

Identification of Essential Metabolites in Metabolite Networks

by

Cai Long

B.Sc., Jilin University, 2009

Minor in B.Econ, Jilin University, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE

in

The Faculty of Graduate Studies

(Biomedical Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

October 2012

© Cai Long 2012

Abstract

Metabolite essentiality is an important topic in systems biology and as such there has been increased focus on their prediction in metabolic networks. Specifically, two related questions have become the focus of this field: how do we decrease the amount of gene knock-out work loads and is it possible to predict essential metabolites in different growth conditions? Two different approaches to these questions: interaction-based method and constraints-based method, are conducted in this study to gain in depth understanding of metabolite essentiality in complex metabolic networks.

In the interaction-based approach, the correlations between metabolite essentiality and the metabolite network topology are studied. With the idea of predicting essential metabolites, the topological properties of the metabolite network are studied for the *Mycobacterium tuberculosis* model. It is found that there is strong correlation between metabolite essentiality and the degree and the number of shortest paths through the metabolite. Welch's two sample T-test is performed to help identify the statistical significance of the differences between groups of essential metabolites and non-essential metabolites.

In the constraint-based approach, essential metabolites are identified *in-*

Abstract

silico. Flux Balance Analysis (known as FBA), is implemented with the most advanced *in-silico* model of *Chlamydomonas Reinhardtii*, which contains light usage information in 3 different growth environments: autotrophic, mixotrophic, and heterotrophic. Essential metabolites are predicted by metabolite knock out analysis, which is to set the flux of a certain metabolite to zero, and categorized into 3 types through Flux Sum Analysis. The basal flux-sum for metabolites is found to follow an exponential distribution, it is also found that essential metabolites tend to have larger basal flux-sum.

Table of Contents

| | |
|-----------------------------|------|
| Abstract | ii |
| Table of Contents | iv |
| List of Tables | viii |
| List of Figures | ix |
| List of Acronyms | xi |
| Acknowledgements | xii |
| 1 Introduction | 1 |
| 1.1 Metabolite Essentiality | 1 |
| 1.2 Outline | 3 |
| 2 Literature Review | 6 |

Table of Contents

| | | |
|----------|--|-----------|
| 2.1 | Systems Biology | 6 |
| 2.1.1 | Basic Steps in Systems Analysis | 7 |
| 2.1.2 | Systems Analysis of Metabolite Essentiality | 9 |
| 2.2 | Interaction-based Approach | 10 |
| 2.2.1 | Graph Theory in Systems Biology | 10 |
| 2.3 | Constraints-based Approach | 12 |
| 2.3.1 | Flux Balance Analysis | 13 |
| 2.4 | Subjects of Applications | 16 |
| 3 | Metabolite Essentiality and Reaction Network Topology . | 17 |
| 3.1 | Graph Theory and Essential Metabolites | 17 |
| 3.1.1 | Graph Theory | 18 |
| 3.1.2 | Categories of Metabolites | 30 |
| 3.2 | Application to <i>Mycobacterium Tuberculosis</i> | 32 |
| 3.2.1 | <i>Mycobacterium Tuberculosis</i> | 32 |
| 3.2.2 | Gaps in the Metabolite Network iNJ661 | 34 |
| 3.2.3 | Metabolite Essentiality and Network Degree | 34 |
| 3.2.4 | Metabolite Essentiality and the Degree of Neighbors . | 36 |

Table of Contents

| | | |
|----------|---|-----------|
| 3.2.5 | Metabolite Essentiality and Clustering Coefficient . . | 38 |
| 3.2.6 | Metabolite Essentiality and Network Betweenness. . . | 38 |
| 3.3 | Conclusion | 40 |
| 4 | Constraint Based Identification of Essential Metabolites . | 42 |
| 4.1 | Application: Microalgae | 43 |
| 4.1.1 | <i>Chlamydomonas Reinhardtii</i> | 44 |
| 4.1.2 | Biofuel from Microalgae | 46 |
| 4.2 | Flux Balance Analysis | 49 |
| 4.2.1 | Mathematical Reconstruction of a Biochemical Net- work | 50 |
| 4.2.2 | Model Validation | 51 |
| 4.2.3 | Mass Balance | 52 |
| 4.2.4 | Constraints | 53 |
| 4.2.5 | Objective Function | 56 |
| 4.2.6 | Linear Program Solver | 58 |
| 4.2.7 | Identification of Essential Metabolites | 59 |
| 4.3 | Flux Sum Analysis | 61 |

Table of Contents

| | |
|---|-----------|
| 4.3.1 Procedure for Flux Sum Analysis | 61 |
| 4.3.2 Conclusion | 69 |
| 5 Conclusion | 70 |
| Bibliography | 72 |
| Appendix | 85 |
| A.1 Appendix 1: ELM in Mycobacterium Tuberculosis | 85 |
| A.2 Appendix 2: Universal Metabolites | 86 |
| A.3 Appendix 3: Root No-production Metabolites in iNJ661 | 92 |
| A.4 Appendix 4: Root No-consumption Metabolites in iNJ661 . . | 93 |
| A.5 Appendix 5: Common Essential Metabolites in All 3 Growth Conditions | 95 |
| A.6 Appendix 6: Biomass Function(Objective Function) for Dif- ferent Growth Conditions | 98 |
| A.7 Appendix 7: Matlab Codes | 102 |
| A.7.1 Interaction-based Approach Code | 102 |
| A.7.2 Constraint-based Approach Code | 108 |

List of Tables

| | | |
|-----|--|----|
| 4.1 | Oil yield from algae and from other sources,(Chisti, 2007) | 44 |
| 4.2 | Oil content from microalgae (Chisti, 2007)(Li et al., 2010) | 48 |
| 4.4 | Constraints for different growth conditions | 55 |
| 4.5 | Number of different types of essential metabolites in different growth conditions | 67 |

List of Figures

| | | |
|-----|---|----|
| 1.1 | Interaction-based approach and constraints-based approach are both implemented to study metabolite essentiality. | 4 |
| 2.1 | Linear Programming | 15 |
| 3.1 | Pathway diagram from a simple biosystem consists of 7 metabo- lites | 19 |
| 3.2 | Examples of Orphan reaction and Gap. A: the missing re- action (Gap) creates two dead-end reactions; B: the reaction catalyzed by unknown gene product can be a orphan reac- tion (Reprinted from Orth, Jeffrey D, 2010(Orth and Palsson, 2010), with permission from 2010 Wiley Periodicals, Inc.) . . | 27 |
| 3.3 | Characterization of problem metabolites in metabolic net- works (Satish Kumar et al., 2007) | 28 |
| 3.4 | Probability distribution of degree of metabolites | 35 |
| 3.5 | Probability distribution of neighbor's degree | 36 |

List of Figures

| | | |
|-----|--|----|
| 3.6 | Average sum of neighbor's degrees for EM, EUM and NEM | 37 |
| 3.7 | Probability distribution of Clustering Coefficient | 39 |
| 3.8 | Average betweenness of EM, EUM and NEM | 39 |
| 3.9 | Probability distribution of betweenness | 40 |
| 4.1 | Reconstructed metabolic network of <i>C. reinhardtii</i> , (Reprinted from (Boyle and Morgan, 2009)) | 46 |
| 4.2 | Mathematically reconstruction of a biochemical network | 51 |
| 4.3 | Model validation | 52 |
| 4.4 | Mass balance definition | 53 |
| 4.5 | The total basal flux-sum for C.Reinhardtii in 3 different conditions. The blue part represents the total basal flux-sum for Universal Metabolites. | 63 |
| 4.6 | Probability distribution of metabolites with certain basal flux-sum. | 64 |
| 4.7 | 2 types of essential metabolites: Type AE and Type BE | 66 |
| 4.8 | Number of different type of essential metabolites in different growth conditions | 67 |

List of Acronyms

EM: Essential Metabolite

EUM: Essential Unusual Metabolite

NEM: Non-Essential Metabolite

ORF: Open Reading Frame

EC number: Enzyme Commission number

KEGG: Kyoto Encyclopedia of Genes and Genomes

SBML: Systems Biology Markup Language

FBA: Flux Balance Analysis

FSA: Flux Sum Analysis

LP: Linear Programming

DFBA: Dynamic Flux Balance Analysis

M.T: *Mycobacterium tuberculosis*

C.R: *Chlamydomonas Reinhardtii*

Acknowledgements

I would like to express my most sincere appreciation and gratitude to my supervisor, Prof. Bhushan Gopaluni for his excellent supervision and precious advice throughout the whole period of my study at the University of British Columbia. His motivation and inspiring attitude is exemplary. I learned how to conduct a research project from him, which I am sure will be a life time beneficial.

My thanks go to Dr. Ezra Kwok and whole Process Modeling and Control lab, their ideas, experience and generously sharing help me grow.

Special thanks go to Dr. Roger Chang and Dr. Nathan Lewis in University of California, San Diego, for releasing their data from their experiments. I am also grateful for Dr. Pan-Jun Kim in University of Illinois at Urbana Champaign for his kindly help. Moreover, I would like to convey my thanks to all the faculty, staff and fellow postgraduate students in Chemical and Biological Engineering department at UBC.

Last, I leave the warmest part of my heart for my beloved parents, who gave birth to me, enlightened me and educated me with their unconditional support and continuous love.

Chapter 1

Introduction

Every cell is characterized by the presence of a complex network of metabolites connected by chemical reactions. These reactions are catalyzed by specialized proteins called enzymes. There are usually thousands of reactions inside the cell, and at the same time, there are thousands of metabolites (Samal et al., 2006). It is well-known that certain reactions are vital to the survival and maintenance of essential functions of a cell. These are called “essential” reactions. Notably, the essentiality of reactions or metabolites may change depending on the environmental conditions.

1.1 Metabolite Essentiality

The metabolites involved in the reaction network can be classified into two categories: essential metabolites and nonessential metabolites. While cells are known to be quite robust to perturbations in the reaction network, the absence of essential metabolites could cause serious damage or even death. On the other hand, recent investigations have shown that non-essential metabolites cause very little or no impact on the living cells (Jeong et al., 2003).

1.1. Metabolite Essentiality

The study of essential metabolites has received significant interest from the systems biology community due to several reasons:

First, the loss of essential metabolites will diminish cell viability. Most drugs exert therapeutic effects by binding and regulating the activity of a particular metabolite, set of proteins or nucleic acid targets in the pathogenic microbes. Therefore identification of essential metabolites will be beneficial to investigate new inhibitors of disease and potential drug targets as inhibitors, the identification and validation of essential metabolites compose an important step in drug discovery process (Samala, 2006).

Second, analysis of essential metabolites will help researchers understand the complex metabolite networks, which may yield better predictions in *in vivo* cellular behavior, and have better insight into the complex relationship between cell components and systems-level cellular phenotypes (Jamshidi and Palsson, 2007).

Third, many drugs that are highly successful in human clinical use mimic a substrate or product of essential metabolites. For example, folic acid is an essential biomolecule, which needs to be synthesized *de novo* by many bacteria, and dihydropteroate synthase, an enzyme in the folic acid biosynthesis pathway, synthesizes dihydrofolate from p-aminobenzoate. Sulfonamide-based drugs are structural analogs of p-aminobenzoate and act by inhibiting dihydropteroate synthase. Many bacterial infections are effectively treated with sulfonamides, as they mimic an essential substrate and competitively inhibit an essential enzyme. There are lots of other examples of inhibition of essential metabolites by mimicking their substrates (Bermingham and Derrick, 2002).

Hence, the study of metabolite essentiality will be beneficial not only to the understanding of systems biology (especially with complex metabolite networks), but also is expected to play an important role in helping to identify drug targets.

The systems biology approach, with its combination of computational, experimental and observational enquiry, is highly relevant to drug discovery and the optimization of medical treatment regimes. Particularly, computer simulation and analysis, along with traditional bioinformatics approaches, have frequently been proposed to significantly increase the efficiency of drug discovery (Kitano, 2002).

Currently, the main drawback is due to the cost and time consumption of the approaches taken to identify essential metabolites, which is mainly gene knock-out experiments.

With the objective to reduce the time and cost of determining essential metabolites, we are going to study the correlation between metabolite essentiality and metabolite network topology, and try to predict essential metabolites using constraint-based modeling.

1.2 Outline

In Chapter 2, we will review recent progress made on the topic of correlation between metabolite essentiality and network topology, the lethality-centrality rule, and other findings. We will also discuss the importance of choosing *C.Reinhardtii*, which is a model organism of microalgae, as our in-

1.2. Outline

vestigation object. Finally, the basic concepts of systems biology and linear programming will be discussed here.

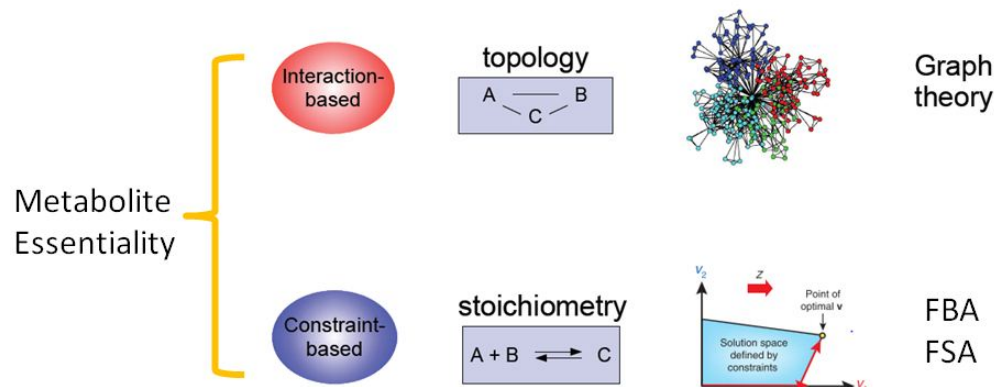


Figure 1.1: Interaction-based approach and constraints-based approach are both implemented to study metabolite essentiality.

As indicated in Figure 1.1, two modeling approaches: interaction-based approach and constraints-based approach are both implemented to study metabolite essentiality.

In Chapter 3, **interaction-based approach** model *iNJ661* of *Mycobacterium tuberculosis* is used to identify essential metabolites. First, we categorize the essential metabolites into 3 different types: Essential Unusual Metabolites, Universal Metabolites, and Non-Essential Metabolites. Secondly, we introduce a method based on adjacency matrix to find the gaps in the model and fill the model with GapFill, a method developed by Orth Jeffrey to fill the gaps (Orth and Palsson, 2010). Finally, we study the correlations between metabolite essentiality and the topology parameter of metabolic networks. The metabolite degree, degree of neighbors, clustering

coefficient of each metabolite, and betweenness of the metabolite network is discussed, respectively.

In Chapter 4, **constraints-based approach** model organism, *Chlamydomonas Reinhardtii*, is chosen to conduct the study of predicting essential metabolites by constraints based modeling. With the light usage information, we are able to predict essential metabolites in different growth conditions, and find the common essential metabolites. We also propose the categorization of essential metabolites by using Flux Sum Analysis.

In Chapter 5, we summarize the results and discuss possible future work.

Chapter 2

Literature Review

At the core of our understanding of biological processes and underlying systems, is a characterization of function and interactions of their constituent parts. Systems biology, which takes into account the key characteristics of complex systems, including essentiality, emergence, robustness and modularity, is one of the essential topics. Today, systems biology is established as a fundamental interdisciplinary science that focuses on detailed studies of the complex mechanisms, which orchestrate the interactions between various biomolecules that compose life.

2.1 Systems Biology

Systems biology, broadly speaking, is a subject that attempts to investigate the behavior and relations of all the ‘elements’ in a given functioning biological system (Kitano, 2002). It aims at system-level understanding of biological processes and biochemical networks as a whole. This “system-oriented” new biology is shifting our focus from examining particular molecular details to studying the information flow at all biological levels: genomic DNA, mRNA, proteins, informational pathways, and regulatory networks

(Price and Lee, 2010). Systems biology approaches seek to study the complexity of life to help in understanding how the cellular networks work together. It requires a broad interdisciplinary knowledge of molecular and cell biology, biochemistry, informatics, mathematics, computing, and engineering. It provides tools to understand the various functions and properties of biological systems, and predicts systems behavior under various physiological conditions.

2.1.1 Basic Steps in Systems Analysis

A widely used *in silico* quantitative systems biology tool to relate the genotype to the phenotype comprises of four steps:

1. **Collection of information from ‘omics’ and literature data on the target organism**

Genome sequencing is the starting point for the systems analysis. After that, the genome is annotated to define genes and transcribed elements, and open reading frame (ORF)s are delineated. The most challenging part of genome annotation, which is assigning molecular function, can be done through comparison of related genes and proteins with known functions, for instance, by predicting protein function based on sequence similarity with proteins of previously annotated function in database such as Uniprot or Metacyc databases. This approach generates a genome annotated with Enzyme Commission(EC) numbers which contains the catalytic information of the gene product.(Francke et al., 2005)

2. Reaction network model

After genomic sequencing, the reaction network reconstruction process are performed. This process is carried out by assigning reactions to annotated genes using metabolic databases such as Kyoto Encyclopedia of Genes and Genomes (KEGG). Reaction properties that include reversibility and localization to cellular compartments are also built into the network model. Incomplete reaction pathways or lack of metabolic functions are quite common in network models. Often, reorganization of reactions is required to make the model consistent with the known physiological and biochemical characteristics.

3. Mathematical description of the network model

The reaction network model is described by a set of reaction rate equations so as to allow quantitative analysis. Stoichiometric matrix is a popular representation of the network model and is rather straightforward to generate. The large number of reactions in these models makes it almost impossible to develop models manually. A variety of software programs are available for automatically building the mathematical models based on reaction network information. *Antimony* is one such software that generates a model in Systems Biology Markup Language (SBML) (Smith et al., 2009).

4. Evaluation and refinement of the model

Metabolomic and transcriptomic data from high-throughput experiments is used to evaluate and refine the model and iteratively improve its capacity to predict phenotypes. Different types of analysis can be performed on the refined model to optimize or predict the prop-

erties of the network. In this context, constraint based modeling approaches such as flux balance analysis (FBA) have been widely studied to predict flux through metabolic path ways, optimal growth media, product yields, and other factors relevant to bioprocess design and optimization (Hatzimanikatis et al., 2005; Hjersted and Henson, 2009; Hucka, 2003; Kauffman et al., 2003; Krieger et al., 2004; Lee et al., 2006; Meadows et al., 2010)

2.1.2 Systems Analysis of Metabolite Essentiality

Serval attempts, both *in vivo* or *in silico*, have been made to study the metabolite essentiality. Among *in silico* methods, systems biology is the most popular one. Rigoustos states that “Systems biology is an integrated approach that brings together and leverages theoretical, experimental, and computational approaches in order to establish connections among important molecules or groups of molecules in order to aid eventual mechanistic explanation of cellular processes and systems.” (Rigoutsos, 2007). Aiming at a system-level understanding of biological systems, systems biology provides a tool to understand the various properties of biological systems and predict system behavior under different physiological conditions (Palsson, 2009). Just as theoretical and mathematical biology deal with the mathematical modeling of certain aspects of biology, systems biology deals with the prediction of various function from the metabolic networks and provides a mechanistic bridge between phenotype and genotypes.

Flux Balance Analysis (Ghim et al., 2005; Imieliski et al., 2005; Kim et al., 2007; Li et al., 2011; Palsson, 2003) and Flux-sum analysis (Chung and Lee,

2009) are two popular systems biology approaches that are used in understanding metabolite essentiality. Metabolite essentiality is commonly determined *in silico* by monitoring cell growth while changing the concentration of a given metabolite to zero.

An *in vivo* method for studying metabolite essentiality is to implement wet-lab gene knock out experiments to find out the essential enzymes, and determine the essential metabolites based on the knock-out results. These experiments often provide more reliable models, however, there is usually missing information about reactions or mechanisms in the *in silico* network (Lamichhane et al., 2011).

2.2 Interaction-based Approach

2.2.1 Graph Theory in Systems Biology

Graph theory has been used for analyzing data for protein interaction network, and is receiving more and more attention in predicting essential metabolites.

Metabolite essentiality has gained enormous interest in the recent years. One of the most intriguing questions in the study of metabolite essentiality is to understand the connection between biological and topological importance of metabolite networks. One of the first attempts at studying this topic was made in 2001 on the *S. cerevisiae* protein-protein interaction network (Bro et al., 2006). It was also investigated under the topic “centrality and lethality” by Jeong and colleagues (Jeong et al., 2001). Since then, many

2.2. Interaction-based Approach

efforts have been put into the protein-protein interaction network, the correlation between protein-protein network topology and protein essentiality was confirmed by many researchers (Coulomb et al., 2005; Hahn and Kern, 2005; Yu et al., 2004, 2007; Zotenko et al., 2008). The recent availability of large protein interaction databases has fueled the analysis of protein interaction networks and it has been demonstrated that protein essentiality could be strongly related to some topological parameters of these networks. For example, protein networks are found vulnerable when a highly connected “hub” is removed (He and Zhang, 2006). Computational analysis shows that removing hubs increases the proportion of unreachable pairs of nodes (metabolites) and the mean shortest path length between all pairs of reachable nodes in the network. (Albert et al., 2000)

However, not much work has been reported on the correlation between metabolite essentiality and topology. Mahadevan *et al* (Mahadevan and Palsson, 2005) conjectured that low degree metabolites (metabolites connect with small number of other metabolites) are just as likely to be recognized as essential metabolites as high degree metabolites (metabolites connect with large number of other metabolites). Areejit Samal generated a random matrix to explain this phenomenon (Samal et al., 2006). Other graph driven methods to analyze complex cellular networks are emphasized by many researchers (Aittokallio and Schwikowski, 2006a).

Traditional methods to study the essential metabolites mainly rely on creating random mutants of a gene and therefore require a large amount of work. For *in silico* metabolite network predictions like flux balance analysis, the complexity and integrity of the metabolite model would greatly affect

the accuracy of the prediction. Although a lot of progress has been made in studying the topological and functional properties of metabolite networks, very little effort has been put into understanding the correlations between metabolite essentiality and topology. We are trying to involve more topological parameters of the metabolite network, which would help to increase the accuracy of addressing essential metabolites, and to better understand the metabolite network structures.

2.3 Constraints-based Approach

Another approach used in predicting essential metabolites is constraints-based, in which Flux Balance Analysis(FBA) and other linear programming based tools are implemented with biology mathematic models.

