUST\textsuperscript{UL} and MUST\textsuperscript{LL} set of reactions for overproducing 1-butanol in *E. coli* using the thiobutanoate pathway.
Figure 3.5: Reactions that needs to be actively manipulated (FORCE sets) for overproducing 1-butanol. The figure on the left shows the metabolic map of *E. coli* recruiting the thiobutanoate pathway while the one on the left is the metabolic map of *E. coli* with heterologous pathways using ketoacid intermediates. Above each of these figures are the set of reactions that needs to be manipulated.
**Figure 3.6:** MUST\textsuperscript{U} and MUST\textsuperscript{L} set of reactions identified by OptForce for overproducing succinate in *E. coli.*
Figure 3.7: MUST set of reactions considered two-at-a-time (pairs) for overproducing succinate in *E. coli*. 
Figure 3.8: MUST$^{UUU}$, MUST$^{UUL}$, MUST$^{ULL}$, and MUST$^{LLL}$ set of reactions for overproducing succinate in E. coli.
Figure 3.9: FORCE set of reactions for succinate overproduction in *E. coli* without (figure A on the left) and after (figure B on the right) addition of heterologous pyruvate carboxylase reaction.
Chapter 4

Validation of OptForce: Improving the yield of malonyl-CoA and overproducing flavanones in *Escherichia coli*‡

4.1. Background

The biosynthesis of plant-specific secondary metabolites such as aromatic polyketides (Askenazi et al., 2003; Zhang et al., 2008), flavanones (Chemler and Koffas 2008; Winkel-Shirley 2001) and fatty acids (Cahoon et al., 2007; Schwender 2008) has recently become the focus of extensive research efforts due to their pharmaceutical potential in chronic diseases such as diabetes, cancer, obesity and Parkinson’s disorder. As a result, tremendous effort has been dedicated to the development of cost-efficient processes for the synthesis of these compounds which are still manufactured by extraction from their native plant sources. The major bottleneck in such extraction processes is their relatively low abundance and the complicated downstream purification processes (Cragg et al., 1997). Though simple starting materials can be used to chemically synthesize these compounds, such chemical processes involve toxic intermediates and extreme reaction

‡ This chapter has been published as: Xu et al. (2011) “Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA”. *Metab Eng* (doi:10.1016/j.ymben.2011.06.008)
conditions that significantly limit their scalability (Chemler and Koffas 2008; Harborne and Williams 1975). Recently however, industrial microbial hosts (e.g. Escherichia coli and yeast) have been engineered with multi-gene heterologous pathways to combinatorially produce these chemical compounds from renewable resources. In general, the biochemical routes that lead to the synthesis of these compounds involve acyl-CoA monomers (e.g. acetyl-CoA, malonyl-CoA, propanoyl-CoA) as precursors. For example, polyketides are formed in a multi-step decarboxylative condensation by the enzyme polyketide synthase (PKS) which recruits one molecule of malonyl-CoA in each step (Khosla and Keasling 2003; Pfeifer and Khosla 2001). On the other hand, the pathway responsible for the synthesis of flavanones involves the enzymes 4-coumaroyl-CoA ligase (4CL), chalcone synthase (CHS) and chalcone isomerase (CHI) and requires three molecules of malonyl-CoA (Leonard et al., 2008; Yan et al., 2005). Malonyl-CoA also serves as a starting point for the synthesis of microdiesel (Kalscheuer et al., 2006) and a host of various other value-added chemical compounds such as polyunsaturated omega-3 fatty acids (Qi et al., 2004). Therefore, the transfer of malonyl-CoA-dependent metabolic pathways into heterologous hosts for various applications depends on resourceful utilization of malonyl-CoA that is derived from primary metabolism.

In both E. coli and yeast the intracellular availability of malonyl-CoA is limited (Takamura and Nomura 1988) due to its direct association with cell growth and synthesis of phospholipids and fatty acids (Magnuson et al., 1993). With the introduction of new genes in the biosynthetic cascade, it is essential to engineer the bacteria so that a metabolic balance can be achieved between the requirement of malonyl-CoA for cell
growth and product synthesis. Hence, efficiently harnessing malonyl-CoA in E. coli continues to be one of the largest obstacles for producing natural products such as polyketides and flavanones. Metabolic engineering efforts aimed at improving intracellular malonyl-CoA based on metabolic network inspection have resulted in tuning pathways that are closer in proximity to malonyl-CoA synthesis. For example, overexpression of acetyl-CoA carboxylase (Leonard et al., 2007; Wattanachaisaereekul et al., 2008) that leads to the formation of malonyl-CoA and deleting enzymes involved in its depletion, such as acetate kinase and alcohol dehydrogenase (Zhang et al., 2009), have been reported. While genetic manipulations implemented in these studies have significantly improved the synthesis of malonyl-CoA, numerous metabolic engineering possibilities are yet to be explored.

In this chapter, we demonstrate the utility of OptForce (Ranganathan and Maranas 2010; Ranganathan et al., 2010a) in predicting genetic interventions resulting in significantly improved yields that are distal from the target pathways. OptForce procedure identifies the complete set of minimal genetic intervention strategies using a Boolean description consistent with an overproduction target. Here, we first characterize the wild-type strain of E. coli, BL21 Star™ based on available flux measurements and then identify computationally the interventions that cooperatively force carbon flux towards malonyl-CoA while at the same time preventing the drain towards byproducts. We combinatorially analyzed the impact of gene knockouts and overexpression strategies and demonstrate the sequential aggregation of beneficial interventions. The combined use of
computations and experiments to increase the availability of malonyl-CoA is demonstrated for the synthesis of the flavanone molecule naringenin in *E. coli*.

### 4.2. Materials and Methods

#### 4.2.1. OptForce Predictions

The essence of OptForce procedure is to compare and contrast the maximal range of flux variability for a wild-type strain (or a base strain) against the ones consistent with a prespecified overproduction objective. As outlined in the previous chapters, we have deployed OptForce here to identify the minimal set of genetic manipulations to improve the malonyl-CoA availability for flavanone synthesis. We have used wild-type flux measurements (Noronha et al., 2000; van de Walle and Shiloach 1998) for *E. coli* strain BL21 Star™. The latest genome-scale metabolic model for *E. coli* (Feist et al., 2007) was used to perform computations. The following experiments were performed by Prof. Mattheos Koffas group at the State University of New York, Buffalo. The gene knockouts and overexpressions were performed based on the results predicted by OptForce.

#### 4.2.2. Strains and Media

All recombinant plasmid construction was performed using *E. coli* XL1-Blue (Stratagene) as a host strain following standard cloning protocols and BL21 Star™ (DE3) (Invitrogen) was used for flavanone production. Plasmids pKD4, pKD46 and pCP20 for gene knockout construction were obtained from the *E. coli* Genetic Resource Center,
Yale University. The duet vectors pCoLA-Duet and pCDF-Duet (Novagen) were used for cloning and subcloning. Restriction enzymes and T4 DNA ligase were purchased from New England Labs. Genomic DNA was prepared using PureLink genomic DNA purification kit (Invitrogen), and colony PCR was performed using GoTaq HotStart DNA polymerase (Promega); all other PCRs were performed using Phusion High Fidelity Master Mix (Finnzymes). Plasmid DNA preparation and fragment DNA recovery were conducted using the Zyppy miniprep kit and Zymoclean gel recovery kit (Zymo Research), respectively. Luria-Bertani (LB) broth (Sigma) and M9 minimal salts (Difco) were routinely used for culture growth and flavanone production. Recombinant cultures were grown in media containing the appropriate antibiotics: ampicillin (70 µg/ml), streptomycin (40 µg/ml), kanamycin (40 µg/ml) and chloramphenicol (20 µg/ml). Culture glucose levels were measured using a Wako LabAssay™ Glucose kit (B-Bridge) purchased from Japan. All chemicals were purchased from Sigma unless otherwise specified.