The development of high-throughput experimental techniques in recent years has led to an explosion of genome-scale data sets for a variety of organisms. Considerable efforts have yielded complete genomic sequences and gene-annotation based metabolite models for dozens of organisms. A prudent approach to gain biological understanding from these complex data involves the development of mathematical models, simulation, and analysis and techniques (Kim et al., 2008). In these complementary efforts, many analytical tools have been developed to use these models in computational investigations of model organisms. One method in particular, Flux Balance Analysis (FBA), is a powerful mathematical approach to assess the ability of an organism to grow on a particular substrate or in particular environment and also be used to assess the effect of metabolic gene deletions under various

growth conditions (Palsson, 2009).

2.3.1 Flux Balance Analysis

Flux balance analysis is a widely used constraint based approach for studying biochemical networks (Orth et al., 2010). A reaction network is assumed to be at steady state in order to overcome the lack of knowledge of metabolite concentration or details of enzyme kinetics of the system (Edwards et al., 2001). It is difficult and in some cases impossible to provide real time metabolite concentration or enzyme kinetics using current experimental techniques. The model of the steady state reaction network is defined by a linear matrix equation that contains reaction stoichiometric coefficients.

Constraints are typically of two types, one is the stoichiometry matrix, which is generated from mass balance equations (Kauffman et al., 2003). These matrix-based constraints ensure the total amount of any compound being produced must be equal to the total amount being consumed at steady state. The other type of constraints are given by the reactions, which define the maximum and minimum allowable fluxes of the reactions.

However, the dynamics of the metabolic networks sometimes are too important to be neglected, Dynamic Flux Balance Analysis (DFBA), a widely used approach for studying biochemical networks and phenotype optimization method, was introduced to generate dynamic prediction of substrate, biomass and concentrations in batch culture (Meadows et al., 2010). Many tools have been developed to perform FBA and DFBA, for instance, FBA-

2.3. Constraints-based Approach

SimVis(Grafahrend-Belau et al., 2009), SurreyFBA((Gevorgyan et al., 2010), and CobraToolbox(Becker et al., 2007).

With the network reconstruction data from Nanette R Boyle (Boyle and Morgan, 2009), and Kyoto Encyclopedia of Genes and Genomes (KEGG), DFBA is utilized to predict the biomass production and lipid concentration of *C.Reinhardtii*. (Hucka, 2003)(Becker et al., 2007), the simulation and optimization results will be compared with existing experimental results (Smith et al., 2009).

Linear programming(LP) is used to identify single or multiple optimal solutions from constraints in constraints based modeling.

Linear Programming

Linear Programming (also known as LP, or Linear Optimization) is a mathematical method to determine the optimal solution (such as maximum or minimum) in a given mathematical model with a list of constraints represented as linear relationships. The linear objective function, subject to linear equality and linear inequality constraints is used to find the optimal point. The optimal solution normally lies in a corner of the constraint polytope. Occasionally, the objective function has the same value along a whole edge and all the points on that edge are optimal values. In this rare case the objective function is "parallel" to the edge of the polytope.

The figure below represents a simple example of linear programming problem.

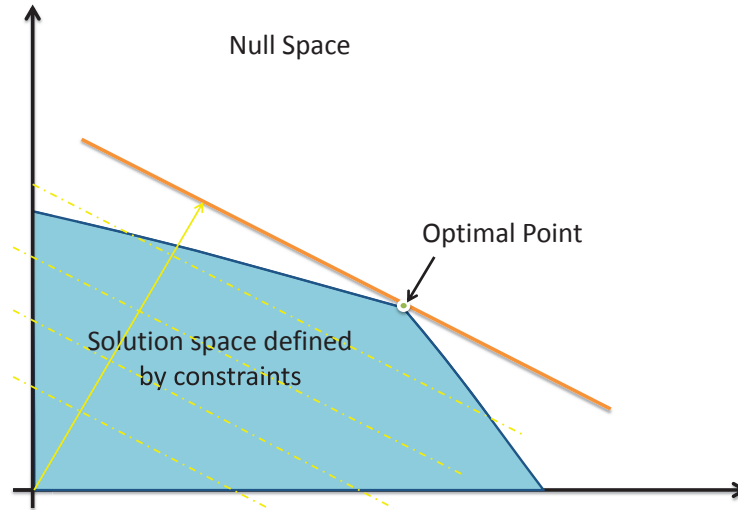


Figure 2.1: Linear Programming

LP problems can usually written into form:

$$\begin{aligned} &\text{Maximize } c^T x \\ &\text{subject to } Ax \leq b \\ &\text{and } x \geq 0 \end{aligned}$$

where x represents the vector of variables, c and b are vectors of coefficients, A is the coefficient matrix. Most of the metabolic engineering LP problems are convex under-determined. An under-determined system means there are less equations than variables, while an over-determined system means there are more equations than unknowns.

2.4 Subjects of Applications

Two modeling approaches, interaction-based and constraints-based , are applied on different model organisms.

Mycobacterium tuberculosis, model *iNJ661*, is used in the interaction-based approach, with a list of essential metabolites from G.Lamichhane, J.Freundlich et al. in 2011 through a wet-lab approach. The correlations between metabolite essentiality and the topology parameter of metabolic networks are being studied, to improve the accuracy of the essential metabolites predication. The main reason to use this model is that it's the first organism with a full list of essential metabolites with wet-lab experiemental results.

Constraints-based approach is applied on *Chlamydomonas Reinhardtii*, model *iRC1080*, as it is the latest and only model with light usage, which enable us to implement simulation under three different growth conditions. Flux balance analysis is utilized to identify the essential metabolites, and flux sum analysis is used to categorize the essential metabolites.

Chapter 3

Metabolite Essentiality and Reaction Network Topology

One of the most interesting questions in the study of metabolite essentiality is to understand the connection between biological and topological importance of metabolite networks. In this chapter, we investigated the degree, neighbor's degree, clustering coefficient and betweenness of the essential metabolites and unessential metabolites, try to find the correlation between essential metabolites and reaction network topology.

3.1 Graph Theory and Essential Metabolites

Before we study the correlation between metabolite essentiality and reaction network topology properties, the basic concepts of graph theory, and the methodologies we used to classify essential metabolites are discussed.

3.1.1 Graph Theory

Graph

A graph is a mathematical abstraction of structural relationships between discrete objects. A graph usually refers to a collection of “nodes” and “edges” that connect the vertices. An edge could be either directed, meaning there is a distinction from one node to another or undirected, which means there is no direction from one node to another. Several methods or data structures can be used to describe the nodes and edges, an easy and widely used one is adjacency matrix M . An adjacency matrix is an n by n matrix, where n is the number of nodes in the graph. If there is an edge from node x (in metabolite network, metabolite X) to node y (in metabolite network, metabolite Y), then the element $M(x, y)$ is 1 (or in general the number of edges between x and y), otherwise it would be zero.

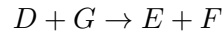
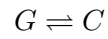
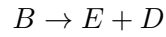
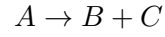
$$M(x, y) = n$$

n is the number of reactions in which metabolite X acts as a reactant and metabolite Y is a product.

The representation of complex cellular networks as a graph has made it possible to systematically investigate the topology and function of these networks using well-understood graph-theoretical concepts that can be used to predict the structural and dynamical properties of the underlying network (Aittokallio and Schwikowski, 2006b).

3.1. Graph Theory and Essential Metabolites

A simple biosystem, which consists of 4 reactions and 8 metabolites, is constructed for demonstration:



While \rightarrow means non-reversibility, the symbol \rightleftharpoons in the reaction indicates it's reversible. A pathway diagram representing this simple system is shown as Fig 3.1,

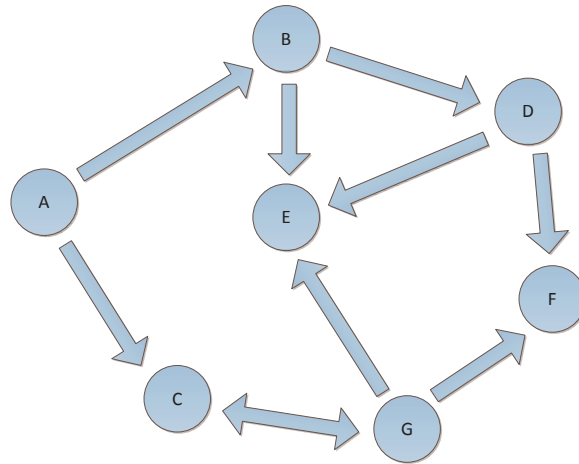


Figure 3.1: Pathway diagram from a simple biosystem consists of 7 metabolites

The adjacency matrix X can be derived for the above reaction system in a straightforward way. So the Figure 3.1 could be interpreted as :

$$X \triangleq \begin{matrix} & \begin{matrix} A & B & C & D & E & F & G \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \\ E \\ F \\ G \end{matrix} & \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \end{pmatrix} \end{matrix}$$

A very interesting and useful property of adjacency matrix is that the (i, j) element of X^k gives the number of k -step edge sequences from node i to node j (Jiang et al., 2009). For instance, element $(2, 5)$ represents that there are two 2-step edge sequences from node b to node e; as it is clear that we can find in the graph that there are two 2-step edge sequences from node b to node e: $\{b \rightarrow c \rightarrow e\}, \{b \rightarrow d \rightarrow e\}$

For a digraph with N nodes and an adjacency matrix X , the following matrix

$$R = (X + X^2 + X^3 + \dots + X^N)$$

is defined as a connectivity matrix, the (i, j) th element of R indicates the number of directed paths from node i to node j . In our research, we only focus on two-step connections, which means,

$$R = X + X^2 + X^3$$

3.1. Graph Theory and Essential Metabolites

The connectivity matrix for the digraph in Fig 3.1 is

$$X \triangleq \begin{matrix} & \begin{matrix} A & B & C & D & E & F & G \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \\ E \\ F \\ G \end{matrix} & \begin{pmatrix} 0 & 1 & 2 & 1 & 3 & 2 & 1 \\ 0 & 0 & 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 2 & 2 & 1 \end{pmatrix} \end{pmatrix}$$

$X_{1,3} = 2$, it means from node A to node C , there are 2 pathways with less than 2 nodes in between. The connectivity matrix is used to find the gaps in our study, as well as to study the nature of metabolite reaction network topology.

Stoichiometric and Adjacency Matrices For large systems, especially complex metabolite networks, the adjacency matrix can be obtained from the corresponding stoichiometric matrix. The stoichiometric matrix is widely used in the computational systems biology, the matrix S stores the stoichiometric coefficients associated with each reaction flux in a network. In the above formulation, both internal fluxes and boundary fluxes, which transport material into or out of the system, are included in S . Typically, a number of inequalities are introduced to constrain the boundary (also called injection) fluxes depending upon the external media (Edwards, 2000) (Beard et al., 2002). Stoichiometric matrix can be obtained from databases

like MetaCyc(Caspi et al., 2010), CSB.DB (Kopka et al, 2005) quite easily. More details about stoichiometric matrix can be found in chapter 4.

In the Stoichiometric matrix, the i th reaction $A + B \rightarrow C + D$ showing that A and B will be consumed to produce C and D , so both A and B are adjacent to C and D . For any metabolite X in stoichiometric matrix S , j^A is the row number of the metabolite X . For i th reaction, we define the boolean equivalent of any reachability between any two metabolites A and B as follows:

$$K(j^A, j^B) = \begin{cases} 0, & \text{if } S(j^A, i) \cdot S(j^B, i) = 0, \\ 1, & \text{if } S(j^A, i) \cdot S(j^B, i) \neq 0, \end{cases}$$

For a system with i reactions, the adjacency matrix would be:

$$R(j^A, j^B) = \sum K(j^A, j^B)^{\#i}$$

The MATLAB code can be found in the Appendix 6.

Network Topology Definitions and Notations

For a directed graph G , we shall write $D(x)$ as the degree of a node x in $V(G)$, which is the total number of edges (both in- or out- of the vertex) of x .

Degree The degree of a certain metabolite in the metabolite network is equal to the number of reactions it is included, either as a reactant or product.

$$D(X) = \sum_{i=1}^n M_{x,i} + \sum_{j=1}^n M_{j,x}$$

The degree distribution of the metabolite network measures the proportion of nodes in the network having degree k . We have

$$P(k) = \frac{n_k}{n}$$

where n_k is the number of nodes in the network of degree k , and n is the size of the network.

Neighbor's Degree The sum of the degrees of a certain metabolite's neighbors, which reveal the numbers of metabolites connected to the metabolite indirectly but very still very close to that metabolite, is also very important.

An interesting and useful property of adjacency matrix is: (i, j) element of X^k gives the number of k -step edge sequences from node i to j . So $ND(X)$, the number of degrees of the neighbors of metabolite X is:

$$ND(X) = \sum_{i=1}^x M_{i,x}^2 + \sum_{i=1}^x M_{x,i}^2$$

The average of the neighbors' degrees of metabolite X $Avg_ND(X)$ is calculated as:

$$Avg_ND(X) = \frac{ND(X)}{D(X)}$$

Clustering Coefficient Next, in graph theory, clustering coefficient represents how the nodes tend to cluster together. Here we study the local clustering coefficient for each node, which quantifies how close its neighbors are to being a clique(a complete circle), is defined as the proportion of links between the vertices within its neighborhood divided by the number of links that could possibly exist between them. For a directed graph, e_{ij} is distinct from e_{ji} and therefore for each node N_i there are $k_i(k_i - 1)$ links that could exist among the nodes within the neighborhood, here k_i is the degree(in and out) of the node.(Mason and Verwoerd, 2007)

$$C_i = \frac{|\{e_{jk}\}|}{k_i(k_i - 1)}$$

Betweenness Another important topological feature of the network has received much attention - betweenness, which measures the total number of nonredundant shortest paths going through a certain node or edge (Girvan and Newman, 2002). For node k , the betweenness can be defined as following:

$$P_k = \sum N_{ij}$$

$$N_{ij} = \begin{cases} 0, & \text{if no shortest path through node } k, \\ 1, & \text{if the shortest path through node } k, \end{cases}$$

Missing Information in the Biological Models

The genomes of several microorganisms have been completely sequenced and annotated in the past decade, however, even the most complete genomes are not perfect; they have missing information, which may lead to inaccurate predictions of the model. A key challenge in the automated generation of genome-scale reconstructions is the elucidation of the gaps and the subsequent generation of hypotheses to bridge them. This challenge has already been recognized and a number of computational approaches have been under development to resolve these issues. Feist et al. (2009); Oh et al. (2007); Orth and Palsson (2010); Satish Kumar et al. (2007)

There are two types of missing information (Orth and Palsson, 2010):

- **Gaps:** Gaps are created by dead-end reactions. When a reaction that consumes or produces a metabolite is missing, it creates a dead-end. For instance, experiments reveal a producing reaction but no consuming reaction, or no producing reaction but a consuming reaction). Example A in Figure 3.2 is a common type of gap. In FBA, these reactions carry no fluxes and therefore can lead to wrong predictions. There are several reasons for gaps in the metabolic network:

1. *Biological*: An enzyme in a completed reaction pathway is missing in the biochemical network. For example, iAF1260 for *E.coli* K-12 MG1655 (Edwards, 2000).
2. *Scope*: Metabolites produced in metabolism but then enter other systems not included in the network models like transcription and, translation may leave gaps in the models. For example, tRNAs in iAF1260 (Chavali et al., 2008).
3. *Knowledge*: It is not known what biochemical reaction produces or consumes a certain metabolite. A new biological discovery must be made to fill this gap.

- **Orphan reactions**: There are two different types of orphan reactions:

1. Reactions known to exist but are catalyzed by unknown gene product. They are the result of missing knowledge of the metabolism of an organism, (which gene or genes code for their enzymes.)
2. Reactions catalyzed by gene products with unknown functions. Even most well-studied organisms have many gene with unknown functions, eg: *E.coli* K-12 MG1655 has 981 partially or fully uncharacterized. A database named ORENZA lists global orphan reactions recently found.

Example B in Figure 3.2 shows one type of orphan reactions, which is catalyzed by a unknown gene product.

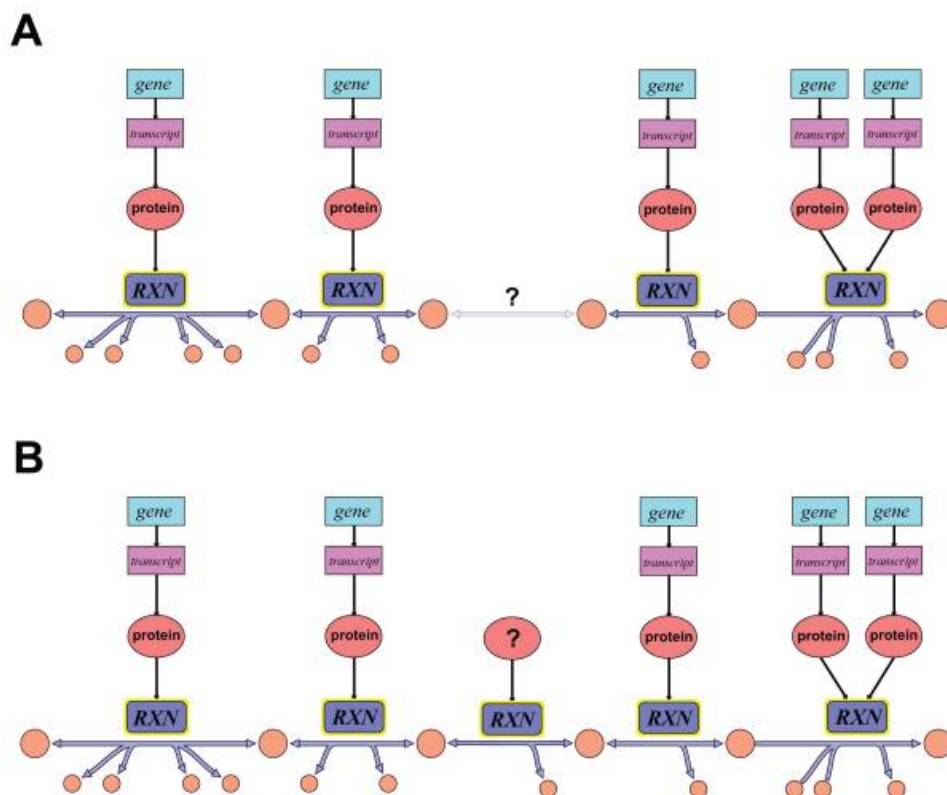


Figure 3.2: Examples of Orphan reaction and Gap. A: the missing reaction (Gap) creates two dead-end reactions; B: the reaction catalyzed by unknown gene product can be a orphan reaction (Reprinted from Orth, Jeffrey D, 2010(Orth and Palsson, 2010), with permission from 2010 Wiley Periodicals, Inc.)

Identifying the Gaps in a Reaction Network

Gaps exist in almost every metabolic reaction network due to lack of information. In this thesis, a novel approach to find these gaps using what is called an adjacency matrix is proposed. The adjacency matrix contains information about interactions between metabolites. Gaps in metabolic re-

3.1. Graph Theory and Essential Metabolites

constructions are defined as (i) metabolites which cannot be produced by any of the reactions or imported through any available uptake pathways in the model; or (ii) metabolites that cannot be consumed by any of the reactions or exported by any secretion pathways in the network. The first kind of metabolites are recognized as root no-production metabolite (e.g.; metabolite A in Figure 3.3) and the second situation is recognized as root no-consumption metabolites (e.g.; metabolite B in Figure 3.3). There will be no flow through these metabolites at steady state due to their inability to connect with the rest of the network. Consequently, the metabolites directly related to them will be affected as well, which are defined as downstream no-production metabolites (e.g.; metabolite C in Figure 3.3) and upstream no-consumption metabolite (e.g.; metabolite D in Figure 3.3) respectively (Satish Kumar et al., 2007).

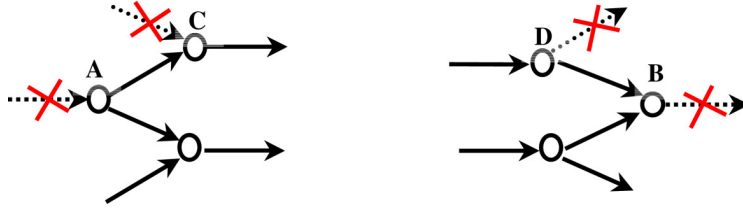


Figure 3.3: Characterization of problem metabolites in metabolic networks (Satish Kumar et al., 2007)

The root no-production metabolites and root no-consumption metabolites are caused by the gaps in the system, while they introduce more downstream or upstream no flux metabolites simultaneously. In the connectivity matrix, the value of element $X(i, j)$ shows the number of pathways from node i to node j , if $X(i, j) = 0$, there is no flux from metabolite i to metabolite j . Set

$$K_j = \sum_{i=1,2\dots n} X(i,j)$$

Clearly, if $K_j = 0$, the j th metabolite is a root no-production metabolite. Similarly, set

$$C_i = \sum_{j=1,2\dots n} X(i,j)$$

C_i represent the number of pathways producing metabolite i , so if $C_i = 0$, it would be a root no-consumption metabolite.

Gaps could be filled by different methods like BNICE (Hatzimanikatis et al., 2005), GapFill (Satish Kumar et al., 2007) , SMILY (Reed et al., 2006), etc.

Current gap-filling methods: In computational biology, gap-filling methods are quite useful as they improve the predictive capabilities of models by making them more realistic by characterizing a previously unknown gene, a model refinement tool.

- a) Computational methods: (to filling the gaps, reactions from database , KEGG, etc are used)

1. GapFind and GapFill: minimize the total number of gaps in a metabolic network model. Gapfind: a mixed integer linear programming algorithm that can identify every gap in a network by identifying blocked metabolites (cannot be produced or consumed at steady-state under any conditions)

GapFill: another mixed integer linear programming(MILP) method

3.1. Graph Theory and Essential Metabolites

to minimizing the gaps by reversing the existing reactions, adding new reactions or transport reactions, or reactions between compartments, with minimal number of model modifications.

2. SMILEY: predicts reactions that are likely missing from a network when the model predicts no growth but experiment predicts growth (based on the OptStrain algorithm).
 3. GROWMATCH: uses experimentally determined gene essentiality data to identify incorrect model predictions.
 4. other methods. OMNI, for example.
- b) Experimental methods. Several experimental methods could also be introduced to filling the gaps.

After refining the model by find and fill the gaps, we categorize metabolites into 3 different types novelly: Universal Metabolites, Essential Unusual Metabolites, and Non-Essential Metabolites.