4.2.3. Cloning and Pathway Construction

E. coli K-12 (MG1655) genomic DNA was used as a template for amplification of pgk, gapA, aceE, aceF and lpdA genes using the forward and reverse primers specified in Table S2 of Xu et. al. (2011). Subsequently, the pgk and gapA PCR products were double digested, gel purified and cloned into the BamHI/SaiI and BglII/XhoI restriction sites of pCoLA-Duet vector, respectively. The pgk fragment was then subcloned into the BamHI/SaiI sites of pCoLA-gapA or pCDF-4CL2, to form pCoLA-pgk-gapA or pCDF-pgk-4CL2 constructs. For overexpression of pyruvate dehydrogenase multienzyme
complex, three rounds of cloning were performed. First, the lpdA and aceF gene were inserted into the BamHI/EcoRI and BglII/XhoI restriction sites of pCoLA-Duet vector, forming plasmids pCoLA-lpdA and pCoLA-aceF, respectively. Then the T7-aceF fragment was amplified and cloned into the Sall/NotI restriction site of pCoLA-lpdA to form pCoLA-lpdA-T7aceF, using pCoLA-aceF as template. Finally, aceE fragment was cloned into the MfeI/XhoI restriction sites of pCoLA-lpdA-T7aceF to form recombinant plasmid pCoLA-PDH; each of the three subunits (lpdA, aceE and aceF) was under the control of a separate T7 promoter. All clones were screened by restriction digestion analysis and subsequently verified by gene sequencing.

Gene deletions were performed using the red recombinase based chromosomal gene inactivation protocol developed by Datsenko and Wanner (Datsenko and Wanner 2000). Deletion primers (Table S2) with 40 nt homologous regions were used to create the FRT-flanked kanamycin resistance cassette from pKD4, which was then transformed into the red-recombinase expressing BL21 Star\textsuperscript{TM} (DE3) strain by electroporation. Positive knockout strains were screened by colony PCR and verified by gDNA PCR. Finally, the resistance marker was eliminated by expressing the flippase recombination enzyme from pCP20.

4.2.4. Fermentation Procedure

Flavanone production was performed based on a two-step fermentation protocol. Strains were first cultivated in 40 ml LB broth at 37 °C with orbital shaking. Induction of heterologous pathway expression was performed during the mid-exponential phase of
cultivation (approximately OD 0.6-0.8) by addition of 1 mM IPTG and the cultures were left to grow at 30 °C for an additional 6-8 hours. After induction, the bacterial pellet was harvested by centrifugation and suspended with 16 ml M9 modified medium (1 × M9 salts, 10 g/L glucose, 1 mM MgSO₄, 0.1 mM CaCl₂, 6 μM biotin, 10 nM thiamine, 2.6 mM p-coumaric acid, 1 mM IPTG). 4 g/L glucose (0.4%) was supplemented once the glucose was completely exhausted (approximately at 12 hour and 20 hour of cultivation). Flavanones were extracted with 50% ethanol after 24 hours of fermentation, then the E. coli pellet was removed by ultracentrifugation and the supernatant was analyzed for flavanone quantification. Fermentation kinetics were investigated every 2 or 3 hours by measuring biomass concentration, glucose and p-coumaric acid consumption and flavanone production throughout the 40-hour fermentation process. Cell growth rate, glucose consumption rate and flavanone production rate were obtained by analyzing the steady state fermentation kinetic curves (usually achieved for the first 10 hours fermentation) using linear progression, with a correlation coefficient more than 0.95 used as the selection criteria.

4.2.5. Analytical procedures

Flavanones were analyzed by the Agilent 1100 HPLC system equipped with a ZORBAX SB-C18 column (5μm, 4.6 × 150 mm) kept at 25 °C and a diode array detector (DAD). The mobile phase contains 35% acetonitrile (with 0.1% formic acid) and 65% water (with 0.1% formic acid). The retention time for p-coumaric acid and naringenin were around 2.6 min and 6.9 min, respectively. Extraction and quantification of coenzyme A
compounds were performed according to the procedures as previously described (Fowler et al., 2009).

4.3. Results

4.3.1. Genetic interventions suggested by OptForce

In this section, we describe all possible genetic interventions identified by OptForce for improving the intracellular malonyl-CoA levels. Calculations were based on wild-type flux measurements (Noronha et al., 2000; van de Walle and Shiloach 1998) for *E. coli* strain BL21 Star™. A fixed glucose uptake flux at 100 µmol/gDW/hr and a yield of 70 µmol/gDW/hr for naringenin while allowing for 50% yield on biomass were imposed as constraints and targets, respectively. Figure 4.1 illustrates the results for the FORCE set of genetic interventions.

Based on the above, it is important to note that the naringenin pathway requires the synthesis of two important precursors natively produced in *E. coli* - 1) acetyl-CoA that is produced at the end of glycolysis and pyruvate metabolism, and 2) malonyl-CoA which is derived from the carboxylation of acetyl-CoA by acetyl-CoA carboxylase (ACCOAC). Acetyl-CoA serves as a key branching metabolite in redistributing resources to the citric acid cycle, fermentative byproducts and towards fatty acid and amino acid metabolism. On the other hand, malonyl-CoA participates in biomass formation directly as a reactant and through phospholipid biosynthesis and the synthesis of acyl carrier protein (ACP). However, the molar conversion ratio between glucose and malonyl-CoA is 1:1 under the
assumption that half of the acetyl-CoA flux goes into TCA cycle. The requirements for malonyl-CoA are constrained by the allowances for cell growth, which in turn restricts the availability of malonyl-CoA for naringenin. Furthermore, the availability of malonyl-CoA imposes a restriction on the consumption of \(p\)-coumaric acid.

OptForce suggested the up-regulation of glycolytic reactions, namely glyceraldehyde-3-phosphate dehydrogenase (GAPD) and phosphoglycerate kinase (PGK) that would result in an increased flux towards pyruvate produced through glycolysis. In addition, up-regulation of pyruvate dehydrogenase (PDH) and acetyl-CoA carboxylase (ACCOAC) increase the pool of precursors acetyl-CoA and malonyl-CoA. In contrast, down-regulation of reactions in the citric acid cycle, namely malate dehydrogenase (MDH), fumarase (FUM) and aconitase (ACONT\(\text{a/b}\)), were suggested by OptForce to reduce the drain of carbon towards TCA cycle products. In order to allow for biomass production yield at 50% of the theoretical maximum, OptForce suggested reducing the activity of TCA reactions instead of completely eliminating them. Interestingly, OptForce predicted knockouts for the reactions succinyl-CoA synthetase (SUCOAS) and propanoyl-CoA:succinyl-CoA carboxylase (PPCSCT) that consume coenzyme A as a cofactor leading to the formation of succinyl-CoA. It is to be noted that the oxidation of pyruvate to acetyl-CoA by pyruvate dehydrogenase using coenzyme A is one of the key precursor reactions in the biosynthesis of malonyl-CoA. While the up-regulation of pyruvate dehydrogenase and acetyl-CoA carboxylase increase the utilization of coenzyme A towards the formation of malonyl-CoA, the remaining amount of coenzyme A is used up in other parts of the metabolism. Notably, the wild-type flux values for reactions
propanoyl-CoA:succinyl-CoA carboxylase and succinyl-CoA synthetase indicate that a considerable amount of coenzyme A is utilized as a cofactor in the formation of succinyl-CoA and methylmalonyl-CoA. Hence, OptForce suggested completely eliminating these reactions in order to provide the maximum quantity of coenzyme A for the synthesis of malonyl-CoA.

4.3.2. Implementing OptForce interventions

The interventions suggested by OptForce are at the metabolic flux level. In order to implement these interventions at the genetic level, it is essential to identify the genes encoding each one of these reactions. Using the gene-protein-reaction (GPR) associations provided with the iAF1260 metabolic model, we identified all genes (isozymes and subunits) encoding these reactions (see Figure 4.1). We selected the genes that were not regulated under our experimental conditions. For instance, \textit{fumA}, \textit{fumB} and \textit{fumC} are three biochemically distinct fumarases (FUM) that catalyze the interconversion of fumarate to malate in the TCA cycle. Synthesis of these three isozymes in \textit{E. coli} is controlled in a hierarchical manner depending on oxygen level (Tseng et al., 2001). Under our experimental conditions (see section 4.2), RT-PCR revealed that there is no detectable \textit{fumA} expression, which is consistent with the findings that \textit{fumA} expression can only be activated under strict anaerobic conditions (Tseng et al., 2001). Similarly, the succinyl-coA synthetase of \textit{E. coli} is encoded by two polycistronic genes, \textit{sucC} (β subunit) and \textit{sucD} (α subunit), which are synthetic lethal pairs (Motter et al., 2008; Suthers et al., 2009). After thoroughly scrutinizing the gene regulation under the experimental conditions and GPR associations, we selected the targets for genetic
interventions for subsequent experiments. Figure 4.2 shows the effects of each genetic intervention predicted by OptForce for overproducing naringenin.

4.3.3. Overexpression of acetyl-CoA carboxylase and precursor pathways

In accordance with previous findings (Davis et al., 2000; Leonard et al., 2007), OptForce suggested the overexpression of acetyl-CoA carboxylase as the most important step towards improving intracellular malonyl-CoA. In addition, OptForce also suggested the up-regulation of phosphoglycerate kinase (PGK) or glyceraldehyde-3-phosphate dehydrogenase (GAPD), and pyruvate dehydrogenase (PDH). Up-regulation of glyceraldehyde-3-phosphate dehydrogenase (gapA) and phosphoglycerate kinase (pgk) can maximize the conversion of glucose to pyruvate while preventing the accumulation of glyceraldehyde-3-phosphate, which would otherwise be channeled into the pentose phosphate pathway (PP) or methylerythritol phosphate (MEP) pathway. Subsequently, up regulation of pyruvate dehydrogenase multienzyme complex (PDH) increases the amount of acetyl-CoA available for conversion into malonyl-CoA and at the same time reduces the formation of other by-products (e.g. lactate, ethanol and acetate). In the recombinant strain containing overexpressed accA, pgk or gapA and the three-subunit pyruvate dehydrogenase, we observed that the yield of malonyl-CoA increased by about 220% (see Figure 4.3) compared to the wild-type strain.
4.3.4. TCA Cycle Disruptions on Cell Growth and Flavanone Biosynthesis

Chromosomal inactivation of TCA cycle genes was carried out to identify flavanone-overproducing mutants with least effect on cell growth (Figure 2). Results shown here indicate that single knockouts of *sucC*, *fumB* or *fumC* increased the production of naringenin, a common flavanone, by 30% compared to the wild type strain, whereas knockouts of *mdh*, *acnA* or *acnB* decreased naringenin production by up to 30-50%. These results suggest that partially impairing the TCA activity by deletion of *fumB/C* or *sucC* can successfully redirect acetyl-CoA flux towards flavanone synthesis while maintaining an acceptable cell growth, which can serve as a guideline for the optimization of the biosynthesis of other acetyl-CoA-derived compounds.

4.3.5. Construction of recombinant strain by combining beneficial interventions

In this section, we analyzed the cumulative effect of multiple genetic manipulations (Alper et al., 2005a; Park et al., 2007) by combining the beneficial knock-out mutants together with overexpression targets. Although the use of OptForce procedure resulted in the prediction of several combinatorial recombinant strains, we investigated the production potential of the best strains out of the several alternative genetic interventions proposed. As shown in Figure 4.3, the two double mutants Δ*fumBΔsucC* and Δ*fumCΔsucC*, can be categorized as promising combinations for the improvement of flavanone synthesis, while other double mutants including model-identified combinations (i.e. Δ*mdhΔsucC*) are less promising candidates. Strains overexpressing acetyl-CoA carboxylase (ACC), phosphoglycerate kinase (PGK) and pyruvate dehydrogenase (PDH) in the Δ*fumCΔsucC* double mutant background led to the maximum final volumetric
production of 474 mg/L of naringenin, which is a 5.6-fold increase compared with the parental strain (85 mg/L). It was interesting to find that combinations of top single knockouts can result in suboptimal strains (i.e. ΔfumBΔfumC double mutant strain) that yielded as low as 2/3 of the flavanone production obtained from the best producer. Similarly, any additional modifications failed to further increase naringenin production compared with the fumC and sucC double mutant strain, probably due to the unbalanced metabolism between cell growth and product formation.

4.3.6. Enhanced Cellular Malonyl-CoA Level in Constructed Strains

Cells tend to maintain a relatively constant level of acetyl-CoA and malonyl-CoA, something that severely limits their flavanone production potential (Fowler et al., 2009; Pitera et al., 2007). In order to verify our hypothesis that the constructed recombinant and mutant strains produced flavanones more efficiently due to higher levels of intracellular acetyl-CoA and malonyl-CoA, we measured the intracellular concentrations of these two cofactors in the engineered strains. Strains carrying acetyl-CoA carboxylase (ACC), phosphoglycerate kinase (PGK) and pyruvate dehydrogenase (PDH) overexpression in the fumC and sucC double mutant background were found to contain the highest levels of both malonyl-CoA and acetyl-CoA, which is consistent with the naringenin production profile (Figure 4.3). Furthermore, the model-identified optimal strains (fumC and sucC double mutants) exhibit a relatively high level of acetyl-CoA (~11.6 µmol/gDW) and malonyl-CoA (~10.4 µmol/gDW), roughly a 4-fold increase over the wild type strain, indicating that the carbon flux was successfully rerouted to the acetyl-CoA and malonyl-CoA synthesis. Moreover, for the suboptimal strains (ΔacnB, Δmdh or ΔmdhΔsucC
mutants), the changing pattern of malonyl-CoA levels at different post-induction phase resembled the ones from the wild type strain, indicating that other possible limiting factors, such as lack of precursor availability, unbalanced NAD\(^+\)/NADH ratio or impaired glyoxylate shunt pathway, may cancel the beneficial effects of these genetic interventions. As mentioned before, reduced cell growth associated with these interventions is a genuine factor that may affect the malonyl-CoA level.