3.1.2 Categories of Metabolites

In this study, the metabolites are divided into three groups:

Universal Metabolites (UM): Some inorganic or cofactor metabolites, such as H_2O , ATP, or NADP+, have been found to exist universally more than 90% organisms whether they are prokaryotes or eukaryotes. These metabolites are called universal metabolites.

Essential Unusual Metabolites (EUM): The metabolites whose absence will cause cell death, but are not UM are called Essential Unusual

3.1. Graph Theory and Essential Metabolites

Metabolites. In order to find out the essential metabolites, a large amount of transposon insertion mutants are created to represent the disruption and therefore the loss of function of more than 2000 genes. UM and EUM are usually seen as essential metabolites together, in most of the studies.

The list of EUM in *M. Tuberculosis* can be find in Appendix 1.

Non-Essential Metabolites(NEM): All other metabolites are called non-essential metabolites.

The universal metabolites are usually treated as essential metabolites because most living matter cannot survive without the metabolites like H_2O and ATP. However, this definition could bring confusion and misunderstanding in the research, especially in the drug target studies. For example, metabolites as H_2O and ATP are to be recognized as essential because very few living cell can live without H_2O and ATP, but they can hardly be used as a drug target. (Martelli et al., 2009)

We are trying to find a method to predict the metabolites which are not common metabolites, but still, the fact without them will significantly eliminate the cell growth. A innovative idea is to filter all the common seen metabolites, in other words, to pick out the Essential Unusual Metabolites (EUMs).

Obtain EUM and UM With a database of 250 species of organism, we define metabolites those could be found in more than 90% of the organisms are universal metabolites. Some of the list of metabolites in different species are obtained from a database investigated by Kim (Kim et al., 2007), other

3.2. Application to *Mycobacterium Tuberculosis*

are from KEGG pathway database. The comprehensive list of the universal metabolites are listed in Appendix 2.

All the UM metabolite are found to be essential metabolites in most of the recent studies about essential metabolites in different organisms (Martelli et al., 2009). The next main step is to study the correlation between the topology of the metabolite network and the metabolite essentiality for each type. Before that, it's very important to refine the model we are going to use, as there are missing information as gaps and orphan reactions.

3.2 Application to *Mycobacterium Tuberculosis*

A list of essential metabolites for *Mycobacterium Tuberculosis*(MTB) was obtained from G.Lamichhane, J.Freundlich et al., (Lamichhane et al., 2011) from a *in vivo* approach. 5126 independent, genotyped and archived mutants with disruption in both intra- and intergenic regions were created, followed by a statistical analysis to predict the essentiality of the genes. The molecules produced by reactions encoded by essential enzymes are classified as essential metabolites. This is also the first comprehensive report of a large number of essential molecules so far.(Duarte et al., 2004)

3.2.1 *Mycobacterium Tuberculosis*

Mycobacterium tuberculosis(MTB) is a pathogenic bacterial species in the genus *Mycobacterium* and the causative agent of most cases of tuberculosis, it was first discovered in 1882 by Robert Koch. However, with 1.77 million

3.2. Application to *Mycobacterium Tuberculosis*

deaths from TB in 2007, this disease ranks second only to human immunodeficiency virus as a cause of death from an infectious agent. The estimate that more lives may be lost in 2011 due to tuberculosis than in any year in history is alarming. In 1993, the gravity of the situation led the World Health Organisation (WHO) to declare tuberculosis a global emergency in an attempt to heighten public and political awareness. Complete genome sequence of the best-characterized strain of *Mycobacterium tuberculosis* has been determined in 1998 by S.T. Cole, R.Brosch *et al*, (Cole et al., 1998a) to enhance the understanding of the biology of the slow-growing pathogen and to help the conception of new prophylactic and therapeutic interventions. New-resistant tuberculosis appear almost every year, so new drugs are needed to treat the infections caused, the attempt to determine essential metabolites would benefit the drug target filtration. Gyamu, Joel, et al, identified essential metabolites and enzymes for *M.tuberculosis* using a genetics-based approach,(Lamichhane et al., 2011) which provide a new blueprint for developing effective chemical probes of *M. tuberculosis* metabolism.

The cell envelope of *M. tuberculosis*, contains an additional layer beyond the peptidoglycan that is exceptionally rich in unusual lipids, glycolipids and polysaccharides. Cell-wall components such as mycolic acids, myco-cerosic acid, phenolthiocerol, lipoarabinomannan and arabinogalactan, are generated by novel biosynthetic pathways, and several of these may contribute to mycobacterial longevity, trigger inflammatory host reactions and act in pathogenesis. Little is known about the mechanisms involved in life within the macrophage, or the extent and nature of the virulence factors produced by the bacillus and their contribution to disease.(Cole et al., 1998b) In addition to the mycolic acids, the cell envelope contains a wide array of

3.2. Application to *Mycobacterium Tuberculosis*

distinctive lipids and glycolipids that confers extreme hydrophobicity to the outer surface of the organism.(Sibley et al., 1988, 1990)

The model of Tuberculosis we used is *iNJ661* for *Mycobacterium tuberculosis H37Rv*, developed by N. Jamshidi. (Jamshidi and Palsson, 2007)

3.2.2 Gaps in the Metabolite Network iNJ661

Using graph theory stated in 3.1, there are two different types of gaps found in *iNJ661* model for MTB. For the list of root no-production metabolites, please see Appendix 3. For a comprehensive list of root no-consumption metabolites, please see Appendix 4.

3.2.3 Metabolite Essentiality and Network Degree

It has been found that essential metabolites have higher degree than non-essential metabolites in *E.coli* (He and Zhang, 2006). However, in *M.tuberculosis*, we calculated the average degree of essential metabolites and non-essential metabolites, respectively. The average degree of essential metabolites is found to be 83, much higher than the non-essential ones, which is just 9. It is mainly because the universal metabolites, which are counted as essential metabolites, usually have much higher degree than the others with a noticeably average degree of 95.89.

In order to find out if there is statistically significant difference between essential metabolites and non-essential metabolites, Welch two sample test is implemented on the essential metabolites and non-essential metabolites,

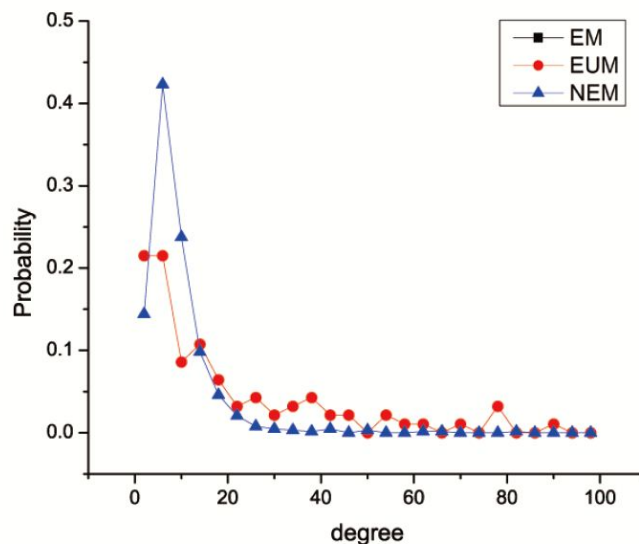


Figure 3.4: Probability distribution of degree of metabolites

with a p value of 0.00066. When comparing with the t -test result of EUMs and NEMs, which has a p value of 0.1588 shows there is no statistically significant difference existing if UMs are not included. It is concluded that the the higher degree of UMs is the reason for the difference between EMs and NEMs, and this supports He's finding.

Another interesting fact is the fraction of essential metabolites among the 10% most connected is 64.8% and there is no essential metabolites in the least connected. However, it is interesting to see that the t -test shows there is a significant difference between the downstream degree of EUMs and NEMs, with a p -value of 0.00014, it means usually EUMs has smaller downstream degree, so there is a higher possibility that a metabolite with fewer products is EUM.

Figure 3.4 is the degree distribution of *iNJ661*. The horizontal axis is the degree of the metabolite, while the vertical axis is the probability of the metabolite, so for any given spot, it shows the probability of metabolites with a certain degree. It shows that essential metabolites have a higher probability with higher degrees, especially larger than 20. It also shows that most of the non-essential metabolites have degrees under 20, and barely any NEMs larger than 20.

3.2.4 Metabolite Essentiality and the Degree of Neighbors

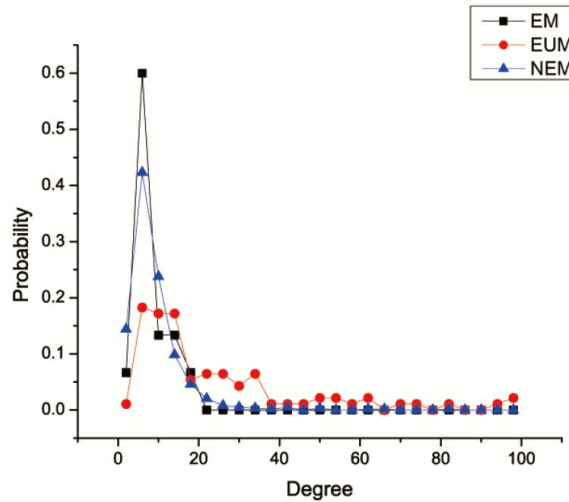


Figure 3.5: Probability distribution of neighbor's degree

Here we examine the total degree of neighbors and the average degree of neighbors for EM, EUM, NEM, respectively. The average sum of the neighbor's degrees for EM, EUM and NEM are shown in Figure 3.5.

With a Welch's two sample t -test, it is clear that both EM and EUM

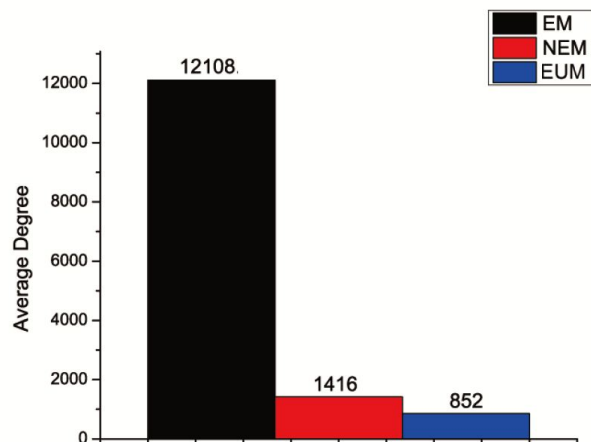


Figure 3.6: Average sum of neighbor's degrees for EM, EUM and NEM

have a distribution with larger degree of their neighbors compared to NEMs, with p values of 0.0206 and 0.0003. The mean of EM is 12108, 8 times larger than that of NEM, which has a mean of 1416. The main reason is that EM has incredibly high indirectly-connected neighbors. The mean of EUM is 852, and we can see from Figure 3.6 that they have much higher probability with neighbor's degree larger than 10, and almost all the NEMs's neighbor's degrees are under 20.

Interestingly, we found there is no significant statistical difference between both the average degrees of EM and NEM (p value = 0.3952), EUM and NEM (p value = 0.9455), which means for all the metabolites, the average degrees of their neighbors are not related to the fact it's essential or not, statistically.

3.2.5 Metabolite Essentiality and Clustering Coefficient

With the model of *iNJ661*, when it comes to clustering coefficient, we found that there is no true difference between EM and NEMs (p value = 0.256), the averages of them are also quite close, 0.272 for EM and 0.234 for NEM .

We observed that EUMs, the means of which is only 0.07, shows a visible difference from the NEMs. t -test results show the EUMs do have a smaller clustering coefficient, with a p -value of 0.0051. The fraction of metabolites with 0 clustering coefficient is much higher in the EUMs than other 2 groups. Figure 3.7 shows the prolixity distribution of clustering coefficient for all 3 type of metabolites, in which more EUMs have a clustering coefficient of 0.

This interesting result shows that we can reliably associate metabolite essentiality with this parameter, but is just limited to EUMs, which is useful as the UMs can be derived from the database straightforwardly. Small clustering coefficient could be used as an indicator for the EUMs.

3.2.6 Metabolite Essentiality and Network Betweenness.

According to our investigation, both UMs and EUMs are shown to have shortest path through, the means of which are 8924 and 1310, respectively, while the average of NEMs is just 666, the p value for Welch's two sample t -test is 0.001 for UMs and NEMs. There is a significant difference between UMs and NEMs. It's important to note that NEMs have more shortest path through them. According to Figure 3.8, it can be concluded that EM and EUMs have great probability with higher betweenness. So the network

3.2. Application to *Mycobacterium Tuberculosis*

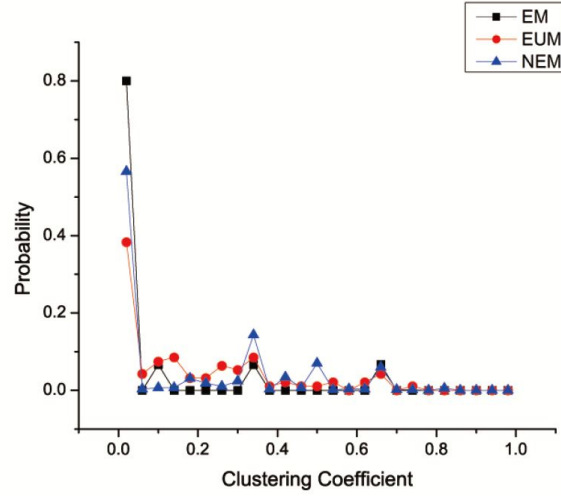


Figure 3.7: Probability distribution of Clustering Coefficient

betweenness could also be used as an indicator for the metabolite essentiality.

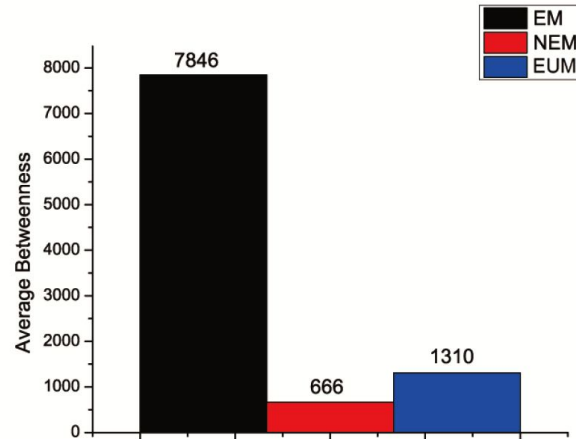


Figure 3.8: Average betweenness of EM, EUM and NEM

Figure 3.9 is about the probability distribution of betweenness, we could

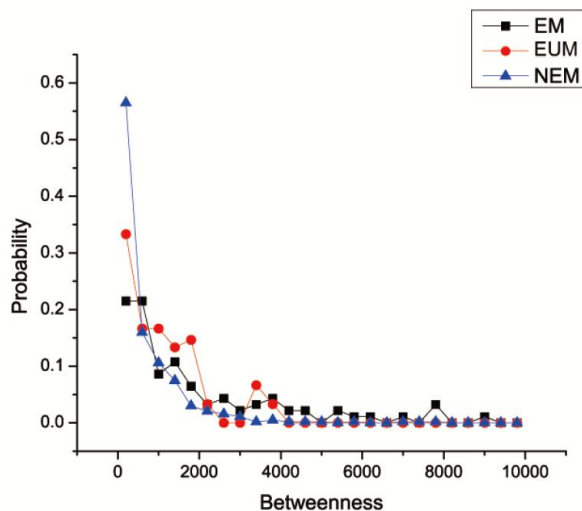


Figure 3.9: Probability distribution of betweenness

find the distribution follows a exponential distribution, and when the betweenness is larger than 3000, only probabilities of EM and EUM are above 0, and NEMs are all 0.

3.3 Conclusion

We looked systematically for correlations between the essentiality of genes and their topological characteristics in interaction networks. We have found that the metabolite essentiality is significantly related to the parameter of the metabolite in the metabolic network. The EMs are usually with larger degree, more neighbors' degree and more shortest path through, notably, the EUMs have smaller clustering coefficient.

While the essential metabolites are derived from the essential genes and

3.3. *Conclusion*

approved by the experiments, it is possible that gene essentiality is also related to metabolite topology parameters, this could be evaluated by future studies.

Chapter 4

Constraint Based Identification of Essential Metabolites

Flux Balance Analysis and Flux Sum Analysis are two alternate approaches to graph theory that are often used to identify the essential metabolites. Unlike graph theory, which is a generic statistical predication, the constraint based approaches (Flux Balance Analysis and Flux Sum Analysis) identify the essential metabolites in-silico, and would further decrease the amount of wet-lab experiments for validating essential metabolites. With the most advanced model of *C. Reinhardtii*, we identified essential metabolites under three different growth conditions, and categorized the essential metabolites using Flux Sum Analysis.

4.1 Application: Microalgae

Microalgae are ubiquitous sunlight driven cell factories in fresh water or marine systems, they convert CO_2 to food, biofuels or other high value bioactive products, and even cosmetic products (Spolaore et al., 2006). The number of algal species have been estimated to be more than one million with a majority being microalgae (Metting, 1996). Among all the potential sources, microalgae are now recognized as the only source of renewable biodiesel that is capable of meeting the global demand for transport fuels.

Compared to the first generation sources of biofuel, microalgae have greater potential as a reliable alternate energy source. Table 4.1 about oil yield from algae and other sources below demonstrates the advantage of cultivating microalgae. The higher concentration of lipid content in microalgae is one reason for this, as lipid contains quite high energy. The lipid concentration can often exceed 80% while 20%-50% are quite common. (Beer et al., 2009) Moreover, the fast doubling time of microalgae makes it possible to generate large quantities of biomass, which could be further processed to get different types of biofuels.

Currently, several species of microalgae have gained public and scientific attraction. However, for the following reasons there is still enormous scope for engineering micro algae to increase their production:

1. Little experience with the development of closed large scale photobioreactors.
2. High material costs for closed, highly efficient bioreactor systems.

4.1. Application: Microalgae

| Crop | Oil Yield(L/ha) |
|------------|-----------------|
| Corn | 172 |
| Soybeans | 446 |
| Jatropha | 1892 |
| Coconut | 2689 |
| Oilpalm | 5950 |
| Microalgae | 5000-15000 |

Table 4.1: Oil yield from algae and from other sources,(Chisti, 2007)

3. High energy requirement for cultivation (e.g. mixing). Expensive harvesting (cells need to be separated from medium which is time and/or energy consuming) (Metting, 1996).

4.1.1 *Chlamydomonas Reinhardtii*

Among many types of microalgae, green algae *C. Reinhardtii* is selected for this study for the following reasons:

- *C.Reinhardtii* is a model organism for the process of photosynthesis in plants (Harris, 2001), and a model for photosynthetic hydrogen production (Melis and Happe, 2004). Model organisms are simplified representative systems whose study enables researchers to extrapolate their understanding to other complex organisms. A number of efforts have been made on studying *C.Reinhardtii* and full nuclear genome sequence has been assembled in 2007 (Merchant et al., 2007) (Maul et al., 2002) (Vahrenholz et al., 1993) (Boer et al., 1985).

- *C.Reinhardtii* can be cultivated under different conditions, either autotrophic (from simple inorganic molecular and using energy from light), auxotrophic (relying on organic acid and light) or heterotrophic (with organic acids only).
- In addition, the time for *C.Reinhardtii* to grow to a mature individual is 5 to 6 hours under laboratory conditions, with a total fatty acid content of the isolated strain of 25%. The composition of fatty acids in the species of microalgae was mainly docosanoic acid methyl ester, tetradecanoic acid methyl ester, hexadecanoic acid methyl ester and nonanoic acid methyl ester.

Cells of *C. reinhardtii* are oval-shaped, typically 10 μ m in length and 3 μ m in width with two flagella at their anterior end. This algae contains several mitochondria and a unique chloroplast which occupies 40% of the cell volume and partly surrounds the nucleus(May et al., 2008). Figure 4.1 shows the reconstructed metabolic network of *C.Reinhardtii*. This unicellular green algae, closely related to photoreceptors of multicellular organisms, offers a simple life cycle, easy isolation of mutants, and a growing array of tool and techniques for molecular genetic studies (Li et al., 2010; Rupprecht, 2009). Recently, *C. Reinhardtii* have received more attention, because of its potential to generate biofuel to meet the growing clean energy demands.

In our study, model iRC1080, the newly reconstructed genome-scale metabolic network for *C.Reinhardtii* with a novel light-modelling approach that enables quantitative growth prediction for a given light source, is chosen to investigate the essential metabolites in *C.Reinhardtii*.

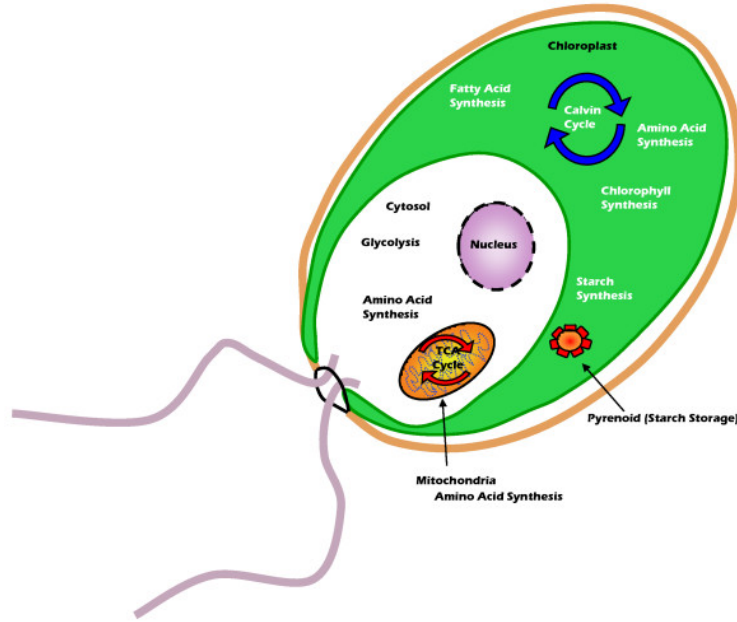


Figure 4.1: Reconstructed metabolic network of *C. reinhardtii*, (Reprinted from (Boyle and Morgan, 2009))

4.1.2 Biofuel from Microalgae

A biofuel is a solid, liquid or gaseous fuel derived from any biological carbon source including treated municipal and industrial wastes. Biofuels can be derived either from land-based crops or marina plants as microalgae. Three main types of biofuels are now produced from microalgae: biohydrogen, biodiesel, ethanol from fermentation of biomass.

Biohydrogen from Microalgae As a fuel, hydrogen causes less environmental impact whether in stationary engines, gas turbines or automotive vehicles. Microalgae have the genetic, metabolic and enzymatic charac-

teristics for hydrogen which cannot be provided by any land-based plants. During photosynthesis, the microalgae convert water molecules into hydrogen ions H^+ and oxygen. The hydrogen ions are then converted into H_2 by the enzyme hydrogenase (Hahn et al., 2004). The photosynthetic production of O_2 results in rapid inhibition of the enzyme hydrogenase and the production of H_2 is inhibited. Therefore, cultivation of microalgae for the production of hydrogen must take place under anaerobic conditions (Brennan and Owende, 2010).

Hydrogen production in *Chlamydomonas* has to take place at an efficiency of 7% under outdoor conditions to be commercially viable. While maximum efficiency for this process has been calculated to be between 6% to 10%. (Rupprecht et al., 2006)

Biodiesel from Microalgae Microalgae has shown great potential in the economical biodiesel production. Microalgae commonly double their biomass within 24h, which makes it possible to produce enough biomass for production of oil. There are two main large producing methods for the biomass: raceway pond and photobioreactors. Photobioreactors provide much greater oil yield compared with raceway ponds, but raceways ponds are cheaper. Both are technically feasible.

Currently, some naturally isolated microalga *Chlamydomonas* (for instance, sp MCCS 026) have been proven to be valuable candidates for biodiesel production as they have high growth rate and lipid content. They require a simple and comparatively low cost culture medium (Morowvat et al., 2010). The oil content in different kinds of microalgae can be found in the

4.1. Application: Microalgae

| Microalga | Lipid content (%dry weight) |
|----------------------------------|-----------------------------|
| <i>Botryococcus braunii</i> | 25-75 |
| <i>Chlorella sp.</i> | 28-32 |
| <i>Cryptocodinium cohnii</i> | 20 |
| <i>Cylindrotheca sp.</i> | 16-37 |
| <i>Dunaliella primolecta</i> | 23 |
| <i>Isochrysis sp.</i> | 25-33 |
| <i>Monallanthus salina N</i> | 20 |
| <i>Nannochloris sp.</i> | 20-35 |
| <i>Phaeodactylum tricornutum</i> | 20-30 |
| <i>Chlamydomonas Reinhardtii</i> | 30 - 60 |
| <i>Schizochytrium sp.</i> | 50-77 |

Table 4.2: Oil content from microalgae (Chisti, 2007)(Li et al., 2010)

table below:

Biomethane from Microalgae Microalgae has been investigated for biomethane production for a long time, it can be grown in large amounts (150 -300 tons per ha per year (Degen, 2001)), which leads to a theoretical yield of 200,000 - 400,000 m^3 of methane per ha per year. However, due to the high cost of biomass, and the low production capacity compared to the high demand of commercial gas, biogas is now usually a mixture of carbon dioxide gas and biomethane (Schenk et al., 2008).

Despite the advantages of algae as a source of biofuels, there are still significant challenges that must be addressed before algal biofuels can be

widely used. One of the main concerns is the biodiesel from algae is not yet economically competitive with fossil fuels or corn ethanol: the cost to producing gasoline is about \$ 1.86 per gallon (according to retail price in 2009), while for algal biodiesel, it will be \$2.5 - \$25(range depends on algae productivity) (Schmidt et al., 2010).

4.2 Flux Balance Analysis

Flux Balance Analysis(FBA) calculates the flow of metabolites (also known as flux), and is widely used as a tool to predict metabolite behavior such as growth rate of an organism or the rate of production of a bio-technologically important metabolite. With the assumption that the system will reach a steady state under any given environmental condition, the regulated metabolite network is set to satisfy a set of feasible constraints. Once the constraints and fluxes are identified, optimization techniques could be used to evaluate the performance of the biological system under different conditions, such as varying objective functions or bounds on certain reactions, growth on different media, or of bacteria with different gene knockouts. FBA can be further used to predict the yields of important cofactors such as ATP, NADH or NADPH (Kauffman et al., 2003; Lee et al., 2006).

Flux Balance Analysis can be divided into 4 steps as follows:

4.2.1 Mathematical Reconstruction of a Biochemical Network

Metabolite network reconstruction is the fundamental step in FBA, it involves generating a model that describes the system of interest. This process can be further decomposed into three parts typically performed simultaneously during model construction: data collection, metabolic reaction list generation, and gene-protein relationship determination .

After genome-scale metabolic reconstruction, a stoichiometric matrix S could be generated from the metabolic reactions, S is an $m \times n$ matrix of stoichiometric coefficients that captures the underlying reaction of the biochemical network. The rows of S correspond to the compounds, while the columns of S correspond to reactions. The entries in each column are the stoichiometric coefficients of the metabolites participating in a reaction. Negative elements of the matrix represent the consumption of compounds and positives elements denote production, for the metabolites not participating in a particular reaction, the coefficient is zero (Palsson, 2003). Figure 4.2 shows the basic procedures for mathematically reconstruction of a biochemical network. The reactions are obtained from the complex gene annotation database, and then converted into stoichiometric matrix.

The genome-scale *C.reinhardtii* metabolic network used in this study consists of 1080 genes, associated with 2190 reactions and 1068 unique metabolites, and encompasses 83 subsystems distributed across 10 compartments (Chang et al., 2011).

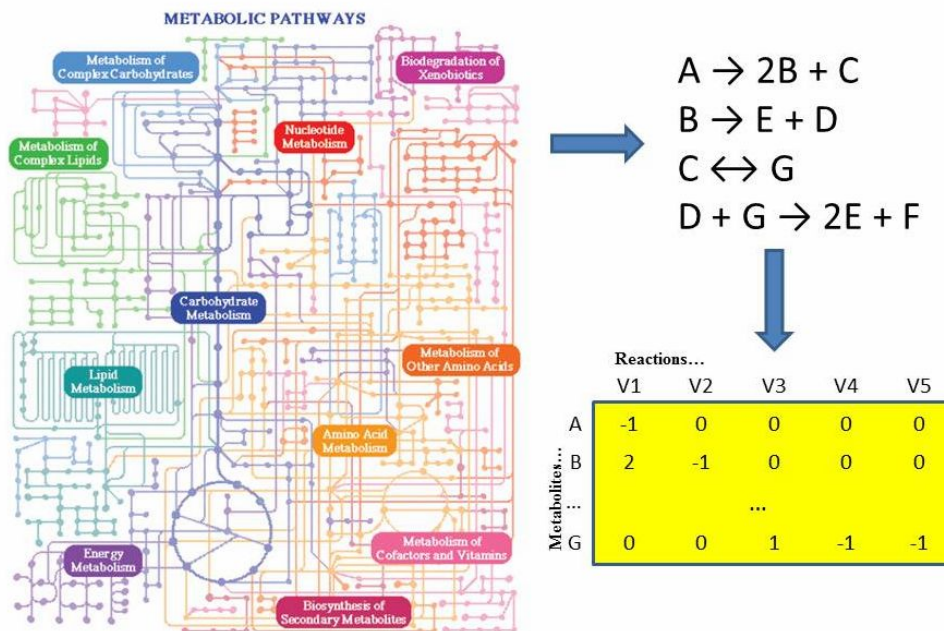


Figure 4.2: Mathematically reconstruction of a biochemical network

4.2.2 Model Validation

Even the most complete models are not perfect; they might contain missing information, which are called "gaps", the incomplete reconstructions may lead to prediction of erroneous genetic interventions for a targeted over-production or the elucidation of misleading organizational principles and properties of the metabolic network. Several computational and experimental methods can be used to address the gaps to help make more realistic predictions. As Figure 4.3 shows, the dead-end metabolites are identified.

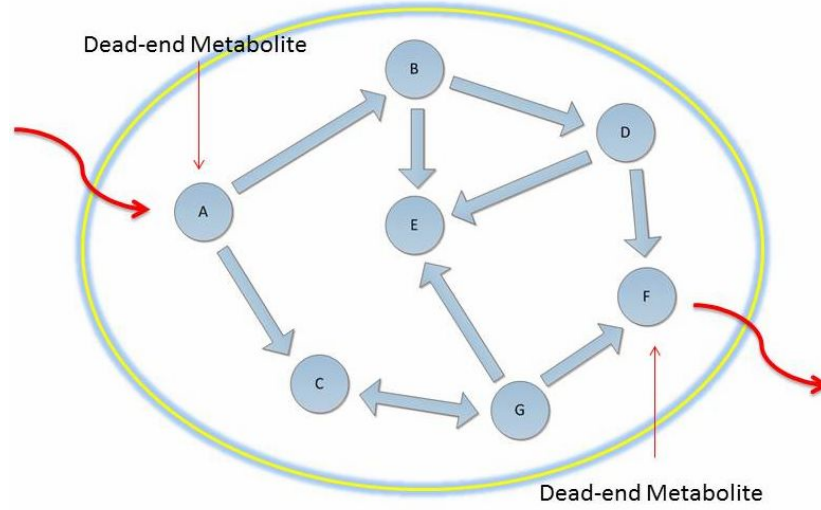


Figure 4.3: Model validation

4.2.3 Mass Balance

After the network matrix is reconstructed, mass balance can be defined in terms of the flux through each reaction and the stoichiometry of that reaction in the following form

$$\frac{\partial x}{\partial t} = Sv$$

v is the vector of fluxes with elements corresponding to the fluxes in given reactions. In steady state, the change amount of a metabolite x over time t within the whole system becomes zero, yielding :

$$Sv = 0$$

Figure 4.3 explains the basic mechanism of mass balance definition.

$$\begin{aligned}\frac{dA}{dt} &= -v1 + v6 \\ \frac{dB}{dt} &= 2v1 - v2 \\ &\dots \\ \frac{dG}{dt} &= v3 - v4 - v5\end{aligned}$$

→

$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dG}{dt} \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 1 \\ 2 & -1 & 0 & 0 & 0 & 0 \\ \dots & & & & & \\ 0 & 0 & 1 & -1 & -1 & 0 \end{bmatrix} \begin{bmatrix} v1 \\ v2 \\ v3 \\ v4 \\ v5 \\ v6 \end{bmatrix} = 0$$

Stoichiometric Matrix(S) v

In steady state, $Sv = 0$

Figure 4.4: Mass balance definition

4.2.4 Constraints

One way to add additional constraints to the metabolic network and calculate the fluxes in the network is to measure fluxes in the metabolite network. Usually, it's hard to measure the exact flux values, so ranges of allowable flux values are incorporated as additional constraints. Constraints could be physicochemical, topological or environmental. Physicochemical constraints are physical laws like conservation of energy and mass; topological constraints contains information of metabolites within different cellular compartments; and environmental constraints include nutrient availability,

pH and temperature that vary over time and space. The constraints imposed by the thermodynamics (e.g. effective reversibility or irreversibility of reactions) and enzyme or transporter capabilities (e.g. maximum uptake or reaction rates) are considered and incorporated into the model. It should be emphasized that these constraints are based on what may be considered “hard-wired” constraints the metabolic system must obey.

$$\alpha_i \leq v_i \leq \beta_i$$

The following constraints several of which are obtained from Roger Chang and Nanette Boyle (Boyle and Morgan, 2009; Chang et al., 2011) are often used:

1. Fluxes of all reversible reactions are left unbounded.
2. Irreversible reactions are given a lower bound of zero to preserve directionality.
3. Different environmental conditions are modeled by appropriately setting reaction flux constraints in *iRC1080*. These reactions consist of environmental exchanges, non-growth associated ATP maintenance, O₂ photoevolution, starch degradation, and light or dark-regulated enzymatic reactions (Table 4.4).
4. Constraint values are derived from published sources unless otherwise noted and imposed only under appropriate environmental conditions.
5. Minimal condition signifies a constraint that is used under all environmental conditions. The appropriate biomass reaction was set as

4.2. Flux Balance Analysis

| Metabolite | A | B | C |
|-----------------------|--------|--------|--------|
| Ex_photonVis | | | 0 lb |
| Ex_CO2 | | | 0 lb |
| EX_Oxygen(e) | -10 lb | -10 lb | -10 lb |
| EX_ac(e) | 0 lb | -10 lb | |
| EX_starch(h) | 0 both | 0 both | |
| PCHLDR | 0 both | 0 both | |
| PFKh | 0 both | 0 both | |
| G6PADHh | 0 both | 0 both | |
| G6PBDHh | 0 both | 0 both | |
| FBAh | 0 both | 0 both | |
| H2Oth | 0 ub | 0 ub | 0 lb |
| BIOMASS_Chlamy_auto | 1.00 | | |
| BIOMASS_Chlamy_hetero | | | 1.00 |
| BIOMASS_Chlamy_mixo | | 1.00 | |

Table 4.4: Constraints for different growth conditions

the objective function for optimization depending on environmental conditions.

For the list of constraints, please see below:

A(Autotrophic):light, aerobic, no acetate

B(Mixotrophic):light, aerobic with acetate

C(Heterotrophic):dark, aerobic, with acetate

In addition, GLPThi, ATPSh, BFBPh, GAPDH(nadp), MDH(nadp)hi, MDHC(nadp)hi, PPDKh, IDPh, PRUK, RBPCh, rRBCh, SBP are set to be zero flux in the heterotrophic growth condition, as there are no photosynthesis reaction in this growth condition. In the light growth conditions (autotrophic and mixotrophic), the light is assumed to have the same composition as solar light when measured from the surface of the earth. According to the literature, the conversion rate from emitted energy (Em^2s) to incident ($mmolgDW hr$) is found to be 3.83.(Costa and de Moraes, 2010)

4.2.5 Objective Function

The model is under-determined as the number of linear equations is far less than the number of unknown reaction fluxes. Therefore, additional constraints should be incorporated into FBA so as to optimize a particular cellular objective. Objective functions usually take on a linear form

$$Z = cv$$

where c denotes the coefficient for weights indicating how much each reaction (v) contributes to the objective. In practice, when only one reaction, such as biomass production, is desired for maximization or minimization, c is a vector of zeros with a value of 1 at the position of the reaction of interest. Objective functions can take on many forms, commonly used objective functions include:

Maximizing biomass: the objective is to simulate the optimal cell growth.

Minimize ATP production: the objective is to determine conditions of optimal metabolic energy efficiency.

Maximize metabolite production: this objective function has been used to determine the biochemical production capabilities of *Escherichiacoli*. In this analysis, the objective function was defined to maximize the production of a chosen metabolite or desired product (e.g: lysine or phenylalanine)

According to the literature, the *in silico* predictions of the maximizing biomass production are consistent 86% of the time for *E.coli*, and approximately 60% of the time for *Helicobacter pylori*, approximately 91% for the *E.coli* when transcriptional regulation was accounted for (Ibarra et al., 2002)(Edwards et al., 2001).

Biomass Objective Function for C. Reinhardtii The biomass formation equations used for Flux Balance Analysis were derived according to previous methods (Chavali et al., 2008). The idea is to estimate the proportion of dry weight biomass composed of protein, DNA, RNA, carbohydrate, fatty acid, glycerol, lipids, chlorophyll, etc., using available literature. At first, concentration of DNA, RNA, retinal, chlorophyll and xanthophylls in the cell have been found in the literature to be about 0.40% (Valle et al., 1981), 11.1%, 0.00002795%(Beckmann and Hegemann, 1991), 2.4% and 0.37%(Niyogi, 1997).

Then composition of the remaining cellular components was estimated from previously published data, components reported at less than 0.1g/L are omitted, the remaining components (carbohydrates, including starch;

glycerol; lipid, including triglyceride; protein; and volatile fatty acids, representing the sum of acetic, propionic, butyric, and valeric acids) are obtained from R.Chang in UCSD.

Finally, the data above are integrated into different full biomass equations for each growth condition. All the values are converted into $mmol/gDW$

The biomass function for 3 different growth conditions can be found in the Appendix 6.

4.2.6 Linear Program Solver

Linear programming is used to find the optimal solution derived from the objective function within the space defined by the mass balance equations and reaction bounds and other constraints. Due to the under-determined nature of the stoichiometric equations, the solution to the above optimization problem maybe non-unique (i.e, the optimal solution lies along an edge, plane, or hyperplane, rather than simply lying at a vertex); thus, several different sets of fluxes may achieve the same optimal objective. Please see Figure 2.1 for Linear Programming.)

In general, lots of computational tools can be implemented to solve the LP problem that arises in FBA, even for large-scale systems.

4.2.7 Identification of Essential Metabolites

With the flux distribution obtained from the initial Flux Balance Analysis, essential metabolites are distinguished from a total of 1215 metabolites. The metabolite essentiality can be found by metabolite knock-out analysis, which is defined as the phenotypic effect on cell growth when the consumption rate of a given metabolite M is set to zero. Only fluxes producing M are allowed, so the constraints are applicable to all the outgoing fluxes that are set to zero. The essentiality of metabolite is defined by the change in scale of cell growth rate compared to the growth rate of wild type,

$$ME = (Basegrowth - Optimal Growth) / Base Growth$$

In this study, an essential metabolite is recognized when its absence leads to decrease in cell growth rate that is at least half of that of the wild type, which means, $ME > 90\%$. We calculated the elimination caused by the reduction of the flux of each metabolite to zero. With the model *iRC1080*, which creatively contains metabolic light usage, we can simulate the growth in three different conditions. The growth conditions includes:

Condition A (Autotrophic) : light, aerobic, no acetate, biomass as objective function.

Condition B (Mixotrophic): light, aerobic, with acetate, biomass as objective function.

Condition C (Heterotrophic): dark, aerobic, with acetate, biomass as objective function.

The same metabolite could exist in seven different compartments in this model, including cytosol, chloroplast, mitochondria, glyoxysome, flagellum, nucleus and extra-cellular. The metabolite essentiality are calculated separately in different compartments. In other words, if a metabolite participates in reactions in different compartments, the flux of that particular metabolite is treated as two different fluxes in their respective compartments. When it comes to analyzing the overall metabolite essentiality, we ignore the compartment difference, it is recognized to be essential as long as it is found to be essential in any one of the compartments.

There are 1215 metabolites in total in C.R, in model *iRC1080*. Among all the 1215 metabolites, 426 are found to be essential in Condition A , and 247 are found to be essential in Condition B , while 260 in Condition C , this demonstrates for different growth conditions, the microalgae use different metabolite pathways to fulfill the basic growth requirements. 189 metabolites show essentiality in all 3 growth conditions (Appendix 5), 38 metabolites are found to be essential in 2 growth conditions, 419 metabolites show essentiality in 1 growth conditions. Less than 15% of metabolites are found to be essential in all three growth conditions, this might be because of the high robustness of biosystems because in different growth conditions different pathways are activated to ensure cell growth.

Although essential metabolites have been identified, it is not yet clear if all the essential metabolites exert the same influence on the biological system. We are going to categorize essential metabolites by Flux Sum Analysis, to better understand how essential metabolites influence the total growth rate of biological systems.

4.3 Flux Sum Analysis

A new variable "flux-sum" is introduced by Bevan Kai Sheng Chung and Dong-yup Lee in 2009 (Chung and Lee, 2009) to describe the absolute rate of consumption and production of each metabolite. For a steady state system, which is also the fundamental assumption of Flux Balance Analysis, flux-sum Φ_i of the metabolite i can be derived from summing up all the incoming and outgoing fluxes around the metabolite (Kim et al., 2007):

$$\Phi_i = \sum_{j \in P_i} S_{ij} v_j = - \sum_{j \in C_i} S_{ij} v_j = \frac{1}{2} \left| \sum_j S_{ij} v_j \right|$$

where S_{ij} is the stoichiometric matrix, and V_j is the flux of reaction j . P_i denotes the set of reactions producing metabolite i , while C_i represents the set of reactions consuming metabolite i . For a system in steady state, in order to maintain a constant concentration of a certain metabolite, the sum of outgoing fluxes should be equal to the sum of incoming fluxes.

Flux sum analysis is known for its capability to help study the differences among essential metabolites, a two-step approach is employed to carry out the flux sum attenuation.

4.3.1 Procedure for Flux Sum Analysis

Step 1 : Evaluate basal flux-sum distribution The wild-type flux distribution is defined as the flux distribution in the wild-type metabolite model (without changing any elements of the mathematic model.) The

4.3. Flux Sum Analysis

basal flux-sum distribution is calculated from the wild-type flux distribution out of FBA, under unperturbed condition. In this case, 3 different growth conditions are simulated, respectively.

$$\max v_{biomass}$$

$$s.t$$

$$\sum_j S_{ij}v_j = 0$$

$$\alpha_j \leq v_j \leq \beta_j$$

The basal flux-sum distribution for metabolite i is achieved after solving the above linear programming question:

$$\Phi_i^B = \frac{1}{2} \left| \sum_j S_{ij}v_j \right|$$

The basal flux-sum distribution for *Chlamydomonas* is listed in the Appendix V. We calculate the total basal flux-sum for the systems in 3 different growth conditions same as Flux Balance Analysis. The total basal flux-sum for mixotrophic growth (with light, with acetate) is found to be larger than other 2 growth conditions. This result is consistent with current studies.

The total basal flux-sum for all the Universal Metabolites are also calculated, it's found that the Universal Metabolites contributes to a very large percentage of the system flux-sum (about 80% - 85%)(Figure 4.5).

It is also noticed that the probability of the basal flux-sum generically

4.3. Flux Sum Analysis

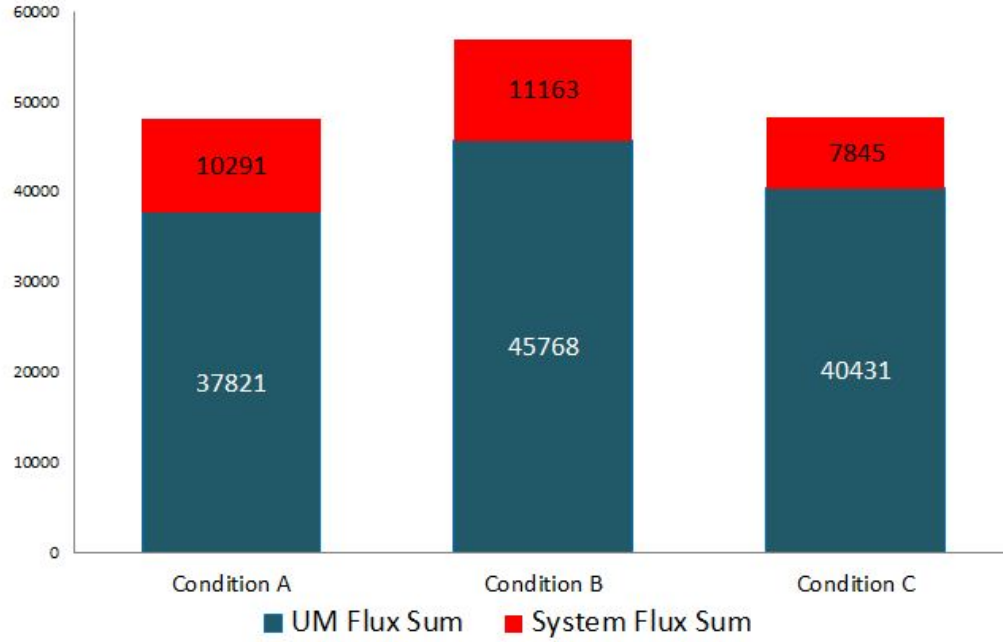


Figure 4.5: The total basal flux-sum for *C.Reinhardtii* in 3 different conditions. The blue part represents the total basal flux-sum for Universal Metabolites.

follows an exponential distribution (as shown in Figure 4.6).

$$y = e^{a+bx+cx^2}$$

y is the probability of a metabolite with basal flux-sum of 10^x . With R^2 larger than 0.99.