### 4.4. Discussion and Conclusions

In this work, we have successfully validated computational predictions from OptForce by designing a strain of *E. coli* with increased levels of intracellular malonyl-CoA. In essence, by implementing a minimal number of direct metabolic interventions (i.e. FORCE sets), we were able to alter the wild-type metabolic network of the microbe towards meeting the changes (encoded within the MUST sets) that conform to our overproduction objective. The set of alternative interventions predicted by OptForce enabled us to avoid genetic interventions that are more difficult to perform or less viable under the experimental conditions. By exhaustively exploring all alternative optimal solutions, FORCE sets can be represented in the form of decision-tree where some interventions can be substituted with others. For example, as shown in Figure 4.2, the knockouts for propanoyl-CoA:succinyl-CoA carboxylase can be substituted by the deletion of succinyl-CoA synthetase. Additionally, interventions predicted by OptForce are rank-ordered based on their quantitative impact towards satisfying the overproduction target. For example, when only three modifications were allowed, OptForce chose the
upregulation of acetyl-CoA carboxylase, pyruvate dehydrogenase and glycolytic reactions that collectively resulted in a yield of 33% of the target yield. As shown in Figure 5, the absolute yields for naringenin increase by about 100% from the wild-type by just including up-regulations for the genes accA and gapA or pgk. For five interventions \((k = 5)\), the original upregulations were conserved and in addition OptForce predicted knockouts for succinyl-CoA synthetase and down regulations in the TCA cycle that enabled meeting the imposed target. In contrast to the OptForce-derived predictions, knockouts for the genes mdh and acnA severely compromised the formation of biomass and naringenin. It might be noted that OptForce suggested down-regulating these enzymes as opposed to completely eliminating the activity of these reactions. However, as we stated earlier, due to the unstable nature of the RNA duplexes (i.e. terminal unpaired region, loop or hairpin degree), the efficacy in down-regulating genes becomes a difficult task. In addition, products of these reactions are closely associated with the formation of biomass. For this reason, we decided to investigate alternative gene knockouts (i.e. fumB and fumC) proposed by OptForce. Furthermore, knockouts for fumB, fumC and sucC in combination with overexpressing three subunits for pyruvate dehydrogenase further boosted the yield of naringenin to 219 mg / g of glucose.
Figure 4.1: Metabolic interventions predicted using OptForce procedure for overproducing naringenin in *E. coli*. In the metabolic map, reactions from the flavanone pathway are shown in light green and *E. coli* reactions are shown in light blue. Up-regulations are shown in dark green, down-regulations are shown in red and knockouts are indicated by \( \times \) symbol in red.
**Figure 4.2:** Plot between naringenin yield and the number of genetic interventions ($k$) that were predicted by OptForce procedure. Overexpressions are shown in green and knockouts are shown in red.
Figure 4.3: Maximum detected levels of flavanone naringenin produced from *E. coli* BL21 Star™ (DE3) mutant strains expressing combinations of precursor pathway gene in batch fermentations. Plus or minus signs indicate presence or absence of the corresponding genetic intervention. Each column represents a specific genotype engineered for naringenin production. The first column in the absence of any genetic modifications corresponds to the parental wild type strain. Error bars represent standard deviation of duplicate data points.
Chapter 5

Validation of OptForce: Comprehensive identification of genetic interventions to understand the fatty acid biosynthesis

5.1. Background and Introduction

Fatty acids of chain length C₆ to C₂₄ are potential precursors to high-energy-density biofuels, including alkanes and alkyl esters (Lennen et al., 2010). So far, metabolic engineering efforts aimed to produce free fatty acids from microbial fermentation have resulted in yields significantly lesser than their theoretical maximums (Steen et al., 2010). Under normal conditions, the fatty acid metabolism is highly regulated because they are important constituents of membranes and act as a source of metabolic energy in most organisms (Fujita et al., 2007). Thus, the biosynthetic gene cascades that control its synthesis are switched off intermittently in accordance with the availability of fatty acids for membrane synthesis. With significant difficulties in controlling the fatty acid pathways, only a few organisms whose genetic map is well-known are used for fatty acid biosynthesis. Specifically, fatty acid regulation and metabolism of Escherichia coli is

---

§ Contents of this chapter are in preparation for publication with collaborators from Center for BioRenewable Chemicals (CBIrc), Iowa State University.
well understood owing to the availability of elaborate metabolic models, flux measurements and other experimental data. In this chapter, we describe preliminary results from a looped metabolic engineering project aimed at improving fatty acids of different chain lengths.

Figure 5.1 shows various branches of fatty acid metabolism in *E. coli*. There are two important fatty acid pathways in *E. coli* metabolism (Fujita et al., 2007; Marrakchi et al., 2002; Schweizer and Hofmann 2004). The type II or dissociated fatty acid biosynthesis (FAB) pathway involves the ATP-dependent acetyl-CoA carboxylase (encoded by *accABC*) as the first step. Acetyl-CoA is converted into malonyl-CoA which is further converted into malonyl-ACP by the enzyme malonyl-CoA:ACP transacylase encoded by the gene *fabD*. First, the fatty acid biosynthesis starts with the C₄ chain where one mole of acetyl-CoA and one mole of malonyl-ACP synthesize a 4-carbon fatty acid acyl carrier protein (i.e. butyryl ACP). Butyryl ACP further elongates into the 6-carbon chain by recruiting one mole of malonyl-ACP to produce hexanoyl-ACP. This process of chain elongation progresses with the availability of malonyl-ACP to form even-numbered fatty acid acyl carrier proteins. Fatty acid ACPs can be converted into fatty acids by thioesterases in a one-step enzymatic conversion process. An alternative biosynthesis route for the production of fatty acids that has gained importance recently (Dellomonaco et al., 2011) is the beta-oxidation pathway. While the natural use of this pathway is to degrade longer chain fatty acid ACPs into smaller coenzyme-A derivatives, the easy reversibility of this pathway has made it a competitive route to fatty acid synthesis. Although this pathway has been proved to be successful in producing C₄ fatty acid and 1-
butanol, the reversibility of pathways that lead to longer chain fatty acids need to be explored for feasibility. Moreover, the maximum theoretical yields from both the pathways are somewhat similar even though the regular fatty acid biosynthesis is dependent on the availability of ATP and the specificity of the termination enzyme (i.e. thioesterase).

Here, we use OptForce for identifying genetic interventions in *E. coli* that result in increased production of free fatty acids using the type II FAB pathways. Importantly, in this work, we use experimentally measured flux values for about ~60 reactions which offers a better characterization for the wild-type strain (MG1655) of *E. coli*. Subsequently, we exhaustively determine computational predictions from OptForce and test them by performing the most important interventions first. In order to initiate a “looped” metabolic engineering approach, we test the genetically modified strain further by estimating the new flux pattern. In subsequent steps, we expect to use the flux data from the improved strain as a feedback control and deploy OptForce further to pinpoint additional interventions that can guarantee the target yield.

### 5.2. Flux changes that MUST happen in the metabolism

In this study, we used the latest genome-scale metabolic model iAF1260 for *E. coli* (Feist et al., 2007). Using metabolic flux analysis (MFA) data for approximately ~60 reactions, we characterized flux distribution in the wild-type strain. Metabolic flux data (yet to be published) was provided by Dr. Jacqueline V. Shanks group from Center for
BioRenewable Chemicals (CBiRC), Iowa State University, USA. Fatty acids with varying carbon chain lengths \((C_6 - C_{16})\) were targeted for overproduction. Target yields for each of the fatty acids were set to 90% of its theoretical maximum. In addition, a biomass production requirement of 10% from the theoretical maximum was imposed for all cases.