Step 2 : Manipulate flux-sum by attenuation Flux-sum of each metabolite is manipulated to evaluate the corresponding metabolite essen-

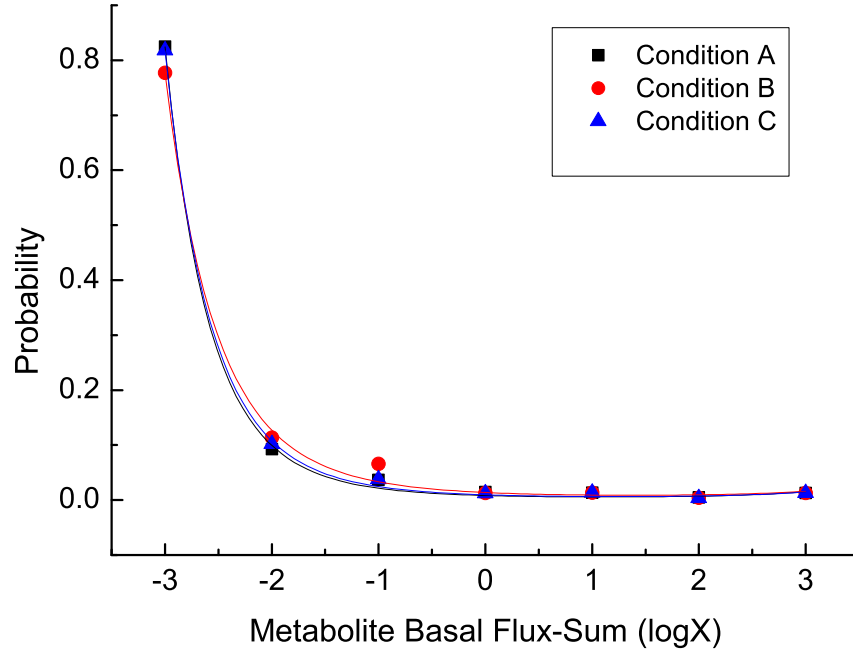


Figure 4.6: Probability distribution of metabolites with certain basal flux-sum.

tiality: the basal flux-sum is considered as a starting point, followed by examining the effects of decreasing the metabolite flux-sum. Same as above, we simulated 3 different growth conditions for each metabolite.

$$\max v_{biomass}$$

$$s.t$$

$$\frac{1}{2} \left| \sum_j S_{ij} v_j \right| \leq k_{att} \Phi_i^B$$

4.3. Flux Sum Analysis

$$\sum_j S_{ij}v_j = 0$$
$$\alpha_j \leq v_j \leq \beta_j$$

Biomass production values for different levels of flux-sum attenuation can be obtained by solving this LP problem. k_{att} control the levels of attenuation of the flux-sum, we set $k_{att} = 1$ initially and then decrease the value of it until $k_{att} = 0$.

While essential metabolites are usually associated with lethal reactions, 3 different types of essential metabolites are determined through the flux-sum attenuation analysis according to the curve trend when we manipulate the flux-sum of different metabolites in Figure 4.7.

Type AE: the most common essential metabolites found in the metabolite network, the biomass production rate varies linearly to the flux-sum of the metabolite.

Type BE: these type of metabolites are attributed to the existence of alternate optimal solutions, which also demonstrates the highly robustness of the bio-system, a small reduction of flux-sum can be compensated by other "equivalent" fluxes.

Type CE: these metabolites showed a rapid drop when the flux-sum was attenuated and reach the 0 flux earlier than other essential metabolites. With a relatively high threshold, the organism would not be able to produce any biomass under the threshold. These metabolites were found to be involved in non-growth associated maintenance.

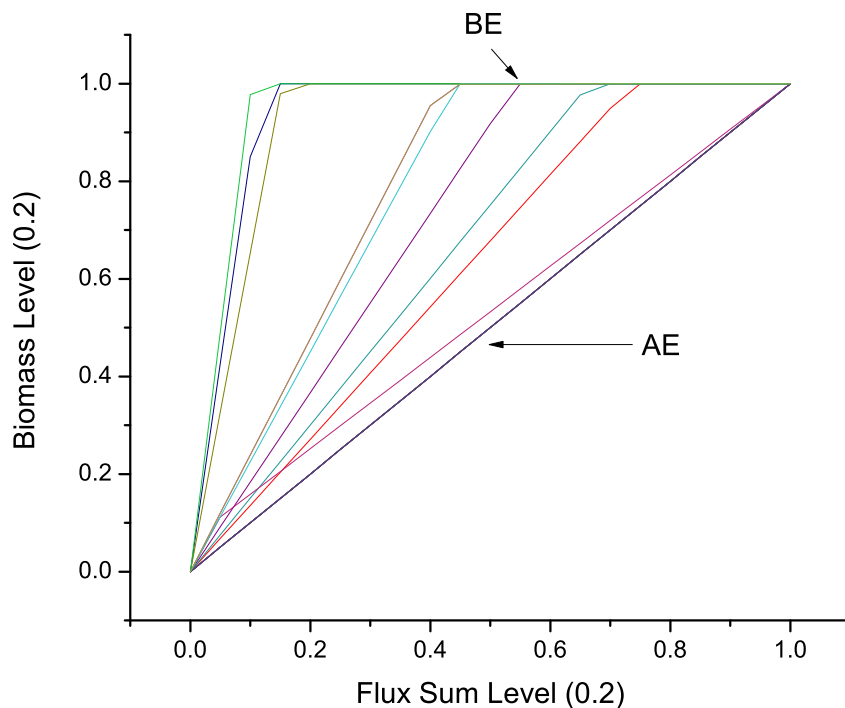


Figure 4.7: 2 types of essential metabolites: Type AE and Type BE

With the model *iNJ1080* for *C.Reinhardtii*, we carried out Flux Sum Attenuation Analysis to study the type of all the essential metabolites in 3 different growth conditions. The table below show the number of different type of essential metabolites in different growth conditions.

We could see from Table 4.5 and in Figure 4.8 that here are much more Type A essential metabolites than Type B essential metabolites, and very a few Type C metabolites. The two essential types, AE and CE, may serve as promising drug targets since the attenuation of their flux-sum will lead

4.3. Flux Sum Analysis

| | Lna | Lwac | Da |
|---------|-----|------|-----|
| Type AE | 301 | 182 | 179 |
| Type BE | 122 | 65 | 79 |
| Type CE | 3 | 1 | 2 |
| Total | 426 | 248 | 260 |

Table 4.5: Number of different types of essential metabolites in different growth conditions

to significant reduction in cell growth.

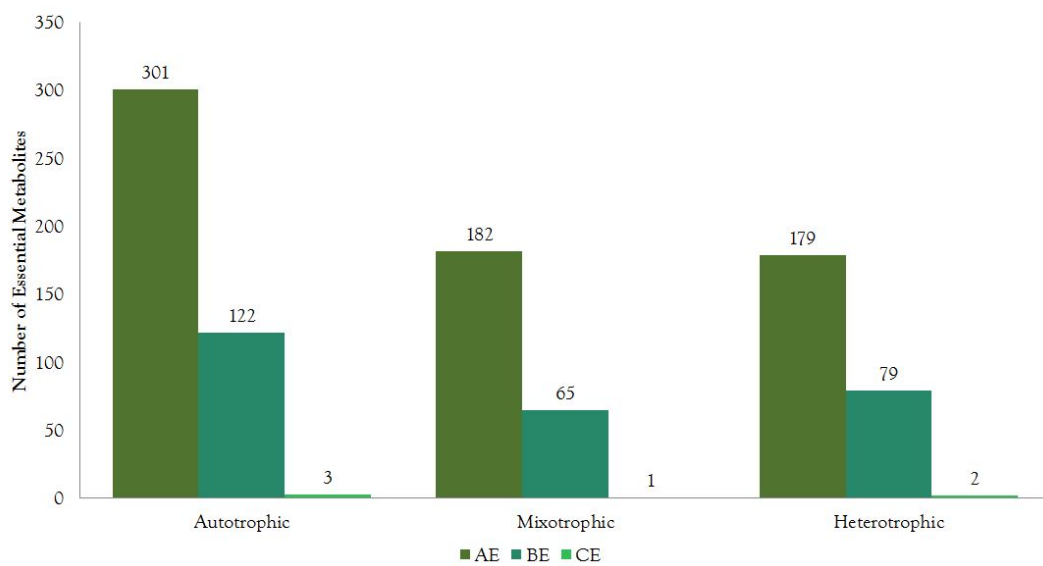


Figure 4.8: Number of different type of essential metabolites in different growth conditions

Biological Discussion

The result shows great consistency with B. Chung's hypothesis that most of the essential metabolites in the cell are type AE (Chung and Lee, 2009).

There are 189 metabolites found to be essential in all three different kind of growth conditions, it demonstrated the high robustness of the biological systems. In different growth conditions, the microalgae will change the metabolite pathway to meet the living requirements. We have found that in autotrophic condition, photosynthesis, porphyrin and chlorophyll metabolism, phenylalanine, tyrosine, and tryptophan biosynthesis were the most essential subsystems, and had most of the essential metabolites. While for mixotrophic condition, phenylalanine, tyrosine, tryptophan biosynthesis, porphyrin and chlorophyll metabolite pathways showed more essentiality than other pathways.

When the simulation is running under the heterotrophic condition, in the dark environment with acetate, photosynthesis pathway does not show essentiality any more. Instead, glycolysis, starch metabolism, amino acids, chlorophyll, and nucleotides still make up a high proportion of required metabolites.

Expectedly, the fact that most of the essential metabolites are Type AE, demonstrates that most of the essential metabolites contribute crucially to the cell growth without any substitute. However, there are still some essential metabolites (BE) that can find an alternative pathway to sustain cell growth for a short period of time.

4.3.2 Conclusion

In this chapter, we implement Flux Balance Analysis as the constraint based modeling tool to identify the essential metabolites, the constraints and biomass formation are conducted from literatures and other resources. 183 metabolites are found to be essential in all 3 growth conditions. This is also the first comprehensive essential metabolites list for *C. Reinhardtii* under all 3 growth conditions. By using Flux Sum Analysis, we categorized all the essential metabolites into 3 different types according to the type of impact when the total flux of a certain metabolite is decreasing. We found that Type AE is the most common essential metabolites.

This study reveals that most of the essential metabolites exert equally influence on the cell growth.

Chapter 5

Conclusion

Understanding and identifying the essential metabolites is important as their absence leads to cell death. The main objective of this study is to identify the metabolite essentiality through two different approaches: an interaction-based and a constraints-based.

In the interaction-based approach, a latest model with essential metabolites from Lamichhane et al. (2011) for *Mycobacterium tuberculosis* is used to study the correlations between metabolite essentiality and the metabolite network topology. The metabolite degree, the degree of neighbors, the clustering coefficient of each metabolite, and the betweenness of the metabolite network is calculated, separately. Based on the statistical tests, we found that the metabolite essentiality is significantly related to the topological characteristics. The essential metabolites usually have larger degree, larger sum of neighbors' degree and smaller shortest path and the essential lite metabolites have smaller clustering coefficient.

In the constraint-based approach, Flux Balance Analysis (known as FBA) is implemented on the most advanced in-silico model of *C. Reinhardtii*, which contains light usage reactions to make it possible to predict essential

metabolites in 3 different growth environments: autotrophic, mixotrophic, and heterotrophic. 403, 223 and 206 essential metabolites were found in these three growth conditions. Flux Sum Analysis is used afterward to classify the essential metabolites, it's found that most of the essential metabolites are Type A, and the distribution of flux sum for all the metabolites tends to follow an exponential distribution and essential metabolites are likely to have larger flux sum.

This work provides a good understanding of essential metabolites through two different approaches. Future work could focus on

- experimental validation, to illustrate the prediction of essential metabolites in *C. Reinhardtii*, the list of essential metabolites can be obtained through gene-knockout experiments.
- further study of the correlations between metabolite topology and metabolite essentiality in more model organisms.
- incorporating dynamic flux balance analysis(DFBA) to predict essential metabolites.
- implement these approaches on one same organism to find out the correlations between the two different approaches.

Bibliography

Aittokallio, T. and Schwikowski, B. (2006a). Graph-based methods for analysing networks in cell biology. *Briefings in bioinformatics*, 7(3):243–55.

Aittokallio, T. and Schwikowski, B. (2006b). Graph-based methods for analysing networks in cell biology. *Briefings in bioinformatics*, 7(3):243–55.

Albert, R., Jeong, H., and Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382.

Beard, D., Liang, S., and Qian, H. (2002). Energy Balance for Analysis of Complex Metabolic Networks. *Biophysical Journal*, 83(1):79–86.

Becker, S. a., Feist, A. M., Mo, M. L., Hannum, G., Palsson, B. O., and Herrgard, M. J. (2007). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nature protocols*, 2(3):727–38.

Beckmann, M. and Hegemann, P. (1991). In vitro identification of rhodopsin in the green alga *Chlamydomonas*. *Biochemistry*, 30(15):3692–3697.

Beer, L. L., Boyd, E. S., Peters, J. W., and Posewitz, M. C. (2009). Engi-

- neering algae for biohydrogen and biofuel production. *Current opinion in biotechnology*, 20(3):264–71.
- Bermingham, A. and Derrick, J. P. (2002). The folic acid biosynthesis pathway in bacteria: evaluation of potential for antibacterial drug discovery. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 24(7):637–48.
- Boer, P. H., Bonen, L., Lee, R. W., and Gray, M. W. (1985). Genes for respiratory chain proteins and ribosomal RNAs are present on a 16-kilobase-pair DNA species from *Chlamydomonas reinhardtii* mitochondria. *PNAS*, 82(10):3340–3344.
- Boyle, N. R. and Morgan, J. a. (2009). Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*. *BMC systems biology*, 3:4.
- Brennan, L. and Owende, P. (2010). Biofuels from microalgaeA review of technologies for production, processing, and extractions of biofuels and co-products. *Renewable and Sustainable Energy Reviews*, 14(2):557–577.
- Bro, C., Regenber, B., Förster, J., and Nielsen, J. (2006). In silico aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. *Metabolic engineering*, 8(2):102–11.
- Caspi, R., Altman, T., Dale, J. M., Dreher, K., Fulcher, C. a., Gilham, F., Kaipa, P., Karthikeyan, A. S., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. a., Paley, S., Popescu, L., Pujar, A., Shearer, A. G., Zhang, P., and Karp, P. D. (2010). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic acids research*, 38(Database issue):D473–9.

- Chang, R. L., Ghamsari, L., Manichaikul, A., Hom, E. F. Y., Balaji, S., Fu, W., Shen, Y., Hao, T., Palsson, B. O., Salehi-Ashtiani, K., and Papin, J. a. (2011). Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. *Molecular Systems Biology*, 7(518).
- Chavali, A. K., Whittemore, J. D., Eddy, J. A., Williams, K. T., and Papin, J. A. (2008). Systems analysis of metabolism in the pathogenic trypanosomatid *Leishmania major*. *Molecular systems biology*, 4(1):177.
- Chisti, Y. (2007). Biodiesel from microalgae. *Biotechnology advances*, 25(3):294–306.
- Chung, B. K. S. and Lee, D.-Y. (2009). Flux-sum analysis: a metabolite-centric approach for understanding the metabolic network.
- Cole, S., Brosch, R., Parkhill, J., and Garnier, T. (1998a). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 396(NOVEMBER).
- Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S., and Barrell, B. G. (1998b). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 393(6685):537–44.

- Costa, J. A. V. and de Morais, M. G. (2010). The role of biochemical engineering in the production of biofuels from microalgae. *Bioresource technology*, 102(1):9–2.
- Coulomb, S., Bauer, M., Bernard, D., and Marsolier-Kergoat, M.-C. (2005). Gene essentiality and the topology of protein interaction networks. *Proceedings. Biological sciences / The Royal Society*, 272(1573):1721–5.
- Degen, J. (2001). A novel airlift photobioreactor with baffles for improved light utilization through the flashing light effect. *Journal of Biotechnology*, 92(2):89–94.
- Duarte, N. C., Herrgård, M. J., and Palsson, B. O. (2004). Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome research*, 14(7):1298–309.
- Edwards, J. S. (2000). The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences*, 97(10):5528–5533.
- Edwards, J. S., Ibarra, R. U., and Palsson, B. O. (2001). In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature biotechnology*, 19(2):125–30.
- Feist, A. M., Herrgård, M. J., Thiele, I., Reed, J. L., and Palsson, B. O. (2009). Reconstruction of biochemical networks in microorganisms. *Nature reviews. Microbiology*, 7(2):129–43.
- Francke, C., Siezen, R. J., and Teusink, B. (2005). Reconstructing the metabolic network of a bacterium from its genome. *Trends in microbiology*, 13(11):550–8.

- Gevorgyan, A., Bushell, M. E., Avignone-Rossa, C., and Kierzek, A. M. (2010). SurreyFBA: A command line tool and graphics user interface for constraint based modelling of genome scale metabolic reaction networks. *Bioinformatics (Oxford, England)*, pages 1–2.
- Ghim, C.-M., Goh, K.-I., and Kahng, B. (2005). Lethality and synthetic lethality in the genome-wide metabolic network of Escherichia coli. *Journal of theoretical biology*, 237(4):401–11.
- Girvan, M. and Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12):7821–6.
- Grafahrend-Belau, E., Klukas, C., Junker, B. H., and Schreiber, F. (2009). FBA-SimVis: interactive visualization of constraint-based metabolic models. *Bioinformatics (Oxford, England)*, 25(20):2755–7.
- Hahn, J. J., Ghirardi, M. L., and Jacoby, W. a. (2004). Effect of process variables on photosynthetic algal hydrogen production. *Biotechnology progress*, 20(3):989–91.
- Hahn, M. W. and Kern, A. D. (2005). Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Molecular biology and evolution*, 22(4):803–6.
- Harris, E. H. (2001). CHLAMYDOMONAS AS A MODEL ORGANISM. *Annual review of plant physiology and plant molecular biology*, 52(1):363–406.
- Hatzimanikatis, V., Li, C., Ionita, J. a., Henry, C. S., Jankowski, M. D.,

- and Broadbelt, L. J. (2005). Exploring the diversity of complex metabolic networks. *Bioinformatics (Oxford, England)*, 21(8):1603–9.
- He, X. and Zhang, J. (2006). Why do hubs tend to be essential in protein networks? *PLoS genetics*, 2(6):e88.
- Hjersted, J. L. and Henson, M. a. (2009). Steady-state and dynamic flux balance analysis of ethanol production by *Saccharomyces cerevisiae*. *IET systems biology*, 3(3):167–79.
- Hucka, M. (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531.
- Ibarra, R. U., Edwards, J. S., and Palsson, B. O. (2002). *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*, 420(6912):186–9.
- Imieliski, M., Belta, C., Halász, A., and Rubin, H. (2005). Investigating metabolite essentiality through genome-scale analysis of *Escherichia coli* production capabilities. *Bioinformatics (Oxford, England)*, 21(9):2008–16.
- Jamshidi, N. and Palsson, B. O. (2007). Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the in silico strain iNJ661 and proposing alternative drug targets. *BMC systems biology*, 1:26.
- Jeong, H., Mason, S. P., Barabási, A. L., and Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, 411(6833):41–2.

- Jeong, H., Oltvai, Z. N., and Barabási, A.-L. (2003). Prediction of Protein Essentiality Based on Genomic Data. *Complexus*, 1(1):19–28.
- Jiang, H., Patwardhan, R., and Shah, S. L. (2009). Root cause diagnosis of plant-wide oscillations using the concept of adjacency matrix. *Journal of Process Control*, 19(8):1347–1354.
- Kauffman, K. J., Prakash, P., and Edwards, J. S. (2003). Advances in flux balance analysis. *Current Opinion in Biotechnology*, 14(5):491–496.
- Kim, P.-J., Lee, D.-Y., Kim, T. Y., Lee, K. H., Jeong, H., Lee, S. Y., and Park, S. (2007). Metabolite essentiality elucidates robustness of *Escherichia coli* metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, 104(34):13638–42.
- Kim, T. Y., Sohn, S. B., Kim, H. U., and Lee, S. Y. (2008). Strategies for systems-level metabolic engineering. *Biotechnology journal*, 3(5):612–23.
- Kitano, H. (2002). Systems biology: a brief overview. *Science (New York, N.Y.)*, 295(5560):1662–4.
- Krieger, C. J., Zhang, P., Mueller, L. a., Wang, A., Paley, S., Arnaud, M., Pick, J., Rhee, S. Y., and Karp, P. D. (2004). MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic acids research*, 32(Database issue):D438–42.
- Lamichhane, G., Freundlich, J., Ekins, S., Wickramaratne, N., Nolan, S., and Bishai, W. (2011). Essential Metabolites of *Mycobacterium tuberculosis* and Their Mimics. *Mbio*, 2(1):1–10.
- Lee, J. M., Gianchandani, E. P., and Papin, J. a. (2006). Flux balance

analysis in the era of metabolomics. *Briefings in bioinformatics*, 7(2):140–50.

Li, Y., Han, D., Hu, G., Sommerfeld, M., and Hu, Q. (2010). Inhibition of starch synthesis results in overproduction of lipids in *Chlamydomonas reinhardtii*. *Biotechnology and bioengineering*, 107(2):258–268.

Li, Z., Wang, R.-S., and Zhang, X.-S. (2011). Two-stage flux balance analysis of metabolic networks for drug target identification. *BMC systems biology*, 5 Suppl 1(Suppl 1):S11.

Mahadevan, R. and Palsson, B. O. (2005). Properties of metabolic networks: structure versus function. *Biophysical journal*, 88(1):L07–9.

Martelli, C., De Martino, A., Marinari, E., Marsili, M., and Pérez Castillo, I. (2009). Identifying essential genes in *Escherichia coli* from a metabolic optimization principle. *Proceedings of the National Academy of Sciences of the United States of America*, 106(8):2607–11.

Mason, O. and Verwoerd, M. (2007). Graph theory and networks in Biology. *Engineering and Technology*.

Maul, J. E., Lilly, J. W., Cui, L., DePamphilis, C. W., Miller, W., Harris, E. H., and Stern, D. B. (2002). The *Chlamydomonas reinhardtii* Plastid Chromosome: Islands of Genes in a Sea of Repeats. *PLANT CELL*, 14(11):2659–2679.

May, P., Wienkoop, S., Kempa, S., Usadel, B., Christian, N., Rupprecht, J., Weiss, J., Recuenco-Munoz, L., Ebenhöf, O., Weckwerth, W., and Walther, D. (2008). Metabolomics- and proteomics-assisted genome an-

notation and analysis of the draft metabolic network of *Chlamydomonas reinhardtii*. *Genetics*, 179(1):157–66.

Meadows, A. L., Karnik, R., Lam, H., Forestell, S., and Snedecor, B. (2010). Application of dynamic flux balance analysis to an industrial *Escherichia coli* fermentation. *Metabolic engineering*, 12(2):150–60.

Melis, A. and Happe, T. (2004). Trails of green alga hydrogen research - from *hans gaffron* to new frontiers. *Photosynthesis research*, 80(1-3):401–9.

Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., Terry, A., Salamov, A., Fritz-Laylin, L. K., Maréchal-Drouard, L., Marshall, W. F., Qu, L.-H., Nelson, D. R., Sanderfoot, A. A., Spalding, M. H., Kapitonov, V. V., Ren, Q., Ferris, P., Lindquist, E., Shapiro, H., Lucas, S. M., Grimwood, J., Schmutz, J., Cardol, P., Cerutti, H., Chanfreau, G., Chen, C.-L., Cognat, V., Croft, M. T., Dent, R., Dutcher, S., Fernández, E., Fukuzawa, H., González-Ballester, D., González-Halphen, D., Hallmann, A., Hanikenne, M., Hippler, M., Inwood, W., Jabbari, K., Kalanon, M., Kuras, R., Lefebvre, P. A., Lemaire, S. D., Lobanov, A. V., Lohr, M., Manuell, A., Meier, I., Mets, L., Mittag, M., Mittelmeier, T., Moroney, J. V., Moseley, J., Napoli, C., Nedelcu, A. M., Niyogi, K., Novoselov, S. V., Paulsen, I. T., Pazour, G., Purton, S., Ral, J.-P., Riaño Pachón, D. M., Riekhof, W., Rymarquis, L., Schroda, M., Stern, D., Umen, J., Willows, R., Wilson, N., Zimmer, S. L., Allmer, J., Balk, J., Bisova, K., Chen, C.-J., Elias, M., Gendler, K., Hauser, C., Lamb, M. R., Ledford, H., Long, J. C., Minagawa, J., Page, M. D., Pan, J., Pootakham, W., Roje, S., Rose, A., Stahlberg, E., Terauchi, A. M., Yang, P., Ball, S., Bowler, C., Dieckmann, C. L., Gladyshev, V. N., Green,

- P., Jorgensen, R., Mayfield, S., Mueller-Roeber, B., Rajamani, S., Sayre, R. T., Brokstein, P., Dubchak, I., Goodstein, D., Hornick, L., Huang, Y. W., Jhaveri, J., Luo, Y., Martínez, D., Ngau, W. C. A., Otilar, B., Poliakov, A., Porter, A., Szajkowski, L., Werner, G., Zhou, K., Grigoriev, I. V., Rokhsar, D. S., and Grossman, A. R. (2007). The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science (New York, N.Y.)*, 318(5848):245–50.
- Metting, F. B. (1996). Biodiversity and application of microalgae. *Journal of Industrial Microbiology & Biotechnology*, 17(5-6):477–489.
- Morowvat, M. H., Rasoul-Amini, S., and Ghasemi, Y. (2010). *Chlamydomonas* as a "new" organism for biodiesel production. *Bioresource technology*, 101(6):2059–62.
- Niyogi, K. K. (1997). The roles of specific xanthophylls in photoprotection. *Proceedings of the National Academy of Sciences*, 94(25):14162–14167.
- Oh, Y.-K., Palsson, B. O., Park, S. M., Schilling, C. H., and Mahadevan, R. (2007). Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *The Journal of biological chemistry*, 282(39):28791–9.
- Orth, J. D. and Palsson, B. O. (2010). Systematizing the generation of missing metabolic knowledge. *Biotechnology and bioengineering*, 107(3):403–12.
- Orth, J. D., Thiele, I., and Palsson, B. (2010). What is flux balance analysis? *Nature biotechnology*, 28(3):245–8.
- Palsson, B. (2003). Flux-balance analysis : Basic concepts. *Systems Biology*.

- Palsson, B. (2009). Metabolic systems biology. *FEBS letters*, 583(24):3900–4.
- Price, N. D. and Lee, S. Y. (2010). Editorial: Systems biology for biotech applications. *Biotechnology journal*, 5(7):636–7.
- Reed, J. L., Patel, T. R., Chen, K. H., Joyce, A. R., Applebee, M. K., Herring, C. D., Bui, O. T., Knight, E. M., Fong, S. S., and Palsson, B. O. (2006). Systems approach to refining genome annotation. *Proceedings of the National Academy of Sciences of the United States of America*, 103(46):17480–4.
- Rupprecht, J. (2009). From systems biology to fuel—*Chlamydomonas reinhardtii* as a model for a systems biology approach to improve biohydrogen production. *Journal of biotechnology*, 142(1):10–20.
- Rupprecht, J., Hankamer, B., Mussnug, J. H., Ananyev, G., Dismukes, C., and Kruse, O. (2006). Perspectives and advances of biological H₂ production in microorganisms. *Applied microbiology and biotechnology*, 72(3):442–9.
- Samal, A., Singh, S., Giri, V., Krishna, S., Raghuram, N., and Jain, S. (2006). Low degree metabolites explain essential reactions and enhance modularity in biological networks. *BMC bioinformatics*, 7:118.
- Satish Kumar, V., Dasika, M. S., and Maranas, C. D. (2007). Optimization based automated curation of metabolic reconstructions. *BMC bioinformatics*, 8:212.
- Schenk, P. M., Thomas-Hall, S. R., Stephens, E., Marx, U. C., Mussnug, J. H., Posten, C., Kruse, O., and Hankamer, B. (2008). Second Generation

- Biofuels: High-Efficiency Microalgae for Biodiesel Production. *BioEnergy Research*, 1(1):20–43.
- Schmidt, B. J., Lin-Schmidt, X., Chamberlin, A., Salehi-Ashtiani, K., and Papin, J. a. (2010). Metabolic systems analysis to advance algal biotechnology. *Biotechnology journal*, 5(7):660–70.
- Smith, L. P., Bergmann, F. T., Chandran, D., and Sauro, H. M. (2009). Antimony: a modular model definition language. *Bioinformatics (Oxford, England)*, 25(18):2452–4.
- Spolaore, P., Joannis-Cassan, C., Duran, E., and Isambert, A. (2006). Commercial applications of microalgae. *Journal of bioscience and bioengineering*, 101(2):87–96.
- Vahrenholz, C., Riemen, G., Pratje, E., Dujon, B., and Michaelis, G. (1993). Mitochondrial DNA of *Chlamydomonas reinhardtii*: the structure of the ends of the linear 15.8-kb genome suggests mechanisms for DNA replication.
- Valle, O., Lien, T., and Knutsen, G. (1981). Fluorometric determination of DNA and RNA in *Chlamydomonas* using ethidium bromide. *Journal of Biochemical and Biophysical Methods*, 4(5-6):271–277.
- Yu, H., Greenbaum, D., Xin Lu, H., Zhu, X., and Gerstein, M. (2004). Genomic analysis of essentiality within protein networks. *Trends in genetics : TIG*, 20(6):227–31.
- Yu, H., Kim, P. M., Sprecher, E., Trifonov, V., and Gerstein, M. (2007). The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS computational biology*, 3(4):e59.

Bibliography

Zotenko, E., Mestre, J., O’Leary, D. P., and Przytycka, T. M. (2008). Why do hubs in the yeast protein interaction network tend to be essential: re-examining the connection between the network topology and essentiality. *PLoS computational biology*, 4(8):e1000140.

Appendix

A.1 Appendix 1: ELM in Mycobacterium Tuberculosis

| No. | Abbrev. | Essential Metabolite Name |
|-----|---------|---|
| 1 | 23dhdp | 2,3-Dihydrodipicolinate |
| 2 | 26dap-M | meso-2,6-Diaminoheptanedioate |
| 3 | 3dhq | 3-Dehydroquate |
| 4 | 3dhsk | 3-Dehydroshikimate |
| 5 | 3mob | 3-Methyl-2-oxobutanoate |
| 6 | 3psme | 5-O-(1-Carboxyvinyl)-3-phosphoshikimate |
| 7 | 5aop | 5-Amino-4-oxopentanoate |
| 8 | alaala | D-Alanyl-D-alanine |
| 9 | chor | chorismate |

A.2. Appendix 2: Universal Metabolites

| | | |
|----|---------|--|
| 10 | glu-L | L-Glutamate |
| 11 | glu1sa | L-Glutamate 1-semialdehyde |
| 12 | hmbil | Hydroxymethylbilane |
| 13 | ppbng | Porphobilinogen |
| 14 | skm5p | Shikimate 5-phosphate |
| 15 | sl2a6o | N-Succinyl-2-L-amino-6-oxoheptanedioate |
| 16 | uaagmda | Undecaprenyl-diphospho-N-acetylmuramoyl- (N-acetylglucosamine)-L-ala-D-glu-meso-2,6-diaminopimeloyl-D-ala-D-ala |
| 17 | uaccg | UDP-N-acetyl-3-O-(1-carboxyvinyl)-D-glucosamine |
| 18 | ugmda | UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6- diaminopimeloyl-D-alanyl-D-alanine |

A.2 Appendix 2: Universal Metabolites

| No. | Abbrev. | Universal Metabolite Name |
|-----|---------|---------------------------|
| 1 | utp | UTP |
| 2 | ump | UMP |

A.2. Appendix 2: Universal Metabolites

| | | |
|----|-------|---|
| 3 | udp | UDP |
| 4 | tyr-L | L-Tyrosine |
| 5 | trdrd | Reduced thioredoxin |
| 6 | trdox | Oxidized thioredoxin |
| 7 | thf | 5,6,7,8-Tetrahydrofolate |
| 8 | ser-L | L-Serine |
| 9 | pyr | Pyruvate |
| 10 | pi | Phosphate |
| 11 | phe-L | L-Phenylalanine |
| 12 | nadph | Nicotinamide adenine dinucleotide phosphate - reduced |
| 13 | nadp | Nicotinamide adenine dinucleotide phosphate |
| 14 | nadh | Nicotinamide adenine dinucleotide - reduced |
| 15 | nad | Nicotinamide adenine dinucleotide |
| 16 | mlthf | 5,10-Methylenetetrahydrofolate |
| 17 | his-L | L-Histidine |
| 18 | h2o | H ₂ O |
| 19 | h | H ⁺ |

A.2. Appendix 2: Universal Metabolites

| | | |
|----|-------|-----------------|
| 20 | gtp | GTP |
| 21 | gly | Glycine |
| 22 | glu-L | L-Glutamate |
| 23 | gln-L | L-Glutamine |
| 24 | gdp | GDP |
| 25 | dttp | dTTP |
| 26 | dgtp | dGTP |
| 27 | dctp | dCTP |
| 28 | datp | dATP |
| 29 | ctp | CTP |
| 30 | coa | Coenzyme A |
| 31 | co2 | CO ₂ |
| 32 | atp | ATP |
| 33 | asp-L | L-Aspartate |
| 34 | amp | AMP |
| 35 | adp | ADP |
| 36 | accoa | Acetyl-CoA |

A.2. Appendix 2: Universal Metabolites

| | | |
|----|-------|----------------------------|
| 37 | val-L | L-Valine |
| 38 | trp-L | L-Tryptophan |
| 39 | thr-L | L-Threonine |
| 40 | pro-L | L-Proline |
| 41 | pep | Phosphoenolpyruvate |
| 42 | met-L | L-Methionine |
| 43 | ile-L | L-Isoleucine |
| 44 | dump | dUMP |
| 45 | dtdp | dTDP |
| 46 | cys-L | L-Cysteine |
| 47 | cmp | CMP |
| 48 | arg-L | L-Arginine |
| 49 | ala-L | L-Alanine |
| 50 | lys-L | L-Lysine |
| 51 | leu-L | L-Leucine |
| 52 | gmp | GMP |
| 53 | dhap | Dihydroxyacetone phosphate |

A.2. Appendix 2: Universal Metabolites

| | | |
|----|------------|--|
| 54 | amet | S-Adenosyl-L-methionine |
| 55 | f6p | D-Fructose 6-phosphate |
| 56 | dtmp | dTMP |
| 57 | 3pg | 3-Phospho-D-glycerate |
| 58 | ru5p-D | D-Ribulose 5-phosphate |
| 59 | 3dhsk | 3-Dehydroshikimate |
| 60 | 3dhq | 3-Dehydroquininate |
| 61 | 13dpg | 3-Phospho-D-glyceroyl phosphate |
| 62 | glyc3p | Glycerol 3-phosphate |
| 63 | fad | FAD |
| 64 | cdpc16c19g | CDPdiacylglycerol (E coli) ** |
| 65 | ACP | acyl carrier protein |
| 66 | prpp | 5-Phospho-alpha-D-ribose 1-diphosphate |
| 67 | e4p | D-Erythrose 4-phosphate |
| 68 | gam6p | D-Glucosamine 6-phosphate |
| 69 | g6p | D-Glucose 6-phosphate |
| 70 | xmp | Xanthosine 5'-phosphate |

A.2. Appendix 2: Universal Metabolites

| | | |
|----|--------|--|
| 71 | imp | IMP |
| 72 | dhpt | Dihydropteroate |
| 73 | g1p | D-Glucose 1-phosphate |
| 74 | dhf | 7,8-Dihydrofolate |
| 75 | ribflv | Riboflavin |
| 76 | o2 | O ₂ |
| 77 | oaa | Oxaloacetate |
| 78 | akg | 2-Oxoglutarate |
| 79 | aicar | 5-Amino-1-(5-Phospho-D-ribosyl)imidazole-4-carboxamide |
| 80 | 10fthf | 10-Formyltetrahydrofolate |
| 81 | dpcoa | Dephospho-CoA |
| 82 | aacoa | Acetoacetyl-ACP |
| 83 | phpyr | Phenylpyruvate |
| 84 | fmn | FMN |
| 85 | 34hpp | 3-(4-Hydroxyphenyl)pyruvate |
| 86 | 34hpp | Phosphatidylglycerophosphate (Ecoli) ** |
| 87 | hco3 | Bicarbonate |

A.3. Appendix 3: Root No-production Metabolites in iNJ661

| | | |
|----|--------|--------------------------------|
| 88 | uacgam | UDP-N-acetyl-D-glucosamine |
| 89 | tdeACP | Tetradecenoyl-ACP (n-C14:1ACP) |
| 90 | malACP | Malonyl-[acyl-carrier protein] |
| 91 | dnad | Deamino-NAD+ |
| 92 | ddca | Dodecanoyl-ACP (n-C12:0ACP) |
| 93 | 2obut | 2-Oxobutanoate |

A.3 Appendix 3: Root No-production Metabolites in iNJ661

| | | | |
|-------------|--------------|-----------|----------|
| a23dhba_c | bmn_c | xyluD_c | pmcoa_c |
| a2c25dho_c | cbi_c | fdxrd_c | ppal_c |
| a2dglcn_c | cbl1_c | fol_c | pre2_c |
| a2dr5p_c | cdpdodecg_c | glcn_c | psd5p_c |
| a2mop_c | cl_c | glutrna_c | ptcys_c |
| a2pglyc_c | clpn160190_c | glyc-R_c | pyam5p_c |
| a4h2opntn_c | cobalt2_c | lald-L_c | pydam_c |

A.4. Appendix 4: Root No-consumption Metabolites in iNJ661

| | | | |
|------------|----------|-----------|-----------|
| a5dglcn_c | coba_c | meoh_c | pydxn_c |
| a5odhf2a_c | copre2_c | mettrna_c | ru5p-L_c |
| acgam_c | copre6_c | mhpglu_c | s_c |
| achms_c | dmbzid_c | mi3p-D_c | sdhla_c |
| ad_c | dtc_c | mi4p-D_c | selcys_c |
| alpam_c | dtcOX_c | mppp9_c | seln_c |
| amob_c | dxyl_c | mshfald_c | seramp_c |
| apoACP_c | enter_c | ncam_c | thfglu_c |
| apl_c | fc1p_c | no_c | thym_c |
| aplp_c | fdxox_c | pdx5p_c | trnaala_c |
| | | | uppg1_c |

**A.4 Appendix 4: Root No-consumption
Metabolites in iNJ661**

| | | | |
|----------|----------|-----------|--------|
| a3ddgc_c | copre8_c | omdtria_c | spmd_c |
| a4hba_c | cpppg1_c | pat_c | tat_c |

A.4. Appendix 4: Root No-consumption Metabolites in iNJ661

| | | | |
|-----------------------|------------|---------------|-------------|
| a4hthr_c | crn_c | pdima_c | tmha1_c |
| a4mhetz_c | dttoX_c | peptido-EC_c | tmha2_c |
| a5mtr_c | enter_c | peptido-TB1_c | tmha3_c |
| a5odhf2a_c | etha_c | peptido-TB2_c | tmha4_c |
| Ac1PIM4_c | fmettrna_c | pg160_c | tmha5_c |
| Ac2PIM2_c | gcald_c | pg190_c | tmha6_c |
| acysbm_n_c | gdptp_c | pheme_c | triat_c |
| alatrna_c | glyb_c | PIM6_c | trnaglu_c |
| arabinanagalfragund_c | homtta_c | ptth_c | uaaAgtla_c |
| btamp_c | hpglu_c | rhcys_c | uaaGgtla_c |
| cl_c | maltpt_c | rmyc_c | uaagtmnda_c |
| cobya_c | man_c | seln_c | udpglcur_c |
| copre5_c | mcbts_c | sheme_c | ugagmda_c |
| mfrppdima_c | sl1_c | xyID_c | |

A.5 Appendix 5: Common Essential Metabolites in All 3 Growth Conditions

| | | | |
|---------------|-------------------|-------------|-------------|
| 12dmpo | argsuc | glyc3p | pgp1819Z160 |
| 1hdecg3p | aspsa | h2mb4p | phpyr |
| 1odec11eg3p | B-DASH-ara1p | h2o2 | phytfl |
| 1odec9eg3p | ca | hco3 | phyto |
| 1odecg3p | cocoa | hcys-DASH-L | pi |
| 1pyr5c | caro | hdeACP | ppa |
| 23dhdp | cbasp | hisp | ppad |
| 23dhmb | cdp12dgr18111Z160 | histd | ppbng |
| 23dhmp | cdp12dgr1819Z160 | hmppp9me | ppgpp |
| 25aics | cdpea | hom-DASH-L | pphn |
| 26dap-DASH-LL | chlida | hso3 | ppi |
| 26dap-DASH-M | chldb | imacp | pppg9 |
| 2ahbut | cmp | lyc | pq |
| 2cpr5p | coa | malcoa | pqh2 |
| 2dda7p | ctp | methf | pram |
| 2h3kmtp | cys-DASH-L | mg2 | pran |

A.5. Appendix 5: Common Essential Metabolites in All 3 Growth Conditions

| | | | |
|--------------|------------------|----------------|--------------------------|
| 2ippm | cyst-DASH-L | mgdg1819Z160 | prbamp |
| 2kmb | dcamp | mgdg1819Z1619Z | prbatp |
| 2me4p | dcaro | mi3p-DASH-D | prfp |
| 2mecdp | dghs16018111Z | mlthf | prlp |
| 34hpp | dghs1601819Z | mppp9 | prot dt |
| 3c2hmp | dghs18111Z18111Z | mppp9me | pyr |
| 3c4mop | dghs18111Z1819Z | nadp | r5p |
| 3dhq | dghs1819Z18111Z | norsp | retinal |
| 3dhsk | dghs1819Z1819Z | o2 | retinal-DASH-11-DASH-cis |
| 3hcvac11eACP | dhor-DASH-S | ocdca | s7p |
| 3hmop | dkmpp | ocdcoxa | skm |
| 3mob | dtmp | ocdcea | skm5p |
| 3ocvac11eACP | dump | octeACP | so4 |
| 3psme | dxyl5p | omppp9me | sqdg18111Z160 |
| 4c2me | eig3p | orot5p | sqdg1819Z160 |
| 4pasp | etha | pa | succ |
| 5aizc | ethamp | pa160 | thdp |

A.5. Appendix 5: Common Essential Metabolites in All 3 Growth Conditions

| | | | |
|---------|---------|----------------|-------------|
| 5aop | fdxox | pa16018111Z | thf |
| 5mdr1p | fgam | pa1601819Z | thmpp |
| 5mdru1p | fpram | pa1801819Z | trdox |
| 5mthf | fprica | pa18111Z160 | trnaglu |
| acg5p | fum | pa18111Z18111Z | trp-DASH-L |
| acg5sa | g3p | pa18111Z1819Z | tyr-DASH-L |
| acglu | gal | pa1819Z160 | udp |
| ade | gar | pa1819Z1619Z | udpg |
| adn | gcara | pa1819Z18111Z | udpgal |
| ahcys | gdptp | pa1819Z1819Z | udpsq |
| aicar | glu1sa | pacoa | udpxyl |
| amet | glu5p | pcdme | ump |
| anth | glu5sa | pep | val-DASH-L |
| aps | glutrna | pgp18111Z160 | xu5p-DASH-D |
| | | | zcaro |

A.6 Appendix 6: Biomass Function(Objective Function) for Different Growth Conditions

The biomass function for autotrophic:

Biomass =

$$\begin{aligned}
 & 273.7E^3 \cdot \text{ala-L[c]} + 150.2E^3 \cdot \text{arg-L[c]} + 67.8E^3 \cdot \text{asn-L[c]} + \\
 & 67.8E^3 \cdot \text{asp-L[c]} + 2.4E^3 \cdot \text{cys-L[c]} + 81.2E^3 \cdot \text{gln-L[c]} + \\
 & 81.2E^3 \cdot \text{glu-L[c]} + 103.0E^3 \cdot \text{gly[c]} + 1.2E^3 \cdot \text{his-L[c]} + \\
 & 32.7E^3 \cdot \text{ile-L[c]} + 82.4E^3 \cdot \text{leu-L[c]} + 18.2E^3 \cdot \text{lys-L[c]} + \\
 & 2.4E^3 \cdot \text{met-L[c]} + 33.9E^3 \cdot \text{phe-L[c]} + 47.2E^3 \cdot \text{pro-L[c]} + \\
 & 20.6E^3 \cdot \text{ser-L[c]} + 82.4E^3 \cdot \text{thr-L[c]} + 1.2E^3 \cdot \text{trp-L[c]} + \\
 & 1.2E^3 \cdot \text{tyr-L[c]} + 59.4E^3 \cdot \text{val-L[c]} + 2.2E^3 \cdot \text{datp[c]} + \\
 & 3.9E^3 \cdot \text{dctp[c]} + 3.9E^3 \cdot \text{dgt p[c]} + 2.2E^3 \cdot \text{dttp[c]} + \\
 & 58.6E^3 \cdot \text{atp[c]} + 104.2E^3 \cdot \text{ctp[c]} + 104.2E^3 \cdot \text{gtp[c]} + \\
 & 58.6E^3 \cdot \text{utp[c]} + 6.4E^3 \cdot \text{starch300[h]} + 328.4E^3 \cdot \text{man[c]} + \\
 & 524.1E^3 \cdot \text{arab-L[c]} + 697.0E^3 \cdot \text{gal[c]} + 28.4E^3 \cdot \text{mgdg1839Z12Z15Z1644Z7Z10Z13Z[h]} + \\
 & 3.2E^3 \cdot \text{mgdg1839Z12Z15Z1637Z10Z13Z[h]} + 3.2E^3 \cdot \text{mgdg1839Z12Z15Z1634Z7Z10Z[h]} + \\
 & 269.4E^6 \cdot \text{dgdg1839Z12Z15Z1644Z7Z10Z13Z[h]} + \\
 & 739.2E^6 \cdot \text{dgdg1839Z12Z15Z1637Z10Z13Z[h]} + \\
 & 739.2E^6 \cdot \text{dgdg1839Z12Z15Z1634Z7Z10Z[h]} + 74.3E^6 \cdot \text{dgts18111Z1819Z[c]} + \\
 & 74.3E^6 \cdot \text{dgts18111Z18111Z[c]} + 1.1E^3 \cdot \text{dgts1601829Z12Z[c]} +
 \end{aligned}$$

A.6. Appendix 6: Biomass Function(Objective Function) for Different Growth Conditions

$$\begin{aligned}
& 1.2E^3 \cdot \text{asqdpal819Z160}[c] + 1.2E^3 \cdot \text{asqdpal8111Z160}[c] + \\
& 1.3E^3 \cdot \text{tag16018111Z160}[c] + 1.3E^3 \cdot \text{tag1601819Z160}[c] + \\
& 1.3E^3 \cdot \text{tag1801819Z160}[c] + 1.3E^3 \cdot \text{tag18111Z18111Z160}[c] + \\
& 1.3E^3 \cdot \text{tag18111Z1819Z160}[c] + 1.3E^3 \cdot \text{tag1819Z18111Z160}[c] + 37.1E^3 \cdot \text{ac}[c] + \\
& 30.0E^3 \cdot \text{ppa}[c] + 25.3E^3 \cdot \text{but}[c] + 12.1E^3 \cdot \text{glyc}[c] + \\
& 10.1E^3 \cdot \text{chla}[u] + 16.5E^3 \cdot \text{chlb}[u] + 1.0E^6 \cdot \text{rhodopsin}[s] + \\
& 504.2E^6 \cdot \text{acarol}[h] + 100.8E^6 \cdot \text{anxan}[u] + 1.4E^3 \cdot \text{caro}[u] + \\
& 655.4E^6 \cdot \text{loroxan}[u] + 1.3E^3 \cdot \text{lut}[u] + 554.6E^6 \cdot \text{neoxan}[u] + \\
& 352.9E^6 \cdot \text{vioxan}[u] + 302.5E^6 \cdot \text{zaxan}[u] + 29.9 \cdot \text{ATP maintainance} + \\
& 2.3E^3 \cdot \text{pe1801835Z9Z12Z}[c] + 1.9E^3 \cdot \text{pail18111Z160}[c] + 258.4E^6 \cdot \text{pail1819Z160}[c]
\end{aligned}$$

The biomass function for Mixotrophic:

$$\begin{aligned}
& \text{Biomass} = \\
& 279.3E^3 \cdot \text{ala-L}[c] + 93.7E^3 \cdot \text{arg-L}[c] + 69.5E^3 \cdot \text{asn-L}[c] + \\
& 69.5E^3 \cdot \text{asp-L}[c] + 12.2E^3 \cdot \text{cys-L}[c] + 91.8E^3 \cdot \text{gln-L}[c] + \\
& 91.8E^3 \cdot \text{glu-L}[c] + 113.9E^3 \cdot \text{gly}[c] + 12.7E^3 \cdot \text{his-L}[c] + \\
& 38.0E^3 \cdot \text{ile-L}[c] + 93.0E^3 \cdot \text{leu-L}[c] + 30.6E^3 \cdot \text{lys-L}[c] + \\
& 12.7E^3 \cdot \text{met-L}[c] + 40.0E^3 \cdot \text{phe-L}[c] + 51.9E^3 \cdot \text{pro-L}[c] + \\
& 20.8E^3 \cdot \text{ser-L}[c] + 34.5E^3 \cdot \text{thr-L}[c] + 1.6E^3 \cdot \text{trp-L}[c] + \\
& 1.6E^3 \cdot \text{tyr-L}[c] + 64.3E^3 \cdot \text{val-L}[c] + 2.2E^3 \cdot \text{datp}[c] +
\end{aligned}$$

A.6. Appendix 6: Biomass Function(Objective Function) for Different Growth Conditions

$$\begin{aligned} & 3.9E^3 \cdot \text{dctp}[c] + 3.9E^3 \cdot \text{dgt p}[c] + 2.2E^3 \cdot \text{dttp}[c] + \\ & 58.6E^3 \cdot \text{atp}[c] + 104.2E^3 \cdot \text{ctp}[c] + 104.2E^3 \cdot \text{gtp}[c] + \\ & 58.6E^3 \cdot \text{utp}[c] + 6.4E^3 \cdot \text{starch300}[h] + 328.4E^3 \cdot \text{man}[c] + \\ & 524.1E^3 \cdot \text{arab-L}[c] + 697.0E^3 \cdot \text{gal}[c] + \\ & 28.4E^3 \cdot \text{mgdg1839Z12Z15Z1644Z7Z10Z13Z}[h] + \\ & 3.2E^3 \cdot \text{mgdg1839Z12Z15Z1637Z10Z13Z}[h] + \\ & 3.2E^3 \cdot \text{mgdg1839Z12Z15Z1634Z7Z10Z}[h] + \\ & 269.4E^6 \cdot \text{dgdg1839Z12Z15Z1644Z7Z10Z13Z}[h] + \\ & 739.2E^6 \cdot \text{dgdg1839Z12Z15Z1637Z10Z13Z}[h] + \\ & 739.2E^6 \cdot \text{dgdg1839Z12Z15Z1634Z7Z10Z}[h] + \\ & 74.3E^6 \cdot \text{dgts18111Z1819Z}[c] + 74.3E^6 \cdot \text{dgts18111Z18111Z}[c] + \\ & 1.1E^3 \cdot \text{dgts1601829Z12Z}[c] + 1.2E^3 \cdot \text{asqdp a1819Z160}[c] + \\ & 1.2E^3 \cdot \text{asqdp a18111Z160}[c] + 1.3E^3 \cdot \text{tag16018111Z160}[c] + \\ & 1.3E^3 \cdot \text{tag1601819Z160}[c] + 1.3E^3 \cdot \text{tag1801819Z160}[c] + \\ & 1.3E^3 \cdot \text{tag18111Z18111Z160}[c] + 1.3E^3 \cdot \text{tag18111Z1819Z160}[c] + \\ & 1.3E^3 \cdot \text{tag1819Z18111Z160}[c] + 37.1E^3 \cdot \text{ac}[c] + \\ & 30.0E^3 \cdot \text{ppa}[c] + 25.3E^3 \cdot \text{but}[c] + 12.1E^3 \cdot \text{glyc}[c] + \\ & 7.8E^3 \cdot \text{chla}[u] + 14.3E^3 \cdot \text{chlb}[u] + 1.0E^6 \cdot \text{rhodopsin}[s] + \\ & 4.0E^6 \cdot \text{acaro}[h] + 790.8E^9 \cdot \text{anxan}[u] + 11.1E^6 \cdot \text{caro}[u] + \\ & 5.1E^6 \cdot \text{loroxan}[u] + 9.9E^6 \cdot \text{lut}[u] + 4.3E^6 \cdot \text{neoxan}[u] + \\ & 2.8E^6 \cdot \text{vioxan}[u] + 2.4E^6 \cdot \text{zaxan}[u] + 29.9 \cdot \text{ATP maintainance} + \\ & 2.3E^3 \cdot \text{pe1801835Z9Z12Z}[c] + 1.9E^3 \cdot \text{pail18111Z160}[c] + \\ & 258.4E^6 \cdot \text{pail1819Z160}[c] \end{aligned}$$

A.6. Appendix 6: Biomass Function(Objective Function) for Different Growth Conditions

The biomass objective function for Heterotrophic:

Biomass =

$$\begin{aligned} & 309.1E^3 \cdot \text{ala-L[c]} + 95.0E^3 \cdot \text{arg-L[c]} + 65.2E^3 \cdot \text{asn-L[c]} + \\ & 65.2E^3 \cdot \text{asp-L[c]} + 11.1E^3 \cdot \text{cys-L[c]} + 82.5E^3 \cdot \text{gln-L[c]} + \\ & 82.5E^3 \cdot \text{glu-L[c]} + 99.8E^3 \cdot \text{gly[c]} + 10.6E^3 \cdot \text{his-L[c]} + \\ & 33.3E^3 \cdot \text{ile-L[c]} + 81.3E^3 \cdot \text{leu-L[c]} + 19.7E^3 \cdot \text{lys-L[c]} + \\ & 10.6E^3 \cdot \text{met-L[c]} + 35.4E^3 \cdot \text{phe-L[c]} + 46.9E^3 \cdot \text{pro-L[c]} + \\ & 23.0E^3 \cdot \text{ser-L[c]} + 92.9E^3 \cdot \text{thr-L[c]} + 6.0E^3 \cdot \text{trp-L[c]} + \\ & 6.0E^3 \cdot \text{tyr-L[c]} + 56.0E^3 \cdot \text{val-L[c]} + 2.2E^3 \cdot \text{datp[c]} + \\ & 3.9E^3 \cdot \text{dctp[c]} + 3.9E^3 \cdot \text{dgtp[c]} + 2.2E^3 \cdot \text{dttp[c]} + \\ & 58.6E^3 \cdot \text{atp[c]} + 104.2E^3 \cdot \text{ctp[c]} + 104.2E^3 \cdot \text{gtp[c]} + \\ & 58.6E^3 \cdot \text{utp[c]} + 328.4E^3 \cdot \text{man[c]} + \\ & 524.1E^3 \cdot \text{arab-L[c]} + 697.0E^3 \cdot \text{gal[c]} + \\ & 28.4E^3 \cdot \text{mgdg1839Z12Z15Z1644Z7Z10Z13Z[h]} + \\ & 3.2E^3 \cdot \text{mgdg1839Z12Z15Z1637Z10Z13Z[h]} + 3.2E^3 \cdot \text{mgdg1839Z12Z15Z1634Z7Z10Z[h]} + \\ & 269.4E^6 \cdot \text{dgdg1839Z12Z15Z1644Z7Z10Z13Z[h]} + \\ & 739.2E^6 \cdot \text{dgdg1839Z12Z15Z1637Z10Z13Z[h]} + 739.2E^6 \cdot \text{dgdg1839Z12Z15Z1634Z7Z10Z[h]} + \\ & 74.3E^6 \cdot \text{dgts18111Z1819Z[c]} + 74.3E^6 \cdot \text{dgts18111Z18111Z[c]} + \\ & 1.1E^3 \cdot \text{dgts1601829Z12Z[c]} + 1.2E^3 \cdot \text{asqdp1819Z160[c]} + \end{aligned}$$

$$\begin{aligned}
 &1.2E^3 \cdot \text{asqdpal18111Z160}[c] + 1.3E^3 \cdot \text{tag16018111Z160}[c] + \\
 &1.3E^3 \cdot \text{tag1601819Z160}[c] + 1.3E^3 \cdot \text{tag1801819Z160}[c] + \\
 &1.3E^3 \cdot \text{tag18111Z18111Z160}[c] + 1.3E^3 \cdot \text{tag18111Z1819Z160}[c] + \\
 &1.3E^3 \cdot \text{tag1819Z18111Z160}[c] + 37.1E^3 \cdot \text{ac}[c] + 30.0E^3 \cdot \text{ppa}[c] + \\
 &25.3E^3 \cdot \text{but}[c] + 12.1E^3 \cdot \text{glyc}[c] + 20.2E^3 \cdot \text{chla}[u] + \\
 &8.8E^3 \cdot \text{chlb}[u] + 1.0E^6 \cdot \text{rhodopsin}[s] + 79.7E^9 \cdot \text{acarol}[h] + \\
 &15.9E^9 \cdot \text{anxan}[u] + 223.3E^9 \cdot \text{caro}[u] + 103.7E^9 \cdot \text{loroxan}[u] + \\
 &199.4E^9 \cdot \text{lut}[u] + 87.7E^9 \cdot \text{neoxan}[u] + 55.8E^9 \cdot \text{vioxan}[u] + \\
 &47.8E^9 \cdot \text{zaxan}[u] + 29.9 \cdot \text{ATP maintenance} + \\
 &2.3E^3 \cdot \text{pe1801835Z9Z12Z}[c] + 1.9E^3 \cdot \text{pail18111Z160}[c] + \\
 &258.4E^6 \cdot \text{pail1819Z160}[c]
 \end{aligned}$$

A.7 Appendix 7: Matlab Codes

A.7.1 Interaction-based Approach Code

**Convert stoichiometric matrix to adjacency matrix and
Determine topology property of metabolites**

```

1 %% Reachability analysis and convert stoichiometric matrix to
2 % adjacency matrix get the stoichiometric matrix (which is
3 % saved as a .mat file), and get the
4 % variable stoi (double)

```

A.7. Appendix 7: Matlab Codes

```
5 % read a file, and load a file.
6 [filename, filepath] = uigetfile;
7 fullpath = [filepath filename];
8 load(fullpath);
9 siz = size(stoi.s);
10 % construct a reachability matrix "Rm", and convert
11 % stoichiometric matrix to adjacency matrix.
12 Rm.m = zeros(siz(2), siz(2));
13 Rm.met = stoi.mets;
14 for i = 1:siz(1)
15     a=0;b=0;
16     for j = 1:siz(2)
17         if stoi.s(i,j) < 0
18             a = a+1;
19             met.reactant(a) = j; % get the reactant
20         else if stoi.s(i,j) > 0
21             b = b+1;
22             met.product(b) = j;
23         end
24     end
25 end
26 if stoi.rev(i) == 1
27     met.reactant = [met.reactant met.product];
28     met.product = met.reactant;
29     a = a +b;
30     b = a;
31 end
32 for k = 1:a
33     for m = 1:b
34         Rm.m(met.reactant(k),met.product(m)) = 1;
35     end
36 end
37 met.reactant = zeros;
```

A.7. Appendix 7: Matlab Codes

```
38         met.product = zeros;
39     end
40     % clear the self-linked reachability error. and
41     % get the Rm^2, Rm^3
42     for i = 1:size(Rm.m)
43         Rm.m(i,i) = 0;
44     end
45     Rm.m2 = Rm.m * Rm.m;
46     for i = 1:size(Rm.m)
47         Rm.m2(i,i) = 0;
48     end
49     Rm.m3 = Rm.m^3;
50     for i = 1:size(Rm.m)
51         Rm.m3(i,i) = 0;
52     end
```

Find gaps in the metabolite networks

```
1  %% Find gaps in the metabolite networks.
2  % this program is to convert the matrix from SBML into double
3  % stoichiometric matrix. 761 and 932 and be replaced by the actual
4  % size of the model.
5  initCobratoolbox;
6  sto = model.S;
7  stoi = model;
8  stoi.s = zeros (size(sto));
9  stoi.s = full(sto);
10 stoi.rev = model.rev;
11 stoi.s = stoi.s'; %need to get a matrix with same row same reaction.
12 %%
```

A.7. Appendix 7: Matlab Codes

```
13
14 % get the stoichiometric matrix (which is saved as a .mat file),
15 %and get the variable stoi (double)
16 % read a file, and load a file.
17 % [filename, filepath] = uigetfile;
18 % fullpath = [filepath filename];
19 % load(fullpath);
20 siz = size(stoi.s);
21 % construct a reachability matrix "Rm",
22 Rm.m = zeros(siz(2), siz(2));
23 Rm.met = stoi.mets;
24 Rm.count = zeros(siz(2));
25 Rm.revmet = zeros(siz(2));
26 for i = 1:siz(1)
27     a=0;b=0;
28     for j = 1:siz(2)
29         if stoi.s(i,j) < 0
30             a = a+1;
31             met.reactant(a) = j; % get the reactant
32             Rm.count(j)= Rm.count(j)+1;
33         else if stoi.s(i,j) > 0
34             b = b+1;
35             met.product(b) = j;
36             Rm.count(j)= Rm.count(j)+1;
37         end
38     end
39 end
40
41 for k = 1:a
42     for m = 1:b
43         Rm.m(met.reactant(k),met.product(m)) =
44         Rm.m(met.reactant(k),met.product(m))+1;
45     end
```

A.7. Appendix 7: Matlab Codes

```
46     end
47         met.reactant = zeros;
48         met.product = zeros;
49     end
50     % if it's a reversible reaction.
51     for i = 1:siz(1)
52         a=0; b = 0;
53         if stoi.rev(i) ~= 0
54             for j = 1:siz(2)
55                 if stoi.s(i,j) > 0
56                     a = a+1;
57                     Rm.revmet(j) = 1;
58                 % revmet counts the metabolites in the reversible rxns.
59                 met.reactant(a) = j; % get the reactant
60                 else if stoi.s(i,j) < 0
61                     b = b+1;
62                     met.product(b) = j;
63                     Rm.revmet(j) = 1;
64                 end
65             end
66         end
67         for k = 1:a
68             for m = 1:b
69                 Rm.m(met.reactant(k),met.product(m)) =
70                     Rm.m(met.reactant(k),met.product(m))+1;
71             end
72         end
73     end
74     met.reactant = zeros;
75     met.product = zeros;
76 end
77
78 % clear the self-linked reachability error. and get the Rm^2, Rm^3
```

A.7. Appendix 7: Matlab Codes

```
79 for i = 1:size(Rm.m)
80     Rm.m(i,i) = 0;
81 end
82 Rm.m2 = Rm.m * Rm.m;
83 for i = 1:size(Rm.m)
84     Rm.m2(i,i) = 0;
85 end
86 Rm.m3 = Rm.m^3;
87 for i = 1:size(Rm.m)
88     Rm.m3(i,i) = 0;
89 end
90
91 % find out the dead-end in the reversible reactions.
```

Determine clustering coefficient

```
1 %% Determine clustering coefficient for each metabolite.
2 % Bioinformatics toolbox is used here.
3 siz = 8;
4 Rm.m = sparse(Rm.m);
5 Rm.pcount = zeros(1,siz);
6 Rm.path = num2cell(zeros(siz,siz));
7 for i = 1: siz;
8     [Rm.dist(i,:),Rm.path(i,:),PRED] = GRAPHSHORTESTPATH(Rm.m,i);
9 end;
10 for i = 1: siz;
11     for k= 1: siz;
12         if ~isempty(Rm.path{i,k});
13             n = length(Rm.path{i,k});
14             for ks = 2 : n-1;
```

A.7. Appendix 7: Matlab Codes

```
15         r = Rm.path{i,k}(ks);
16         Rm.pcount(1,r) = Rm.pcount(1,r)+1;
17     end;
18 end;
19 end;
20 end;
21 Rm.pcount = Rm.pcount - 1;
```

A.7.2 Constraint-based Approach Code

Flux Balance Analysis to determine the metabolite essentiality

```
1 %% Flux Balance Analysis to determine the metabolite essentiality.
2 % Note: To use this code, first load iRC1080 into the COBRA
3 % toolbox in Matlab as a variable named "model".
4 % Then this code can work.
5
6
7 % Measures and constants.
8 DW = 48*10^(-12);
9 % avg. dry weight of log phase chlamy cell = 48 pg (Mitchell 1992)
10 CPerStarch300 = 1800;
11 % derived from starch300 chemical formula
12 ChlPerCell = (13.9+4)/(10^7);
13 % 13.9 +- 4 micrograms Chl/10^7 cells (Gfeller 1984)
14 starchDegAnLight = (4.95+1.35)*(1/1000)*(1/CPerStarch300)*
15 (ChlPerCell/1000)*(1/DW);
16 % approx. SS rate of anaerobic starch degradation in light
17 = 4.95 +- 1.35 micromol C/mg Chl/hr (Gfeller 1984)
18 starchDegAerLight = (2/3)*starchDegAnLight;
```