Figure 5.2 summarizes the set of reactions in the network that MUST change in the face of overproduction. Results are organized so as the flux changes common to all fatty acids are shown within the smaller ellipses and additional length-specific modifications in ascending (left) and descending (right) chain length order are shown within the overlapping ellipses. For example, increases in the flux for 3-oxy-acyl-ACP synthase (3OAS40/60/80), enoyl-ACP reductase (EAR40/60/80), 3-oxo-acyl-ACP reductase (3OAR40/60/80) and 3-hydroxy-acyl dehydratase (3HAD40/60/80) are required for all fatty acids as shown within the \(C_6\) ellipse on the left. Similarly, knockouts for isocitrate lyase (ICL) and down-regulations for pentose phosphate pathways, 2-dehydro-3-deoxy-phosphogluconate aldolase (EDA) and 6-phosphogluconate dehydratase (EDD), are required for all fatty acids as shown in the \(C_{16}\) ellipse on the right.

In addition to these universal changes, a number of network modifications need to take place for a given chain length or higher (see left panel) or a given chain length or lower (see right panel). For example, up-regulating the central metabolic reactions glyceraldehyde-3-phosphate dehydrogenase (GAPD), enolase (ENO) and knocking out phosphogluconate dehydrogenase (GND) are required only for fatty acids \(C_8\) or longer.
(see Figure 5.2a). On the other hand, knockouts for reactions along the C\textsubscript{10} fatty acid pathway for acid are needed for overproducing hexanoic acid but not for other fatty acid targets. Interestingly, 3-oxy-acyl-ACP synthase (3OAS), enoyl-ACP reductase (EAR), 3-oxo-acyl-ACP reductase (3OAR) and 3-hydroxy-acyl dehydratase (3HAD) along the C\textsubscript{10}-C\textsubscript{14} fatty acid pathway were identified as necessary knockouts to overproduce hexanoate, octanoate and pentanoate. However, the same reactions need to increase in flux value for the synthesis of C\textsubscript{12} and longer chain fatty acids indicating that the modifications need to be dependent on the chain length.

Reactions along the beta-oxidation pathway for the C\textsubscript{8} fatty acid were classified in the MUST\textsuperscript{U} set of reactions (see Figure 5.2b) only for hexanoic acid. This is because of the lack of fattyacid-acyl-ACP hydrolase (FA60ACPHi) that can hexanoyl-ACP into hexanoic acid. Instead, the favored pathway involves the next step of elongation resulting in the formation of octanoyl-CoA which is further reduced into hexanoyl-CoA along the reversed beta-oxidation pathway. Eventually, hexanoyl-CoA is esterified into hexanoic acid by thioesterase. However for the higher chain fatty acids (C\textsubscript{8} or higher), the beta-oxidation pathway is seldom used. This is because of two reasons. Firstly, for higher fatty acids, the hydrolase reaction exists in the model that can convert the acid-ACP into acid. Secondly, the beta-oxidation pathways for these fatty acids are irreversible unlike the ones for the C\textsubscript{6} pathway. The increase in flux value for mal-CoA:ACP transacylase (MCOATA) appears in the MUST sets for all the fatty acids. This suggests up-regulating this reaction will ensure the availability of mal-ACP available for the chain elongation in each step. Other flux changes suggest strategies that would maximize carbon redirection
towards the fatty acids. Knockouts for alcohol dehydrogenase (ALD2x) and acetaldehyde dehydrogenase (ACALD) and reduction in flux value of pyruvate formate kinase (PFK) prevent the formation of fermentation by-products. We also notice reduction in flux value for reactions in the TCA cycle (citrate synthase (CS), aconitase (ACONTa), 2-ketoglutarate dehydrogenase (AKGDH), isocitrate dehydrogenase (ICDHy), malate dehydrogenase (MDH) and fumarase (FUM)). In addition, increase in fluxes for the glycolytic reactions and knockouts for pathways branching out from glycolysis ensure maximum carbon flow towards pyruvate and acetyl-CoA.

5.3. Genetic interventions for overproducing hexanoate

From the set of flux changes that MUST happen in the network, we then use OptForce to pinpoint minimal set of direct interventions that guarantee the overproduction of fatty acids. First, we explore the strategies for hexanoic acid (Figure 5.3). Not surprisingly, OptForce suggested knockouts for acetate kinase (ACKr) and phosphotransacetylase (PTAr) that negate the formation of by-products, acetate and ethanol. Up regulation of pyruvate dehydrogenase (PDH) ensures the availability of acetyl-CoA to the fatty acid. Instead of up-regulating the multi-subunit gene system for pyruvate dehydrogenase, OptForce predicted the knockout of alcohol dehydrogenase (ALCD2x) that eliminates consumption of acetyl-CoA into the formation of ethanol. To ensure the availability of pyruvate at the end of glycolysis, OptForce predicted down regulating phosphoenolpyruvate carboxylase (PPC). As an alternative to this strategy, OptForce
predicted down regulating fumarase (FUM) would not only result in increasing phosphoenol pyruvate but also reduce formation of TCA cycle by-products.

In the fatty acid metabolism, OptForce predicts the up regulation of malonyl-CoA:ACP transacylase (MCOATA) that ensures the availability of the malonyl-ACP for the chain elongation step. In addition, OptForce predicted up-regulating 3-oxy-acyl-ACP synthase (3OAS), enoyl-ACP reductase (EAR), 3-oxo-acyl-ACP reductase (3OAR) and 3-hydroxy-acyl dehydratase (3HAD) for the C₄, C₆ and C₈ chain. In addition, up-regulating reactions along the beta-oxidation pathway – octanoyl-CoA dehydrogenase (ACOAD3) and 3-oxooctanoyl-CoA dehydrogenase (HACD3) – would result in the degradation of octanoyl-ACP into hexanoic acid.

5.4. Genetic interventions for overproducing octanoate and palmitate

Metabolic interventions predicted by OptForce for octanoic acid (C₈) allude to restricted usage of carbon in the central metabolism (see Figure 5.4). Not surprisingly, in the fatty acid metabolism, OptForce suggests up-regulating reactions specifically along the C₈ chain (3OAS80, 3OAR80, 3HAD80 and EAR80). In order to prevent the consumption of precursors of octanoate further into the fatty acid pathway, OptForce predicts knockouts for reactions in the C₁₄ or C₁₆ chain (3HAD140 or 3HAD160). Notably, with the increase in chain length, OptForce pinpoints up-regulations for glycolytic reactions and eliminating branching pathways from glycolysis. For instance, up-regulations for enolase (ENO) and fructose bisphosphatase (FBA) appear for the C₈ fatty acid. In addition,
OptForce suggests knocking out phosphogluconate dehydrogenase (GND) and 6-phosphogluconate dehydratase (EDD) that depreciate carbon flux from glycolysis. Down-regulations for TCA cycle reactions – citrate synthase (CS) and aconitase (ACONTa/b) – were also predicted by OptForce to reduce the formation of biomass and other TCA cycle by-products. The knockouts predicted to negate the formation of acetate and ethanol (ACKr and PTAr) for overproducing the C$_6$ fatty acid is conserved for the C$_8$ fatty acid. Note that all these interventions considerably increase the acetyl-CoA available in the metabolism. While acetyl-CoA is an important precursor for the fatty acid metabolism, one mole of malonyl-ACP (transaminated in one-step from malonyl-CoA) is incorporated in each of the chain elongation step. Up-regulating acetyl-CoA carboxylase (ACCOAC) may seem to be a straightforward strategy to increase mal-ACP availability. However, in the absence of a metabolic setup that consumes excess mal-CoA into the fatty acid metabolism, most of it may be consumed directly for the formation of biomass. In order to prevent this, OptForce predicts up-regulating ketoacyl-ACP synthase (KAS15) that uses both acetyl-CoA and mal-ACP to produce direct precursors for all the fatty acids. Additionally, Figure 5.5 demonstrates the use of OptForce towards predicting the most important genetic interventions. For instance, in Figure 5.5, the up-regulation of one of the four reactions in the C$_8$ chain can individually account for an increase in the yield to about 83% of theoretical maximum. On the other hand, subsequent deletions and knock-downs result in improving the yield to about 89% of theoretical maximum.