A.7. Appendix 7: Matlab Codes

```
19 % approx. SS rate of aerobic starch degradation in light =
20 2/3 of anaerobic rate (Gfeller 1984)
21 starchDegAnDark = (13.1+3.5)*(1/1000)*(1/CPerStarch300)*
22 (ChlPerCell/1000)*(1/DW);
23 % approx. SS rate of anaerobic starch degradation in dark =
24 13.1 +- 3.5 micromol C/mg Chl/hr (Gfeller 1984)
25 starchDegAerDark = (2/3)*starchDegAnDark;
26 % approx. SS rate of aerobic starch degradation in dark =
27 % 2/3 of anaerobic rate (Gfeller 1984)
28 dimensionalConversion = 3.836473679;
29 % from emitted microE/m^2/s to incident mmol/gDW/hr
30 effectiveConversion = 0.037532398;
31 % from incident mmol/gDw/hr to effective mmol/gDw/hr
32
33
34 %% set constraints.
35 % %%% light, aerobic, no acetate, biomass objective
36 modelLna = model;
37 % The single PRISM reaction being used has to be commented-out
38 %below.
39 modelLna = changeRxnBounds(modelLna,{...
40 % 'PRISM_solar_litho',...
41 'PRISM_solar-exo',...
42 'PRISM_incandescent_60W',...
43 'PRISM_fluorescent_warm_18W',...
44 'PRISM_fluorescent_cool_215W',...
45 'PRISM_metal_halide',...
46 'PRISM_high_pressure_sodium',...
47 'PRISM_growth_room',...
48 'PRISM_white_LED',...
49 'PRISM_red_LED_array_653nm',...
50 'PRISM_red_LED_674nm',...
51 'PRISM_design_growth',...
```

A.7. Appendix 7: Matlab Codes

```
52 },0,'b');
53 modelLna = changeRxnBounds(modelLna,{ 'EX_o2(e) '},-10,'l');
54 modelLna = changeRxnBounds(modelLna,{ 'EX_ac(e) '},0,'l');
55 modelLna = changeRxnBounds(modelLna,{ 'EX_starch(h) '},0,'b');
56 modelLna = changeRxnBounds(modelLna,'STARCH300DEGRA',
57 starchDegAerLight/2,'u');
58 modelLna = changeRxnBounds(modelLna,
59 'STARCH300DEGR2A',0,'u');
60 modelLna = changeRxnBounds(modelLna,
61 'STARCH300DEGRB',starchDegAerLight/2,'u');
62 modelLna = changeRxnBounds(modelLna
63 , 'STARCH300DEGR2B',0,'u');
64 modelLna = changeRxnBounds(modelLna,
65 { 'PCHLDR'},0,'b');
66 % the light-independent protochlorophyllide reductase is not
67 % expressed in light due to translational inhibition caused by
68 % chloroplast redox state [Cahoon 2000]
69 modelLna = changeRxnBounds(modelLna,{ 'PFKh'},0,'b');
70 % plastidic PFKh inactivated by light (Plaxton 1996)
71 modelLna = changeRxnBounds(modelLna,{ 'G6PADHh','G6PBDHh'},0,'b');
72 % light inhibits G6PDHh of oxidative pentose phosphate
73 % pathway (Plaxton 1996)
74 modelLna = changeRxnBounds(modelLna,{ 'FBAh'},0,'b');
75 % light inactivates FBAh (Lemaire 2004; Matsumoto 2008)
76 modelLna = changeRxnBounds(modelLna,{ 'H2Oth'},0,'u');
77 % there is a high h2o requirement in [h]; however,
78 % experiments show that h2o in general goes from [h] to
79 % [c] in light and from [c] to [h] in dark (Packer 1970)
80 modelLna = changeRxnBounds(modelLna,
81 { 'Biomass_Chlamy_mixo','Biomass_Chlamy_hetero'},0,'b');
82 modelLna = changeObjective(modelLna,'Biomass_Chlamy_auto');
83
84 % Base growth.
```

A.7. Appendix 7: Matlab Codes

```
85 solutionLna = optimizeCbModel(modelLna, 'max', 'one');
86
87
88 %% to get the flux sum.
89 solution = solutionLna;
90 siz = size(model.S);
91 sizem = siz(1); % number of mets.
92 sizer = siz(2); % to get the number of rxns.
93
94 %% Identify the essential Metabolites.
95 % find the rxn in which metabolite i is a reactant. r : reactant, p:
96 % product.
97 modeld = modelLna;
98 for i = 1 : sizem;
99     for j = 1: sizer;
100         if modeld.S(i,j) > 0;
101             modeld.lb(j,1)= 0;
102             modeld.ub(j,1) =0;
103         elseif modeld.S(i,j) < 0;
104             modeld.ub(j,1) =0;
105             modeld.lb(j,1)= 0;
106         end
107     end
108     solution_x = optimizeCbModel(modeld, 'max', 'one');
109     s_effectr(i,1) = solution_x.f - solutionLna.f;
110     s_effectr(i,2) = -s_effect(i,1)/solution.f;
111 end
```

Obtain basal flux-sum in different growth conditions

A.7. Appendix 7: Matlab Codes

```

1 % To obtain basal flux-sum in different growth conditions.
2 % Note: To use this code, first load iRC1080 into the COBRA
3 % toolbox in Matlab as a variable named "model". Then this code can be run.
4
5
6 % Measures and constants.
7 DW = 48*10−12;
8 % avg. dry weight of log phase chlamy cell = 48 pg (Mitchell 1992)
9 CPerStarch300 = 1800;
10 % derived from starch300 chemical formula
11 ChlPerCell = (13.9+4)/(107);
12 % 13.9 +- 4 micrograms Chl/107 cells (Gfeller 1984)
13 starchDegAnLight = (4.95+1.35)*(1/1000)*(1/CPerStarch300)*
14 (ChlPerCell/1000)*(1/DW);
15 % approx. SS rate of anaerobic starch degradation in light
16 = 4.95 +- 1.35 micromol C/mg Chl/hr (Gfeller 1984)
17 starchDegAerLight = (2/3)*starchDegAnLight;
18 % approx. SS rate of aerobic starch degradation in light
19 = 2/3 of anaerobic rate (Gfeller 1984)
20 starchDegAnDark = (13.1+3.5)*(1/1000)*(1/CPerStarch300)
21 *(ChlPerCell/1000)*(1/DW);
22 % approx. SS rate of anaerobic starch degradation in
23 dark = 13.1 +- 3.5 micromol C/mg Chl/hr (Gfeller 1984)
24 starchDegAerDark = (2/3)*starchDegAnDark;
25 % approx. SS rate of aerobic starch degradation in dark
26 = 2/3 of anaerobic rate (Gfeller 1984)
27 dimensionalConversion = 3.836473679;
28 % from emitted microE/m2/s to incident mmol/gDW/hr
29 effectiveConversion = 0.037532398;
30 % from incident mmol/gDw/hr to effective mmol/gDw/hr
31
32
33 % %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

A.7. Appendix 7: Matlab Codes

```

34 % %%% light, aerobic, no acetate, biomass objective
35 modelLna = model;
36 % The single PRISM reaction being used has to be commented-out below.
37 modelLna = changeRxnBounds(modelLna, {...
38 %     'PRISM_solar_litho',...
39     'PRISM_solar_exo',...
40     'PRISM_incandescent_60W',...
41     'PRISM_fluorescent_warm_18W',...
42     'PRISM_fluorescent_cool_215W',...
43     'PRISM_metal_halide',...
44     'PRISM_high_pressure_sodium',...
45     'PRISM_growth_room',...
46     'PRISM_white_LED',...
47     'PRISM_red_LED_array_653nm',...
48     'PRISM_red_LED_674nm',...
49     'PRISM_design_growth',...
50 },0,'b');
51 modelLna = changeRxnBounds(modelLna,{'EX_o2(e)'},-10,'l');
52 modelLna = changeRxnBounds(modelLna,{'EX_ac(e)'},0,'l');
53 modelLna = changeRxnBounds(modelLna,{'EX_starch(h)'},0,'b');
54 modelLna = changeRxnBounds(modelLna,'STARCH300DEGRA'
55 ,starchDegAerLight/2,'u');
56 modelLna = changeRxnBounds(modelLna,'STARCH300DEGR2A',0,'u');
57 modelLna = changeRxnBounds(modelLna,'STARCH300DEGRB
58 ',starchDegAerLight/2,'u');
59 modelLna = changeRxnBounds(modelLna,'STARCH300DEGR2B',0,'u');
60 modelLna = changeRxnBounds(modelLna,{'PCHLDR'},0,'b');
61 modelLna = changeRxnBounds(modelLna,{'PFKh'},0,'b');
62 modelLna = changeRxnBounds(modelLna,{'G6PADHh','G6PBDHh'},0,'b'); )
63 modelLna = changeRxnBounds(modelLna,{'FBAh'},0,'b');
64 modelLna = changeRxnBounds(modelLna,{'H2Oth'},0,'u');
65 modelLna = changeRxnBounds(modelLna,
66 {'Biomass_Chlamy_mixo','Biomass_Chlamy_hetero'},0,'b');

```

A.7. Appendix 7: Matlab Codes

```
67 modelLna = changeObjective(modelLna, 'Biomass_Chlamy_auto');
68
69 % Base growth.
70 solutionLna = optimizeCbModel(modelLna, 'max', 'one');
71
72 modelabs = abs(model.S);
73 fluxsum(1,:) = modelabs * solutionLna.x*0.5 ;
74
75
76 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
77 %%% light, aerobic, w/ acetate, biomass objective
78 modelLwac = model;
79 % The single PRISM reaction being used has
80 %to be commented-out below.
81 modelLwac = changeRxnBounds(modelLwac, {...
82 %   'PRISM_solar_litho',...
83   'PRISM_solar_exo',...
84   'PRISM_incandescent_60W',...
85   'PRISM_fluorescent_cool_215W',...
86   'PRISM_metal_halide',...
87   'PRISM_high_pressure_sodium',...
88   'PRISM_growth_room',...
89   'PRISM_white_LED',...
90   'PRISM_red_LED_array_653nm',...
91   'PRISM_red_LED_674nm'...
92   'PRISM_fluorescent_warm_18W'...
93   'PRISM_design_growth',...
94 }, 0, 'b');
95 modelLwac = changeRxnBounds(modelLwac,
96 {'EX_o2(e)', 'EX_ac(e)'}, -10, 'l');
97 modelLwac = changeRxnBounds(modelLwac,
98 {'EX_starch(h)'}, 0, 'b');
99 modelLwac = changeRxnBounds(modelLwac,
```

A.7. Appendix 7: Matlab Codes

```

100 'STARCH300DEGRA',
101 starchDegAerLight/2, 'u');
102 modelLwac = changeRxnBounds(modelLwac,
103 'STARCH300DEGR2A', 0, 'u');
104 modelLwac = changeRxnBounds(modelLwac, 'STARCH300DEGRB',
105 starchDegAerLight/2, 'u');
106 modelLwac = changeRxnBounds(modelLwac,
107 'STARCH300DEGR2B', 0, 'u');
108 modelLwac = changeRxnBounds(modelLwac, {'PCHLDR'}, 0, 'b');
109 modelLwac = changeRxnBounds(modelLwac, {'PFKh'}, 0, 'b');
110 modelLwac = changeRxnBounds(modelLwac,
111 {'G6PADHh', 'G6PBDHh'}, 0, 'b');
112 modelLwac = changeRxnBounds(modelLwac, {'FBAh'}, 0, 'b');
113 modelLwac = changeRxnBounds(modelLwac, {'H2Oth'}, 0, 'u');
114 modelLwac = changeRxnBounds(modelLwac,
115 {'Biomass-Chlamy-auto', 'Biomass-Chlamy-hetero'}, 0, 'b');
116 modelLwac = changeObjective(modelLwac, 'Biomass-Chlamy-mixo');
117
118 % Base growth
119 solutionLwac = optimizeCbModel(modelLwac, 'max', 'one');
120
121 fluxsum(2,:) = modelabs * solutionLwac.x*0.5 ;
122 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
123 %% dark, aerobic, w/ acetate, biomass objective
124 modelDa = model;
125 modelDa = changeRxnBounds(modelDa, 'EX_photonVis(e)', 0, 'l');
126 modelDa = changeRxnBounds(modelDa, {'EX_o2(e)'}, -10, 'l');
127 modelDa = changeRxnBounds(modelDa, 'EX_co2(e)', 0, 'l');
128 modelDa = changeRxnBounds(modelDa,
129 'STARCH300DEGRA', 0, 'u');
130 modelDa = changeRxnBounds(modelDa,
131 'STARCH300DEGR2A'
132 ,starchDegAerDark/2, 'u');

```

A.7. Appendix 7: Matlab Codes

```
133 modelDa = changeRxnBounds(modelDa,
134 'STARCH300DEGRB', 0, 'u');
135 modelDa = changeRxnBounds(modelDa,
136 'STARCH300DEGR2B'
137 , starchDegAerDark/2, 'u');
138 modelDa = changeRxnBounds(modelDa, {'GLPThi'}, 0, 'u');
139 modelDa = changeRxnBounds(modelDa, {'ATPSh'}, 0, 'b');
140 modelDa = changeRxnBounds(modelDa, {'GAPDH(nadp)hi'}, 0, 'b');
141 modelDa = changeRxnBounds(modelDa, {'MDH(nadp)hi',
142 'MDHC(nadp)hr'}, 0, 'b'); % inactive in dark (Buchanan 1980)
143 modelDa = changeRxnBounds(modelDa, {'PPDKh'}, 0, 'b');
144 modelDa = changeRxnBounds(modelDa, {'IDPh'}, 0, 'b');
145 modelDa = changeRxnBounds(modelDa, {'PRUK'}, 0, 'b');
146 modelDa = changeRxnBounds(modelDa, {'RBPCh', 'RBCh'}, 0, 'b');
147 modelDa = changeRxnBounds(modelDa, {'SBP'}, 0, 'b');
148 modelDa = changeRxnBounds(modelDa, {'H2Oth'}, 0, 'l');
149 modelDa = changeRxnBounds(modelDa,
150 {'Biomass_Chlamy_auto', 'Biomass_Chlamy_mixo'}, 0, 'b');
151 modelDa = changeObjective(modelDa, 'Biomass_Chlamy_hetero');
152
153 % Base growth
154 solutionDa = optimizeCbModel(modelDa, 'max', 'one');
155 fluxsum(3,:) = modelabs * solutionDa.x*0.5 ;
```

Flux Sum Analysis for Cobratoolbox

```
1 %% Replace the optimizeCbModel in CobraToolbox with this code to
2 % obtain flux sum attenuation analysis.
3 if (nargin < 2)
4     osenseStr = 'max';
```



```

5  end
6  if (nargin < 3)
7      primalOnlyFlag = true;
8  end
9  if (nargin < 4)
10     minNormFlag = false;
11 end
12 if (nargin < 5)
13     verbFlag = false;
14 end
15
16 % LP solution tolerance
17 if exist('CBTLPTOL','var')
18     tol = CBTLPTOL;
19 else
20     tol = 1e-6;
21 end
22
23 % Figure out objective sense
24 if (strcmp(osenseStr,'max'))
25     LPproblem.osense = -1;
26 else
27     LPproblem.osense = +1;
28 end
29
30 % All constraints are equalities
31 LPproblem.csense = [];
32 %LPproblem.csense = zeros(1707,1);
33 %LPproblem.csense(1707,1) = 'L';
34
35 % Fill in the RHS vector if not provided
36 if (~isfield(model,'b'))
37     LPproblem.b = zeros(length(model.mets),1);

```

A.7. Appendix 7: Matlab Codes

```
38 else
39     LPproblem.b = model.b;
40 end
41 % Rest of the LP problem
42 LPproblem.A = model.S;
43 LPproblem.c = model.c;
44 LPproblem.lb = model.lb;
45 LPproblem.ub = model.ub;
46
47 %% Solve initial LP
48
49 LPsolution = solveCobraLP(LPproblem,primalOnlyFlag);
50 time1 = 0;
51
52 %% Solve secondary LP to minimize |v|
53
54 if (LPsolution.stat ~= 1)
55     if (verbFlag)
56         warning('Optimal solution was not found');
57     end
58
59     FBAsolution.f = 0;
60     FBAsolution.x = [];
61 else
62     % Store results
63     FBAsolution.f = LPsolution.obj;
64     FBAsolution.x = LPsolution.full;
65     if (~primalOnlyFlag)
66         FBAsolution.y = LPsolution.dual;
67         FBAsolution.w = LPsolution.rcost;
68     end
69
70     % Minimize the absolute value of fluxes to avoid
```

A.7. Appendix 7: Matlab Codes

```
71 % loopy solutions
72     if (minNormFlag)
73         if (strcmp(osenseStr, 'max'))
74             FBAsolution.f = floor(FBAsolution.f/tol)*tol;
75         else
76             FBAsolution.f = ceil(FBAsolution.f/tol)*tol;
77         end
78         if (FBAsolution.f ~= 0)
79             [nMets, nRxns] = size(model.S);
80             % Set up the optimization problem
81             % min sum(delta+ + delta-)
82             % 1: S*v1 = 0
83             % 3: delta+ >= -v1
84             % 4: delta- >= v1
85             % 5: c'*v1 >= f (optimal value of objective)
86             %
87             % delta+, delta- >= 0
88             LPproblem2.A = [model.S sparse(nMets, 2*nRxns);
89                             speye(nRxns, nRxns) speye(nRxns, nRxns)
90                             sparse(nRxns, nRxns);
91                             -speye(nRxns, nRxns) sparse(nRxns, nRxns)
92                             speye(nRxns, nRxns);
93                             model.c' sparse(1, 2*nRxns)];
94             LPproblem2.c = [zeros(nRxns, 1); ones(2*nRxns, 1)];
95             LPproblem2.lb = [model.lb; zeros(2*nRxns, 1)];
96             LPproblem2.ub = [model.ub; 10000*ones(2*nRxns, 1)];
97             LPproblem2.b = [LPproblem.b; zeros(2*nRxns, 1); FBAsolution.f];
98             LPproblem2.csense(1:nMets) = 'E';
99             LPproblem2.csense((nMets+1):(nMets+2*nRxns)) = 'G';
100             LPproblem2.csense(nMets+2*nRxns+1) = 'G';
101             LPproblem2.csense = columnVector(LPproblem2.csense);
102             LPproblem2.osense = 1;
103             % Re-solve the problem
```

A.7. Appendix 7: Matlab Codes

```
104         time1 = LPsolution.time;
105         LPsolution = solveCobraLP(LPproblem2,primalOnlyFlag);
106         %[f,x,y,w,solStatus] = solveLPStm(A,b,C,lb,ub,
107 1,columnVector(csense));
108         if (LPsolution.stat > 0)
109             FBAsolution.x = LPsolution.full(1:nRxns);
110         else
111             FBAsolution.x = [];
112         end
113     end
114 end
115 end
116
117 FBAsolution.stat = LPsolution.stat;
118 FBAsolution.solver = LPsolution.solver;
119 FBAsolution.time = LPsolution.time+time1;
```

Draw figures for flux sum attenuation to categorize metabolites

```
1 %%This is to draw figures for each metabolite with the flux sum attenuation
2 %%data to categorize them.
3 for i = 1 : 22;
4     xaxis(i) = 0.05*i;
5 end
6
7 for j = 1:1;
8     for i = 1: 100;
9         if MECR(i,2*j)> 0.5;
10             %if average(fsaatt(1,i,:)) > 0.02;
11                 figure(i);
```

A.7. Appendix 7: Matlab Codes

```
12         fq(1:22) = fsaatt(j,i,1:22);
13         plot(xaxis(1:22),fq(1:22));
14         m = num2str([j i]);
15         print(m,'-djpeg')
16         close(i);
17     %end
18 end;
19 end;
20 end;
```

Flux sum attenuation

```
1  %%%
2  % Manipulate flux-sum by attenuation
3
4  %%set model, and set the first FBA growth conditions.
5  %% set constraints.
6  % %% light, aerobic, no acetate, biomass objective
7  % The single PRISM reaction being used has to be commented-out below.
8  modelLna = changeRxnBounds(modelLna,{...
9  %     'PRISM_solar_litho',...
10     'PRISM_solar_exo',...
11     'PRISM_incandescent_60W',...
12     'PRISM_fluorescent_warm_18W',...
13     'PRISM_fluorescent_cool_215W',...
14     'PRISM_metal_halide',...
15     'PRISM_high_pressure_sodium',...
16     'PRISM_growth_room',...
17     'PRISM_white_LED',...
18     'PRISM_red_LED_array_653nm',...
```

A.7. Appendix 7: Matlab Codes

```
19     'PRISM_red_LED_674nm', ...
20     'PRISM_design-growth', ...
21 }, 0, 'b');
22 modelLna = changeRxnBounds(modelLna, {'EX_o2(e)'}, -10, 'l');
23 modelLna = changeRxnBounds(modelLna, {'EX_ac(e)'}, 0, 'l');
24 modelLna = changeRxnBounds(modelLna, {'EX_starch(h)'}, 0, 'b');
25 modelLna = changeRxnBounds(modelLna,
26 'STARCH300DEGRA', starchDegAerLight/2, 'u');
27 modelLna = changeRxnBounds(modelLna,
28 'STARCH300DEGR2A', 0, 'u');
29 modelLna = changeRxnBounds(modelLna,
30 'STARCH300DEGRB', starchDegAerLight/2, 'u');
31 modelLna = changeRxnBounds(modelLna, 'STARCH300DEGR2B', 0, 'u');
32 modelLna = changeRxnBounds(modelLna, {'PCHLDR'}, 0, 'b');
33 modelLna = changeRxnBounds(modelLna, {'PFKh'}, 0, 'b');
34 modelLna = changeRxnBounds(modelLna, {'G6PADHh', 'G6PBDHh'}, 0, 'b');
35 modelLna = changeRxnBounds(modelLna, {'FBAh'}, 0, 'b');
36 modelLna = changeRxnBounds(modelLna, {'H2Oth'}, 0, 'u');
37 modelLna = changeRxnBounds(modelLna,
38 {'Biomass_Chlamy_mixo', 'Biomass_Chlamy_hetero'}, 0, 'b');
39 modelLna = changeObjective(modelLna, 'Biomass_Chlamy_auto');
40
41 %% Base growth.
42 solutionLna = optimizeCbModel(modelLna, 'max', 'one');
43
44 %% add a flux sum constraints to implement flux sum attenuation
45 % analysis
46 for i = 1 : 1706; %sizem;
47     if MECR(i,2) > 0.5;
48         modelLnax = modelLna;
49         modelLnax.S(1707,:) = abs(modelLnax.S(i,:));
50         for att = 1 : 20;
51             modelLnax.b(1707) = att/20*fluxsum(1,i)*2;
```

A.7. Appendix 7: Matlab Codes

```
52 %because we times 0.5 when get the flux sum.
53     solutionLnax = optimizeCbModel(modelLnax, 'max', 'one');
54     fsaatt(1,i,att) = solutionLnax.f
55     %xaxis(i) = att/20;
56 % end
57
58 end
59 end
```