The genetic manipulations required in *E. coli* for the overproduction of palmitate (C$_{16}$ fatty acid) and their corresponding impacts on the yield are shown in Figure 5.6. Notably,
OptForce suggests that as we move towards longer chain fatty acids, the number of genetic interventions required in the central metabolism become higher in number. This implies the need to strictly regulate the carbon flow from glycolysis to the fatty acid metabolism.

5.5. Discussion and Conclusions

In this chapter, we described computationally derived predictions for overproducing fatty acids in \textit{E. coli}. These targets have greatly expanded in the fuel candidate repertoire. Of particular interest is the microbial production of alkanes and alkenes, whose biosynthetic pathways have been demonstrated in a recent study (Schirmer et al., 2010). Typically, when building a pathway in a highly engineerable host (i.e. \textit{E. coli} and yeast) to generate a target compound of interest, it is advantageous to consider the use of native high flux pathways to provide key precursors. Furthermore, limiting the use of precursors by competing pathways and creating irreversible steps in the engineered pathway can increase flux toward product. In the previous sections of this chapter, OptForce rightly predicts this aspect. In spite of up-regulating direct biosynthesis pathways, we note that the minimum guaranteed yields do not reach the desirable target values. However, OptForce suggests that by tuning the central metabolic genes, we can adjust the metabolism such that the important precursors are provided at adequate quantities.

Contrary to the existing strategies (Davis et al., 2000; James and Cronan 2004; Kutchma et al., 1999; Marrakchi et al., 2002; Subrahmanyam and Cronan 1998) that rely on
augmenting acetyl-CoA and malonyl-CoA as the most important step, OptForce does not suggest the overexpression of acetyl-CoA carboxylase (accABC). Note that malonyl-CoA is a direct participant in the biomass formation. In Chapter 4, we observed experimentally that an increase in the intracellular levels of malonyl-CoA has proved to significantly increase cell growth. The increased availability of malonyl-CoA is used up towards the formation of biomass as the metabolism does not constrain its usage towards fatty acid pathways. On the other hand, the overexpressed fatty acid enzymes pull carbon flux through malonyl-CoA even though it is important for the formation of biomass by forcing a reduction in cell growth.

Interestingly, OptForce suggested the up-regulation of one the three reactions in the reversed beta-oxidation pathway for overproducing hexanoate. In the E. coli metabolic model, acyl-ACP thioesterase enzyme that catalyzes the conversion of hexanoyl-ACP into hexanoic acid is absent. Hence, the desired route to produce hexanoic acid is by elongating the chain further into the C₈ chain and degrading octanoyl-ACP through the reversed beta-oxidation cycle. The suggested up-regulations for hexanoic acid closely adhere to a recent study (Dellomonaco et al., 2011) to overproduce butanoic acid.

In order to scrutinize the chain-dependency of OptForce interventions for various fatty acids, we present a Venn diagram as shown in Figure 5.7. Only the knockouts for acetate kinase (ACKr) and phosphate acetyl transferase (PTAr) are common to all fatty acids suggesting it is imperative to negate the formation of acetate, ethanol and other fermentative by-products that branch out from acetyl-CoA. Not surprisingly, we observe
that the up-regulation of fatty acid reaction are exclusive in nature. For example, for overproducing C₆ fatty acid the up-regulation in fatty aid biosynthesis are only found in the C₆ chain and so on. As we move along in a clock-wise direction, genetic manipulations that reduce the activity of TCA cycle and increase the carbon flow towards the end of glycolysis start to emerge. This suggests that upon targeting a fatty acid of higher chain length, a strict redirection of central metabolic carbon flow towards the precursors become progressively important.

In view of establishing a metabolic engineering loop that includes model suggestions, recombinant strain engineering and metabolic flux mapping, we have designed a set of experiments that work towards smoothly transferring information from one stage to the other. We expect to complete this metabolic engineering loop by re-engineering an already improved strain for higher yields.
**Figure 5.1:** Fatty acid biosynthesis pathways in *E. coli*
Figure 5.2: Venn diagram representing fluxes that MUST increase (shown in green boxes), decrease (shown in red boxes) or be knocked-out (shown in red text) to meet the imposed overproduction yields. The Venn diagram on the left depicts the shared network modification requirements within the C6 ellipse and additional chain length-specific required changes moving from C₈ to C₁₆ fatty acids. Conversely, the Venn diagram on the right shows shared network modifications for all chains lengths within the ellipse for palmitate (C₁₆) and additional chain-specific network changes moving from C₁₄ to C₆ acids, respectively.
Figure 5.3: FORCE set of reactions predicted by OptForce for the overproduction of hexanoic acid in *E. coli*.
Figure 5.4: FORCE set of reactions predicted by OptForce for the overproduction of octanoic acid in *E. coli*. 
Figure 5.5: Impact of each genetic intervention predicted by OptForce on the yield of octanoate. Interventions that are in the left indicate that they are the most important ones while the interventions on the right contribute the least towards the product yield.
Figure 5.6: Impact of each genetic intervention predicted by OptForce on the yield of palmitic acid.
Figure 5.7: Venn diagram representing the shared genetic interventions predicted by OptForce for fatty acids of chain length C6 to C16.
Synopsis

The availability of large amounts of data quantifying various biological processes has led to a systems-based paradigm to analyze and redesign the genetic setup of organisms. This serves as a motivation behind the development of research in this thesis. A number of success stories have come about in areas where systems-level understanding has helped in understanding metabolic behavior. Reconstructions of organism-specific models of metabolism have shifted research in biotechnology from a “trial and error” approach to a more rational approach. In addition, the ever-increasing accuracy of quantitative data from these models has added value to these suggestions. To this end, the primary focus of this thesis was the development of computational tools to be used on genome-scale models of metabolism. The procedures outlined in this thesis aims to bridge the gap between experimental and computational metabolic engineering by borrowing data from different sources.

In this pursuit, we explain a graph-based algorithm that mines all possible metabolic pathways between a starting and a target metabolite of interest. The primary aim of research explained in this chapter is to add non-native functionalities to user-friendly microbial hosts such as E. coli or yeast so that their metabolic confines can be extended to produce bioproducts that they do not naturally synthesize. In view of avoiding bottlenecks faced by experimentalists working with heterologous gene cascades, we ensure pathways mined are closer in proximity to branching metabolites naturally
produced in the microbial hosts. Furthermore, we take care to prune pathways whose yields are not comparable to existing pathways.

Next, in Chapter 3, we introduced the OptForce procedure that is aimed at overcoming the limitations faced by its current counterparts in suggesting genetic manipulation strategies for overproduction. Unlike other procedures, in view of representing real-time experimental data, we have custom-built OptForce to incorporate metabolic flux analysis data within the procedure. OptForce first identifies the flux changes that happen in the network by contrasting the flux ranges for the wild-type strain and a hypothesized overproducing phenotype. We have built the procedure to even scrutinize reactions that are not well defined by experimentally measured flux values. Subsequently, OptForce identifies the minimal set of engineering interventions that coax the wild-type towards satisfying a pre-specified objective for the product. We demonstrated the computational efficacy of OptForce by identifying metabolic interventions for 1-butanol and succinate synthesis in *E. coli*. Interventions predicted by OptForce corroborate existing strategies for overproduction.

Importantly, in Chapter 4, we have validated the computational predictions from OptForce by integrating an experimental study aimed at overproducing flavanones in *E. coli*. As a highlight, OptForce suggested the importance of overexpressing acetyl-CoA carboxylase for increasing the availability of malonyl-CoA. The 100% increase in the yield of naringenin upon overexpressing acetyl-CoA carboxylase falls in line with our computational predictions. Notably, using OptForce, we were able to construct a
recombinant strain of *E. coli*, which is the highest naringenin-producer in a lab-scale fermentation process. Results presented in Chapter 5 suggest that we are the outset of demonstrating the importance of chain-dependent interventions that impacts the outcome of the fatty acid metabolism. While some interventions predicted by OptForce are straightforward and involved eliminating competing by-products, many others suggested complex mechanisms of compensating for precursors and intermediates.
Appendix A

Computing flux variability for the wild-type and overproducing networks

The following set, variable and parameter definitions are introduced to support the development of the optimization formulations needed for the derivation of the MUST and FORCE sets.

Sets:

N = \{i\} set of metabolites

M = \{j\} set of reactions

M^{data} = \{j\} subset of reactions with known flux values or ranges

M^{uptakes} = \{j\} subset of substrate transport reactions

M^{targets} = \{j\} subset of reactions targeted for overproduction

Variables:

v_{j} flux through reaction j

Parameters:

S_{i,j} stoichiometric coefficient of a metabolite i in a reaction j

v_{j}^{data,L}, v_{j}^{data,U} measured flux ranges for a reaction j in the set M^{data}

v_{j}^{uptake} flux value for uptake reaction j in the set M^{uptakes}

v_{j}^{target} target flux values for all reactions within set M^{targets}
Parameters \( v_{j}^{\text{data}, L} \) and \( v_{j}^{\text{data}, U} \) quantify range data available for a subset of reaction fluxes in the network. These data can be derived by directly measuring growth, uptake and/or secretion rates. In addition, MFA can provide information about many internal fluxes in the network either in the form of restraints or exact values. The use of data range parameters instead of fixed values allows us to make use of lower/upper bound data typically inferred from MFA data due to measurement error and other factors. Parameters \( v_{j}^{\text{data}, L} \) and \( v_{j}^{\text{data}, U} \) are set to the same value if an exact estimate is known for a particular flux.

Upon the incorporation of all available data encoded within parameters \( v_{j}^{\text{data}, L} \) and \( v_{j}^{\text{data}, U} \), the flux variability in the wild-type strain is derived by iteratively minimizing and maximizing each flux using a series of linear programming (LP) problems given by:

\[
\begin{align*}
\text{minimize/maximize} & \quad v_j \\ 
\text{subject to} & \quad \sum_{j=1}^{M} S_{ij} v_j = 0 \quad \forall i \in \mathbb{N} \\ & \quad v_{j}^{\text{data}, L} \leq v_j \leq v_{j}^{\text{data}, U} \quad \forall j \in M^{\text{data}} \\ & \quad v_j = v_j^{\text{uptake}} \quad \forall j \in M^{\text{uptakes}} \\ & \quad \text{LB}_j \leq v_j \leq \text{UB}_j \quad \forall j \in M 
\end{align*}
\]

Constraints (2) impose stoichiometric balances on the network whereas constraints (3) incorporates all known flux data for reaction fluxes present in set \( M^{\text{data}} \). Constraints (4) set the uptake of carbon and other substrates and constraints (5) impose global upper and lower bounds for the remaining fluxes. The solution to this sequence of LPs yields the lower \( (v_{j}^{\text{wt}, L}) \) and upper \( (v_{j}^{\text{wt}, U}) \) flux range values for all reactions in the wild-type
metabolic network. Reactions with tight flux ranges allude to direct or indirect coupling with the imposed wild-type network flux data whereas reactions with wide flux ranges tend to be largely insensitive to the available data. Parameters \( v_{j}^{w.t. L} \) and \( v_{j}^{w.t. U} \) estimated from the above LPs represent the maximal flux variability for the wild-type metabolic network.

Similarly, the flux ranges \( v_{j}^{* L} \) and \( v_{j}^{* U} \) consistent with the imposed set of overproduction targets \( v_{j} \geq v_{j}^{target} \) for all \( j \in M_{targets} \) can be derived by iteratively solving (for every \( j \) in \( M \)) the following sequence of LP formulations:

\[
\begin{align*}
\text{minimize/maximize} & \quad v_{j} \\
\text{subject to} & \quad \sum_{j=1}^{M} S_{i,j} v_{j} = 0 \quad \forall i \in N \quad (2) \\
& \quad v_{j} \geq v_{j}^{target} \quad \forall j \in M_{targets} \quad (6) \\
& \quad v_{j} = v_{j}^{uptake} \quad \forall j \in M_{uptakes} \quad (4) \\
& \quad LB_{j} \leq v_{j} \leq UB_{j} \quad \forall j \in M \quad (5)
\end{align*}
\]

This formulation does not include the constraint that imposes the flux data available for the wild-type network, instead through constraint (6) it sets the required production levels.
Appendix B

Bilevel formulation for the identification of MUST sets

The identification of sets MUST\textsuperscript{L} and MUST\textsuperscript{U} is conducted by testing whether inequalities \( v_j^{*,U} \leq v_j^{*,L} \) and \( v_j^{*,L} \geq v_j^{*,U} \) respectively, are satisfied by exhaustively testing one at a time all reactions in the network. Second-order MUST sets such as MUST\textsuperscript{LL}, MUST\textsuperscript{LU} and MUST\textsuperscript{UU} require exhaustively exploring all pairwise sums/differences of reactions which becomes computationally long running for genome-scale models due to the large number of LP problems that need to be solved. The elucidation of higher-order sets (e.g., MUST\textsuperscript{LLL}, MUST\textsuperscript{UUU}, etc.) using a brute force exhaustive search is thus quickly rendered intractable. The observation that only a (very) small fraction of all combinations of fluxes explored are members of the MUST sets motivated the development of a direct enumeration approach that circumvented the need to exhaustively assess all combinations. This underlying mathematical problem is similar to the one we faced before for the targeted enumeration of higher order synthetic lethals [45]. Therefore, we modified the developed bilevel optimization representation to track reaction combinations with membership in MUST sets as opposed to causing synthetic lethality.

We present the bilevel optimization formulation for the identification of the membership of the MUST\textsuperscript{LU} set. The corresponding optimization formulations for other higher-order MUST sets can be recapitulated in a straightforward fashion. A particular reaction pair
$(j_1, j_2)$ belongs to $\text{MUST}^{LU}$ if either flux $v_{j_1}$ must decrease or flux $v_{j_2}$ must increase to meet the imposed overproduction target. Mathematically, this implies satisfying the following inequality: $\max(v_{j_1} - v_{j_2}) < v_{j_1}^{wt,L} - v_{j_2}^{wt,U}$. The ranges for difference in flux values for two reactions in the overproducing network falls completely to the left of the corresponding ranges for the wild-type network (i.e., less than $v_{j_1}^{wt,L} - v_{j_2}^{wt,U}$) only when the flux of reaction $j_1$ decreases or the flux of reaction $j_2$ increases in the face of overproduction. When searching for $(j_1, j_2)$ pairs belonging to $\text{MUST}^{LU}$, reactions $j_1$ in $\text{MUST}^L$ and $j_2$ in $\text{MUST}^U$ are first excluded from consideration. Binary variables $y_{1,j}^L$ and $y_{2,j}^U$ encode whether a particular reaction pair $(j_1, j_2)$ belongs to $\text{MUST}^{LU}$ by both assuming a value of one. The following bilevel formulation is solved iteratively using integer cuts to exclude the previously found solution to identify the complete membership of set $\text{MUST}^{LU}$.

$$\begin{align*}
\max_{y_{1,j}^L, y_{2,j}^U} & \quad z = \sum_j (v_{j_1}^{wt,L} y_{1,j}^L - v_{j_1}^{wt,U} y_{1,j}^L) - (v_{j_2}^{wt,U} y_{2,j}^U - v_{j_2}^{wt,U} y_{2,j}^U) \\
\text{s.t.} & \quad \sum_j y_{1,j}^L = 1 \quad (7) \\
& \quad \sum_j y_{2,j}^U = 1 \quad (8) \\
& \quad \left\{\begin{array}{l}
\max_j \sum_j (v_{j} y_{1,j} - v_{j} y_{2,j}) \\
\text{s.t.} \quad \sum_i S_{i,j} v_j = 0 \quad \forall i \in \mathbb{N} \quad (9) \\
v_j = v_j^{\text{uptake}} \quad \forall j \in \mathbb{M}^{\text{uptakes}} \quad (10) \\
v_j \geq v_j^{\text{target}} \quad \forall j \in \mathbb{M}^{\text{uptakes}} \quad (11) \\
\text{LB}_j \leq v_j \leq \text{UB}_j \quad \forall j \in \mathbb{M} \quad (12) \\
z \geq 0 \\
y_{1,j}^L \in \{0,1\}, y_{2,j}^U \in \{0,1\}
\end{array}\right.
\end{align*}$$
We recursively solve the optimization problem after employing integer cuts until all the reaction pairs in $\text{MUST}^{LU}$ are exhaustively enumerated.
Appendix C

Bilevel formulation for the identifying the FORCE set

The collection of MUST sets encode all combinations (and possibilities) of flux changes that must happen in the network in the face of an imposed overproduction target. They can be represented as a logic diagram containing AND and OR operators. However, not every required network flux change must be imparted through a genetic modification. All MUST set requirements can simultaneously be satisfied by imparting a much smaller number of genetic modifications due to reaction flux couplings through stoichiometry. We refer to the minimal number of such genetic manipulations as the FORCE set. As was the case with the MUST sets, there typically exists multiple ways of constructing the FORCE set leading again to a logic diagram representation of representing the totality of engineering manipulations leading to the satisfaction of an imposed overproduction target.

We introduce an optimization formulation for the enumeration of the FORCE set that identifies engineering modifications that satisfy the imposed overproduction target for any feasible flux combination in the network. This implies that unlike OptKnock results, the identified FORCE set is valid irrespective of whether maximization of biomass or any other objective function drives flux allocation in the network. This conservative posture implies that it is possible that the overproduction target could be met after only a subset of the engineering modification in the FORCE set are imposed. The optimization
formulation is a bilevel problem accounting for three different types of direct engineering interventions leading to an increase, decrease or elimination of a reaction flux. Sets $M^U$, $M^L$ and $M^*$ contain the candidate reactions from the MUST sets that can be forced up, down or off, respectively to pursue the imposed overproduction target(s). Three separate sets of binary variables are defined to model these choices:

$$
\begin{align*}
    y_j^U &= \begin{cases} 
    1, & \text{if reaction flux } j \text{ is directly increased} \\
    0, & \text{otherwise} 
    \end{cases} \\
    y_j^L &= \begin{cases} 
    1, & \text{if reaction flux } j \text{ is directly decreased} \\
    0, & \text{otherwise} 
    \end{cases} \\
    y_j^\# &= \begin{cases} 
    1, & \text{if reaction } j \text{ is knocked out} \\
    0, & \text{otherwise} 
    \end{cases}
\end{align*}
$$

The proposed optimization formulation can be expressed as follows:

$$
\begin{align*}
    \min_{y_j^U, y_j^L, y_j^\#} & \sum_j y_j^U + y_j^L + y_j^\# \\
\text{s.t.} & \quad v_j \geq v_{\text{target}} & \forall j \in M^{\text{targets}} \\
 & \quad y_j^U + y_j^L + y_j^\# \leq 1 & \forall j \in M \\
\end{align*}
\tag{13}
$$

$$
\begin{align*}
    \min_{v_j} & \quad v_{\text{product}} \\
\text{s.t.} & \quad \sum_j S_{ij} v_j = 0 & \forall i \in N \\
 & \quad v_j = v_{\text{uptake}} & \forall j \in M^{\text{uptake}} \\
 & \quad v_j \geq v_j^{\text{max}} y_j^U + \text{LB}_j (1 - y_j^U) & \forall j \in M^U \\
 & \quad v_j \leq v_j^{\text{min}} y_j^L + \text{UB}_j (1 - y_j^L) & \forall j \in M^L \\
 & \quad \text{LB}_j (1 - y_j^\#) \leq v_j \leq \text{UB}_j (1 - y_j^\#) & \forall j \in M^* \\
 & \quad v_j \in \mathbb{R} \\
 & \quad y_j^U \in \{0,1\}, y_j^L \in \{0,1\}, y_j^\# \in \{0,1\}
\end{align*}
\tag{14, 15, 2, 10, 16, 17, 18, 5}
$$
The inner optimization problem finds the “worse-case” scenario for the production of the targeted product while the outer part aims to meet the overproduction target (even for the worst case scenario) with the minimum number of engineering modifications. Constraints (16-18) propagate the effect of engineering modifications encoded within $y_j^U$, $y_j^L$ and $y_j^*$ to the corresponding reaction fluxes. For example, if a reaction is selected for up regulation then the flux for this reaction is set greater than the maximum value allowed in the wild-type strain ($v_j^{\text{max}}$). Subsequently, we employ integer cuts and identify all the alternate solutions for $k$-interventions that guarantee the target yield for the product.
Bibliography:


Academic Background:

- **Ph.D.** in Bioinformatics and Genomics (expected May, 2012), **Pennsylvania State University** GPA: 3.50/4.00
  Thesis: “Using computations to analyze and redesign metabolism”

- **M.S.** (dual) Operations Research and Industrial Engg (2008), **Pennsylvania State University** GPA: 3.50/4.00

- **B.E.** in Mechanical Engineering (2005), **Anna University, Chennai, India** GPA: 9.00/10.00

Publications:

