# UNIVERSITY OF CALIFORNIA, SAN DIEGO

**Systems analysis of energy metabolism elucidates the roles of mitochondria in human health and disease**

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Bioinformatics

by

Thuy D. Vo

**Committee in charge:**

Professor Bernhard Ø. Palsson, Chair
Professor Andrew D. McCulloch, Co-chair
Professor Jack E. Dixon
Professor Philip E. Gill
Professor Alexander Hoffmann

2007

The dissertation of Thuy D. Vo is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____
Co-Chair

_____
Chair

University of California, San Diego

2007

*To John*

*For all your love and support and*
*for always having answers to all of my questions.*

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

## ACKNOWLEDGEMENTS

First I would like to thank all of my teachers who have instilled in me an appreciation for knowledge and education. Your lessons have not only improved my understanding of science but have also made me who I am today. Specifically, I would like to thank my advisor, Dr. Bernhard Palsson, for the privilege of working and studying with him in an exciting area of science. He is my toughest critic, and for that I am a better scientist. Thank you for believing in me and for always providing me opportunities to grow in both scientific pursuit and leadership. I would like to thank Dr. Philip Gill for his generosity with his time and for the many practical and theoretical lessons in optimization. Thank you for developing SNOPT, which enables much of the research under this dissertation. I thank Dr. Paul Lee for his guidance, both scientific and otherwise. I thank Dr. Harvey Greenberg for the opportunity to study and work with him. I thank my committee members, Dr. Alexander Hoffmann, Dr. Andrew McCulloch, and Dr. Jack Dixon for their thought, time, and suggestions that have benefited the work presented here. Lastly, I am extremely grateful for the financial support provided by the San Diego and GREAT fellowships.

Members of the Systems Biology Research Group and fellows in the Bioinformatics Program have also contributed significantly to the making of this dissertation. Thank you, Jennie, for the many productive discussions, without which much of this work might have not materialized. Thank you, Markus and Nathan, for always having an answer to my questions, be they scientific, statistics or formatting Word documents. Thank you also for being great role models. I thank Natalie, Ines, Scott, Neema, Andrew, and Christian for being great lab mates, for always willing to read my papers, and for the mid-afternoon bantering that actually makes research interesting. Thank you, Adam and Jan, for all the technical support and for keeping the computers and printers in the lab running. Thank you, Qiang, Norman, and Shu, for teaching me and helping me with the experimental techniques. The experimental studies in this dissertation would have not been possible without your help. Finally, I thank Leah and

Silpa, for being great girl friends, for always being available when I need them, and for keeping me sane.

Lastly, I would have not been where I am today if not for my family's love and support. I thank my grandma for helping me to see the value of education at an early age. I thank my many uncles and aunts for giving me the opportunity to be in this country to realize my full potential. I thank my parents for always believing in me even when they do not always understand what I do. I think Tom and Lou Mongan for their constant support and proofreading this dissertation. And finally, words can not express how grateful I am to have John in my life. Thank you for your love, support, encouragement, and for always having solutions to all of my problems. I look forward to a lifetime discussing science with you.

Chapter 1, in part, is a preprint of the material appearing in "Building the power house: Recent advances in mitochondrial studies through proteomics and systems biology," Vo TD and Palsson BO. 2006. *Am J Physiol Cell Physiol*, in press. The dissertation author is the primary investigator and author of this paper. Chapter 2, in part, is a reprint of the material appearing in "In silico Analysis of SNPs and Other High Throughput Data", Jamshidi N, Vo TD, and Palsson BO. 2006. *Cardiac gene expression: methods and protocols*, Totowa, N.J.: Humana Press, in press; "Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data," Vo TD, Greenberg HJ, and Palsson BO. 2006. *J Biol Chem* 279(38):39532-40; "Isotopomer analysis of cellular metabolism in tissue culture: a comparative study between the pathway and the network-based methods", Vo TD, Lim SK, Lee PWN, and Palsson BO. 2006. *Metabolomics*, in press; and "Systems analysis of energy metabolism elucidates the affected respiratory chain complex in Leigh's syndrome", Vo TD, Lee PWN, and Palsson BO, *Molecular genetics and Metabolism,* submitted. The dissertation author is the primary investigator and author of these publications. Chapter 3, in part, is a reprint of the material appearing in "Reconstruction and functional characterization of the human mitochondrial metabolic network based on

proteomic and biochemical data," Vo TD, Greenberg HJ, and Palsson BO. 2004. *J Biol Chem* 279(38):39532-40 and "Candidate metabolic network states in human mitochondria: Impact of diabetes, ischemia, and diet," Thiele I, Price ND, Vo TD, and Palsson BO. 2005a. *J Biol Chem* 280(12):11683-95. The dissertation author is the primary investigator and an author of the materials used from these papers. Chapter 4, in part, is a preprint of the material appearing in ""Isotopomer analysis of cellular metabolism in tissue culture: A comparative study between the pathway- and network-based methods," Vo TD, Lim SK, Lee PWN, and Palsson BO. 2006. *Metabolomics,* in press and "Isotopomer analysis of myocardial substrate metabolism: A systems biology approach," Vo TD and Palsson BO. 2006. *Biotechnology and Bioengineering* 95(5):972-83. The dissertation author is the primary investigator and author of these papers. Chapter 5, in part, is a reprint of the material appearing in "Isotopomer analysis of myocardial substrate metabolism: A systems biology approach," Vo TD and Palsson BO. 2006. *Biotechnology and Bioengineering*, 95(5):972-83. The dissertation author is the primary investigator and author of this paper. Chapter 6, in part, is a preprint of the material appearing in "Systems analysis of energy metabolism elucidates the affected respiratory chain complex in Leigh's syndrome, Vo TD, Lee, PWN, and Palsson BO, *Molecular genetics and Metabolism,* submitted. The dissertation author is the primary investigator and author of this paper. Chapter 7, in part, is a preprint of the material appearing in "Building the power house: Recent advances in mitochondrial studies through proteomics and systems biology", Vo TD and Palsson BO. *Am J Physiol Cell Physiol*. 2006. in press. The dissertation author is the primary investigator and author of this paper.

## VITA

2001 B.A, University of California, Berkeley (Berkeley, CA)

2007 Ph.D., University of California, San Diego (La Jolla, CA)

## PUBLICATIONS

1. **Vo TD**, Lee PWN, and Palsson BO, "Systems analysis of energy metabolism elucidates the affected respiratory chain complex in Leigh's syndrome, *Molecular Genetics and Metabolism*, submitted.
2. **Vo TD**, Lim SK, Lee PWN, and Palsson BO, "Isotopomer analysis of cellular metabolism in tissue culture: A comparative study between the pathway- and network-based methods," *Metabolomics*, in press.
3. **Vo TD** and Palsson, BO, "Isotopomer analysis of myocardial substrate metabolism: A systems biology approach," *Biotechnology and Bioengineering* 95(5):972-83.
4. **Vo TD** and Palsson, BO, "Building the power house: Recent advances in mitochondrial studies through proteomics and systems biology," *American Journal of Physiology – Cell Physiology,* in press.
5. **Vo TD**, Greenberg HJ, Palsson BO. 2004. "Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data," *J Biol Chem* 279(38):39532-40.
6. Thiele I, Price ND, **Vo TD**, Palsson BO. 2005a. "Candidate metabolic network states in human mitochondria: Impact of diabetes, ischemia, and diet," *J Biol Chem* 280(12):11683-95.
7. Duarte N, Becker S, Jamshidi N, Thiele I, Mo M, **Vo TD**, Srivas R, Palsson BO, "Global reconstruction of the human metabolic network based on genomic and bibliomic data," *PNAS*, accepted.
8. Jamshidi N*, **Vo TD***, Palsson BO, "In silico Analysis of SNPs and Other High Throughput Data," *Methods in Molecular Biology: Cardiac Gene Expression and Regulation*, Totowa, N.J.: Humana Press, in press.
9. Thiele I*, **Vo TD***, Price ND, Palsson BO. 2005b. Expanded metabolic reconstruction of *Helicobacter pylori* (iIT341 GSM/GPR): An *in silico* genome-scale characterization of single- and double-deletion mutants," *J Bacteriol* 187(16):5818-30.
10. Reed JL, **Vo TD**, Schilling CH, Palsson BO. 2003. "An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR)," *Genome Biol* 4(9):R54.

(*) Authors contributed equally to the work

## ABSTRACT OF THE DISSERTATION

**Systems analysis of energy metabolism elucidates the roles of mitochondria in human health and disease**

by

Thuy D. Vo

Doctor of Philosophy in Bioinformatics
University of California, San Diego, 2007

Professor Bernhard Ø. Palsson, Chair
Professor Andrew McCulloch, Co-chair

Beginning as the science of life, biology has been transformed and expanded throughout the years to give rise to fields such as ecology, molecular biology, biophysics, and biochemistry. The distinction among these disciplines is perhaps artificial as the study of any particular biological process or system often spans multiple fields. The research on energy metabolism to follow involves three fields: bioinformatics, bioengineering, and systems biology. The informatics aspect deals with the mining, organization, and management of genomic, proteomic, microarray and biochemical data for reconstructing metabolic networks. The engineering aspect employs the application of mathematics and scientific fundamentals to construct constraint-based models for examination. The systems biology paradigm allows the analysis and interpretation of the

models and data from a holistic perspective. Specifically, the findings of this work center on energy metabolism as understood by resource allocation and roles or mitochondria in mammalian cells in health and diseased states.

The constraint-based framework provides the facilities for biological discovery using tools from the three mentioned disciplines. This modeling approach describes a biological system as a network of components whose behaviors can be predicted upon the application of constraints. Energy metabolism of four systems – mitochondria, cardiomyocytes, hepatocytes, and fibroblasts – is investigated here. The application of linear constraints on metabolite mass, reaction reversibility and substrate utilization allows an exploration of network capability and topology. The capability and topology of the mitochondrial metabolic network are evaluated based on the theoretical energetic yield of substrates, the number of alternate pathways satisfying the same metabolic objective, and the effects of diabetic and ischemic conditions on feasible steady states. The second part of this dissertation involves the use of stable isotopes to uncover physiological steady states assumed by the cell. Nonlinear constraints are used to balance isotopomers and select for a set of flux distributions that match GC-MS data. Results from studies with cardiomyocytes and hepatocytes elucidate the paths undertaken by substrates as well as effects of media composition on intracellular flux distributions. Most significantly, results from tissue culture study identify complex II as the deficient complex in fibroblasts derived from a patient affected with Leigh's disease.

# Chapter 1
# Introduction

It has become my personal realization that the journey of "self-discovery" is the most challenging intellectual endeavor a person can undertake. The phrase self-discovery is used liberally here to refer to the investigation of the human body's organization, its cellular composition, and the intricate connections among such components rather than psychological pursuit of self-identity. This is not to discredit the importance of such a soul-searching exercise, but it is really the quest for understanding the inner workings of a human cell that is absolutely intriguing and mesmerizing to me. It has been said that "if the brain was simple enough for us to understand, we would be too stupid to understand it." I remember being particularly bothered by this saying when I first learned about it. However, as I delve deeper into the business of research, it becomes apparent to me that not only is the structure of human cell absolutely exquisite, it is mystifyingly so. I often find it uncomfortably difficult to explain to my non-biologically oriented relatives how it is possible to study genes and gene products without ever actually seeing them with one's own eyes. In fact, many discoveries in biology have been accomplished with experiments ingeniously designed to circumvent our inability to visualize biomolecular interactions or to verify events happening millions of years before the *Homo sapiens* species set foot on earth. If we stop thinking about all the diseases that we are not presently able to cure for one minute, we will immediately be overwhelmed by the

fantastic improvement in the quality of life modern medicine has been able to achieve since the invention of penicillin. The conception of molecular biology and the subsequent evolution of high-throughput biology is no less impressive. It is the rare instance that science is able to meet (albeit remotely) the fantasy created by Hollywood in the 1997 science fiction movie Gattaca. Though the work towards a $1000-genome is not yet within reach, it is incredible to think that we are able to witness, in our lifetime, the assembly of three billion base pairs of the human genome – a job harder than attempting to reassemble several thousand copies of the New York Times after they have been splattered with ink and shredded to millions of pieces. The genomic revolution became the driving force, in terms of scientific capability, publicity, and morale, for the advancement of other high-throughput technologies, which gave rise to a myriad of 'omics' datasets and scientific journals that are named after them. The remaining of this chapter provides an overview of the state-of-the-art technologies, their use in biological research, and the challenges as well as the opportunities they present to me as I started my graduate career.

# 1. The advent of high-throughput biology

## 1.1 Genomics

Launched in 1986, the Human Genome Project, funded primarily through the Department of Energy and the National Institutes of Health, set forth the goal to sequence and identify the functions of the 100,000 putative genes[1] identified in the human genome. The hope was that a physical mapping of the human genome would provide new avenues

for advances in medicine and biotechnology including the speedy identification of disease candidate genes, bio-markers, drug targets, and predisposition to illnesses as well as shedding new light on human evolution. Twelve years later, Celera Genomics was found based on Craig Venter's "whole-genome shotgun" technique to compete for essentially the same goal. This shotgun technique quickly became the most speedy and efficient, and thus standard, technology for genome sequencing. In 2001, both the Human Genome Project[2] and Celera Genomics[3] unveiled the first rough draft of the genome with much fanfare. This initial rough draft and its subsequent revision marked two major milestones in the quest for understanding the cellular system in its entirety. These drafts reveal that there are only 30,000-40,000 protein-coding genes, about twice many as in worm or fly, and make up only a few percent of the human genome. Though this estimated number of genes is far from final, the low gene count did end the speculation that human's superior intelligence is a result of the high number of genes in our genome. Unlike worms or flies, however, human genes tend to have small exons (approximately 50 codons on average) separated by highly variable intron lengths (from 40 bp to more than 10 kb), which makes it difficult to identify the location of open reading frames (ORFs). A substantial fraction of ORFs, more than the initial estimate of 35%, were found to contain alternative splicing. Alternative splicing allows multiple proteins to be encoded by the same ORF, and is thought to be used for complex gene regulation. It was also proposed that the high number of transcripts, or the complex transcriptional regulation allowed by alternative splicing, was the mechanism by which higher organism's complexity is generated. Further investigations by Brett and colleagues, however, showed that such a high number of genes containing alternative

splicing is not unique in human as compared to species such as worm, fly, mouse, and rat[4]. Therefore, the mystery about how the human "book of life" sets us apart from other animals continue to elude us[5]. Efforts to complete the genome sequencing is ongoing and attention is now turned more intensively to developing applications and uses for the data generated by the genome.

## 1.2 Transcriptomics or DNA microarrays

The development DNA microarrays or gene chips evolved naturally from the sequenced genomes. The technology relies on the presence of small DNA segments (probes) affixed on a solid surface, which then hybridize to fluorescently labeled cDNA produced from mRNA in the sample. By having hundreds of thousands of probes on a gene chip, microarrays facilitate the measurement of the genome-wide mRNA expression, hence the transcriptome, of all genes subject to a particular experimental condition (Figure 1.1). Assuming the flow of information from DNA to RNA to protein and function (the Central Dogma of Molecular Biology) is true for most protein-coding genes, the hope is that the knowledge of which genes are being transcribed gives clues about the functions of the involved genes as well as how the cell coordinates these genes to respond to a given stimulus. This was a major breakthrough in molecular biology and received considerable attention in the scientific community. Since the first report of a gene expression study with DNA microarrays in Science[6], approximately five thousand articles have been published on the use and analysis of microarrays. More recent studies have started to focus on standardization, statistical analysis, and data interpretation. The

current absence of standardization in arrays presents an interoperability problem, which hinders the exchange of microarray data. In effort to standardize microarray data, thereby increasing their accessibility, the Microarray Gene Expression Data (MGED) Society defines standards and guidelines, described under MIAME (Minimum Information About a Microarray Experiment), to facilitate unambiguous interpretation and improve reproducibility of microarray experiments[7].

The analysis of DNA microarrays also poses a statistical problem surrounding two issues: *i)* low signal-to-noise ratio resulting from low number of replications and *ii)* high rate of false positives resulting from the high number of genes per chip. The lack of sufficient replications of microarray data is due mostly to the relatively high cost of gene chips. With a low number of replicates (sometimes less than three per experiment), statistical confidence can not be assigned to expression results. The widely varied basal expression levels among the different genes exacerbates the low-signal-to noise ratio, making it difficult to estimate the level of up and down regulation of a particular gene relative both to its baseline and to other genes. Secondly, from a hypothesis-testing standpoint, the large number of genes present on a single array increases the false positive probability even when the likelihood of a gene yielding a result of interest is extremely low. To further complicate the matter, an additional problem that is often overlooked is that mRNA expression level is a weak indicator of the encoded enzyme levels[8]. The observed gene expression thus does not necessarily correlate with cellular functional activity, rendering microarray data uninformative in understanding cellular

physiology. Much effort in both the academic and industrial setting is now targeted at resolving these matters in microarray analysis.

## 1.3 Proteomics

In keeping with the spirit of genomics and transcriptomics, the field of proteomics aims to identify, quantify, and functionally characterize the entire set of proteins contained in a particular cell. Unlike the genome, which is unchanged across cell types and conditions, the proteome is a dynamic set of proteins. The proteomic composition varies with different cell types in a multicellular organism and is subject to modification based on the cell's external and internal conditions. Due to the problems discussed earlier with genomics and transcriptomics, the cellular proteome is expected to be instrumental in deciphering the cell's functional state as well as identifying biomarkers for pharmaceutical purposes. Proteomic studies include minimally two steps, protein separation followed by identification. One- and two-dimensional gel electrophoresis are frequently used to determine protein mass and isoelectric point. Proteins are then excised from the gel, enzymatically digested into smaller peptides, and sent to mass spectrometry for identification. Mass spectrometry ionizes and vaporizes the molecules in a magnetic field and then separates them according to their mass-to-charge ratios. Two soft ionization techniques primarily used for large biomolecules such as peptides are electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI). Tandem MS is now being used to identify the amino acid sequence of the peptide ions, which are used to query a database to determine the identity of the protein.

The major challenge encountered by proteomic studies is the wide range of size, pI, and hydrophobicity of the different proteins in the cell, making it impossible to preserve the integrity of the entire proteome in any particular separation condition. Furthermore, the digestion step makes it difficult to determine any posttranscriptional modification a protein carries. Due to these complications, the complete identification of the human cellular proteome will most likely not be completed until new technology is developed. In the mean time, encouraging success has been achieved for organelles (mitochondrion, lysosome, peroxisome, Golgi, etc) and large cellular structures (cytoskeleton, centrosome, mitotic spindle)[9].

## 1.4 Metabolomics and fluxomics

Metabolomics refers to the set of small molecules present in the cell as the cumulative result of the cell's genomic, transcriptomic, proteomic, and environmental stimuli. Metabolic profiling, enabled by MS and NMR, provides an instantaneous 'snapshot' of the functional state of that cell. The tradeoffs between MS and NMR are structural information and sensitivity. Higher sensitivity and coverage are achieved by MS, particularly with Fourier transform ion cyclotron resonance-MS (FTICR-MS) and capillary electrophoresis-MS (CE-MS). FTICR-MS is able to detect ion masses at extremely high accuracy and has been used to characterize cytochrome C variants[10], detect locations of sulfur-containing amino acids in peptides[11], and profile lipooligosaccharides of the gonorrhea bacterium[12]. CE-MS is highly sensitive to low

abundance metabolites and has been applied to identify nearly 2,000 metabolites in *Bacillus subtilis* extracts[13]. More recently, CE-MS has also been used to analyze polypeptide markers from patient's body fluids for diagnostic purposes[14]. The field of metabolomics is faced with similar problems as proteomics in moving towards a comprehensive dataset, or possible worse, as metabolites making up the metabolome are even more divergent in physical and chemical properties. It is thus the area with the most opportunities for improvement, particularly in sample purification and instrumentation development.

Co-evolving with metabolomics is the search for relationships among these metabolites and the rates by which they interconvert. In fact, the quantification of metabolic intermediates and reaction kinetics is the basic technique of classical biochemistry, which, in its larger scale, gives rise to metabolomics and fluxomics. Fluxomics is often defined as the large scale measurement of reaction fluxes in the cell. The word "measurement" is in fact a misnomer because reaction kinetics, or fluxes, are never measured directly; they are inferred through the concentrations (either relative or absolute) of reaction participants. Therefore, fluxomics is essentially a subcategory of metabolomics with an emphasis on quantification. More recently, fluxomic data have been obtained through the use of stable isotopes, particularly those containing $^{13}C$ labels, and the application of MS or NMR to determine the eventual fate of these labels. Understandably, this field of study has been referred to as both fluxomics and tracer-based metabolomics in the literature. The relative distribution of $^{13}C$ among the different metabolites and the different mass isotopomers of a metabolite reflect the rates of

reactions producing and consuming them. Mass isotopomers of a metabolite are isomers that differ only in the isotope distributions (in this case $^{13}C$ vs. $^{12}C$). The presence of multiple isotopomers of a metabolite is a clue that the metabolites are produced by multiple paths in the reaction network. Conversely, the quantification of reaction fluxes is dependent on a model network describing the cellular metabolism and the mathematical tools to calculate reaction fluxes based on tracer data.

## 2. The emergence of Bioinformatics and Systems Biology

The innovative high-throughput techniques, in some sense, can be considered *disruptive* technologies, as they force scientists to transform molecular biology into a systemic and quantitative discipline. The plethora of omics and legacy data has turned biology from a relatively data-poor to a data-rich discipline. The transformation of biological *data* into *knowledge* of cellular physiology is not trivial; it requires a major effort in organization and interpretation of such data. In fact, the launch of bioinformatics and systems biology as disciplines represents a concerted effort among scientists of different research areas to unite biological data, physical principles, and mathematical tools to unravel the complexity of living systems. As a result, *bioinformatics* and *systems biology* are widely described as interdisciplinary studies. However, I think Sean Eddy said it best when he referred to these fields of study as "antedisciplinary" science - the science that precedes the organization of existing disciplines[15]. Similar to biochemistry or bioengineering, it is my hope that, in time, these areas of study will become established and accepted disciplines in their own right. Being

young disciplines, the two areas of study do not have well defined, concrete descriptions. Nevertheless, it is generally agreed among scientists developing the field that bioinformatics refers to the creation and advancement of computational and statistical algorithms to solve theoretical and practical problems arising from the management and analysis of biological data. Systems biology, on the other hand, is characterized by the integration of experimental and computational methods and a philosophical shift from the more traditional reductionist to a more holistic approach in viewing biological systems. Systems biology and bioinformatics are often used synonymously due to the quantitative component they introduce to molecular biology. Nevertheless, most scientists who subscribe to the philosophy of systems biology agree that experimental study is an integral component of the discipline. In fact, successes in this field depend on the interplay of four processes: *i*) identifying key biological components, *ii*) reconstructing networks of interactions among these components, *iii*) quantitatively analyzing these networks, and *iv*) generating testable hypotheses for model validation and further experimental investigations. Together these four steps form a cycle (Figure 1.2) in which every iteration provides better understanding of the biological system than the sum of results individually collected from each process.

# 3. Network reconstruction and constraint-based modeling

## 3.1 Network reconstruction as a tool for data integration and analysis

The unifying theme across the 'omics' studies described here is the search for a physiological context of the cellular response to a particular condition. This theme

reflects the transition of biology from component identification to functional state characterization. The characterization of cellular functional states requires knowledge not only about the individual components and their functions, but also about the interactions among such components. Network abstraction is an obvious and intuitive way to represent these components and interactions. Networks of proteins and metabolites are the most common, and can now be reconstructed at the genome-scale[16-18]. The research described in the remaining chapters of this dissertation is concerned primarily with networks of metabolites (substrates) as nodes and biochemical reactions as edges. These networks are reconstructed through the curation of high-throughput and legacy biological data. High-throughput data used in these networks encompass genomics, DNA micro-arrays, proteomics, and tracer-based metabolomics. Legacy data include structures of biochemical pathways, reaction directionality, and substrate specificity of individual enzymes. Reconstructed networks thus serve as a highly organized repository for results of a large number of experiments, from which testable hypotheses can be drawn (Figure 1.2). Interrogating the properties of these networks allows one to evaluate their accuracy and functions.

## 3.2 Constraint-based modeling

Once an abstraction of the biological system has been accomplished with a reconstructed network, the next step in the systems biology cycle is the making of a predictive model. The use of simplified models to describe biological processes is not a new concept; however, it is not until recently that large-scale quantitative models have become accessible. Unfortunately, quantitative data are extremely scarce and

quantitative formalisms from which models can be built are even rarer in biology. The plea for an engineering approach in studying biology was eloquently and entertainingly made in Lazebnik's "Can a biologist fix a radio?" communication[19]. Though this scientist, in my opinion, made a rather unfair comparison between a radio and a cellular system, his point is legitimate: quantitative analysis is necessary to move biology forward in a meaningful and productive manner.

With respect to most metabolic systems, the focal concerns of this dissertation, quantitative analysis is often applied by relating the reaction rate to the change in concentrations of involved metabolites. The reaction rate primarily depends on the kinetics of the corresponding enzyme, which, in turn, depends on the enzymatic mechanism, regulatory state, and quantity of available enzyme molecules. A kinetics-based description is perhaps the most straight-forward, though not always the simplest or most feasible, approach to quantitatively model biological systems. It is thus recognized that kinetics-driven models, though useful, require a large number of system- and condition-specific parameters for which values are difficult and laborious to obtain. The lack of such data is a major hindrance to kinetics-based modeling, rendering it inaccessible for many systems of interest, particular those not amenable to experimental studies. An alternative, data-driven, constraints-based approach has been developed to partially overcome this difficulty. This modeling approach seeks to narrow the range of possible phenotypes a metabolic system can display by imposing constraints, rather than precisely determining the exact behavior of the system[20,21]. The constraint-based method is typically used in combination with a reconstructed network representing the system of

interest. Quantitatively, the reconstructed network can be represented by a stoichiometric matrix **S**, where each element $\mathbf{S}_{ij}$ represents the stoichiometric coefficient of metabolite *i* in reaction *j*, following the convention that $\mathbf{S}_{ij}$ is positive if the metabolite is the product of the reaction, and negative otherwise. Typical constraints in constraint-based modeling fall into two categories: *hard* and *soft* constraints. Hard constraints refer to the physico-chemical principles that all chemical reactions must obey. Examples are mass conservation[22], energy balance[23,24], and the laws of thermodynamics[22]. Soft constraints arise from the flexibility inherently built into a biological systems so that the organisms can adapt to various environmental conditions. Soft constraints therefore are system (cell type or organism) and condition (substrate availability, physiological state, presence of stimuli) dependent. Examples of soft constraints include regulatory rules and the goodness of fit with experimentally measured data. Soft constraints can be adjusted when comparing model predictions with experimental observations and likely lead to new discoveries when a disagreement is found. It should be noted that the constraint-based modeling approach is neither superior to nor can it completely replace its kinetic counter parts. Advantages and limitations of the constraint-based modeling approach are further discussed in chapter 3.

## 4. Dissertation overview

This introductory chapter aims to provide a layout of the biological and technological research landscape at the beginning of this century, which incidentally is about the time I started graduate school. What has been described here is the state of the

art science and opportunities I have been presented with to carve my own niche and contributions to the scientific community. As hopefully will become clearer in the chapters to come, biology is becoming a technology-driven discipline, and the philosophy of systems biology will likely become the standard practice to study living things. As technology improves, more data will be generated and at a faster rate, our understanding of biological systems will be limited only by our ability to analyze such data. The opportunity is clear: data must be converted knowledge. In working toward this goal, the research in this dissertation seeks to elucidate the metabolic phenotypes of the cells in normal and diseased states. The characterization of cellular metabolic phenotypes will place an emphasis on the biochemical pathways associated with mitochondria and the organelles' role in energy metabolism. Three main themes will be encountered: *i*) Data integration, *ii*) Data analysis, and *ii*) Data interpretation.

Specifically, chapter 2 details the process of metabolic network reconstruction as an effective and scalable method to organize and integrate 'omics' and legacy data for studying cellular metabolism. Three model systems are described in this chapter. The first model system is the mitochondrion, a semi-autonomous organelle that participates in many important cellular functions, is implicated in many diseases, and is the test system of many proteomic studies. Chapter 3 describes the various linear methods of analysis under the constraint-based modeling formalsim that can be applied to identify relevant metabolic steady states in the cell. These methods allow one to quickly and systemically evaluate the capabilities of the mitochondrial network under the constraints set by physical and chemical principles as well as metabolic disturbance. Chapter 4 describes

the process of incorporating carbon tracer data into the constraint-based framework so that physiological steady states, rather than theoretical metabolic capabilities, can be determined. Chapter 5 applies the method developed in chapter 4 to characterize the substrate distributions in the *ex vivo* mouse cardiomyocyte. Considerations for computational and experimental studies with tracer data are also discussed in this chapter. Chapter 6 arrives at the ultimate goal of this dissertation, which is the elucidation of cellular metabolic phenotypes in human health and disease. The first part of the chapter investigates the effects of external stimuli, namely changes in substrate availability, on the steady state assumed by HepG2 cells. The second part uses model predictions to identify the affected enzyme complex in fibroblasts derived from a patient with Leigh's disease. Finally the concluding chapter discusses the lessons learned from the computational and experimental methods undertaken this dissertation, my contribution to the field, as well as my prediction on the future of the bioinformatics and systems biology.

(This chapter, in part, is a preprint of the material appearing in "Building the power house: Recent advances in mitochondrial studies through proteomics and systems biology," Vo TD and Palsson BO. 2006. *Am J Physiol Cell Physiol*, in press. The dissertation author is the primary investigator and author of this paper.)

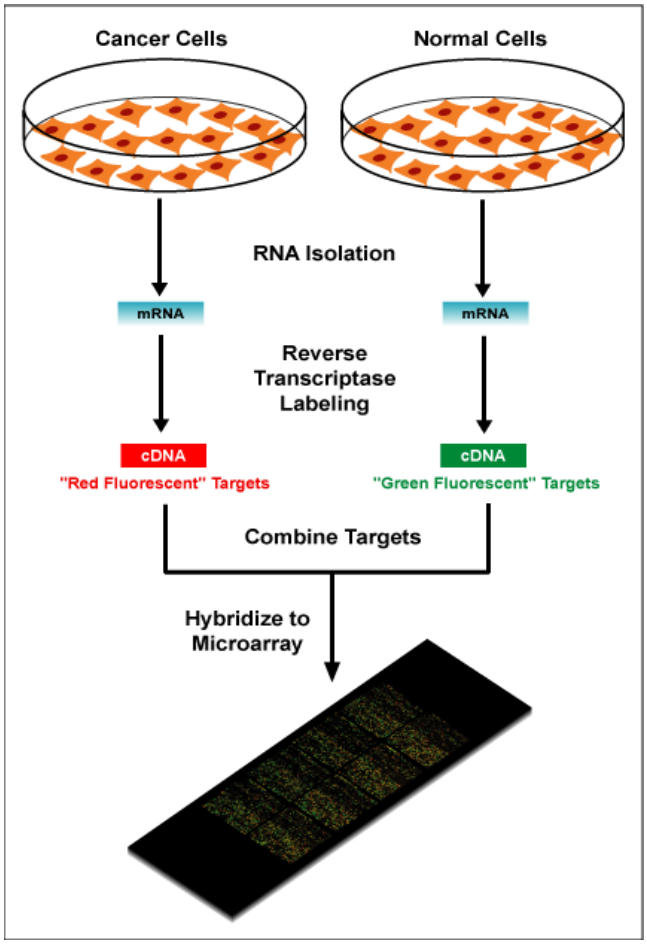



**Figure 1.1: Diagram of a typical dual-color microarray experiment**
mRNA from normal cells (control) and cancer cells are isolated and converted to cDNA that are differentially labeled with fluorescent tags. The cDNA from two samples are combined and are allowed to hybridize with DNA probes affixed on the chip. The combined intensities of the two fluorescent labels indicate the level of mRNA present in the original samples. This figure is reproduced from reference 25 with full permission of the author.

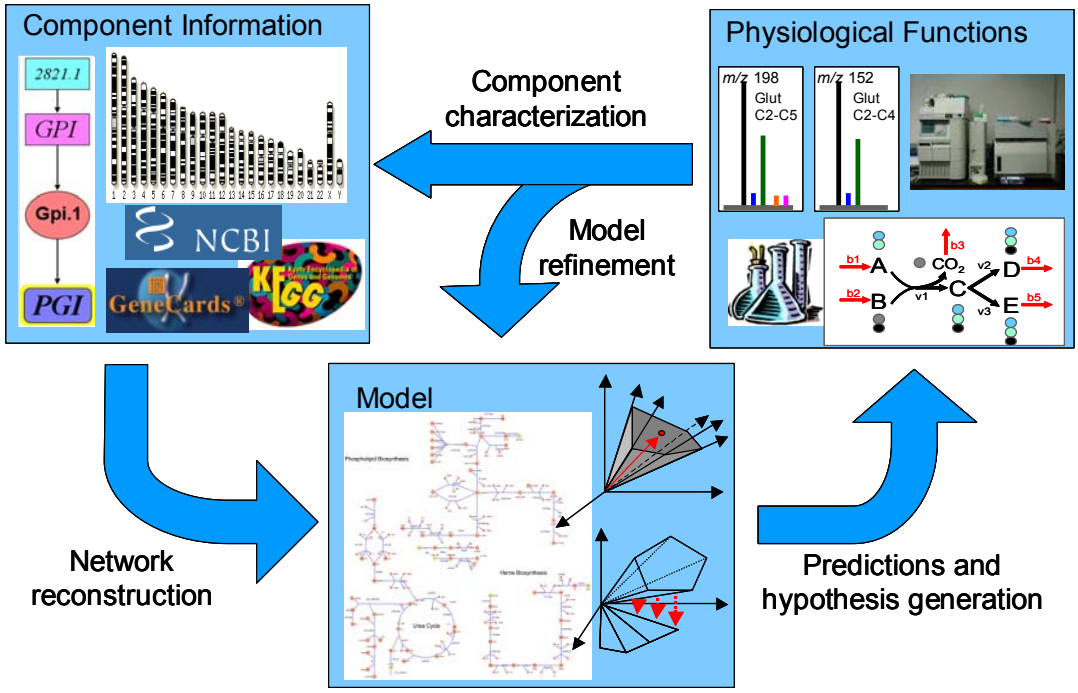



**Figure 1.2: The systems biology cycle**
Systems biology can be defined as the quantitative study of biological processes as whole systems instead of isolated parts. The field is characterized by the synergistic integration of data and theory which can be combined to produce a model. Analysis of the model provides predictions about physiological functions whose measurements are difficult or expensive to obtain. Validation of these predictions helps to identify novel components, which, in turn, refine the model.

# Chapter 2
# Metabolic Network Reconstruction as a Tool for Data Integration and Analysis

With the increasing availability of high throughput data, there is a growing need for integrating and reconciling these heterogeneous data sets to increase their consistency and reliability[26,27]. Metabolic network reconstruction has been referred to as "two-dimensional genome annotation" because it builds on the parts list of the genome (one-dimensional genome annotation) and brings together the connections amongst these components as an additional level of annotation. As these networks are often used as a basis for studying the genotype–phenotype relationship, it is important that they are biochemically and genetically accurate. Specifically, reconstructing a metabolic model includes two steps: *i)* identifying key components and *ii)* specifying biochemical reactions contributing to the metabolism of the system of interest (Table 2.1). Key components include some or all of the following: the model organism's annotated gene index, gene expression data, proteomic data, and legacy biochemical data. The specification of biochemical reactions relies mostly on the cumulative biochemical data from the primary literature and biochemistry textbooks. The present chapter is divided into three main sections. The first section provides a description about a variety of data types, their sources, and how they can be used to construct metabolic networks. The second section describes the steps and considerations to be taken in the reconstruction

process. The last section describes the contents of three metabolic networks whose properties and functions are studied in the remaining chapters of this dissertation.

# 1. Data Acquisition

## 1.1 Genome annotation

A large number of organisms have publicly available genomic sequences with an initial genome annotation. Genome annotation usually includes both experimentally verified functions of known proteins and putative functions annotated with *in silico* methods. These *in silico* methods are computational algorithms that predict the function of open reading frames (ORFs) in the genome. Both *ab initio* and comparative genomics approaches have been used for this purpose. *Ab initio* methods such as GLIMMER (Gene Locator and Interpolated Markov ModelER)[28,29] and GENSCAN[30] predict the location of an ORF based on the presence of regulatory signals such as promoters and stop codons. GLIMMER can find up to 99% of all bacterial protein-coding genes with high accuracy. Eukaryotic *ab initio* gene finders, on the other hand, have achieved only limited success due to the presence of long introns and high numbers of alternative splice sites[31] as discussed in chapter 1. On the other hand, comparative genomics approaches, also known as sequence-homology search tools, include BLAST[32-34], FASTA[35-37], and HMMER[38,39] rely on sequence similarity between closely related organisms to predict the function of a gene. It should be noted that gene annotation based on *ab initio* methods is subject to revision until the gene has been experimentally studied.

Acquiring a complete and annotated gene index of an organism is usually the first step in reconstructing the metabolic network for that organism. This type of data is relatively easy to obtain (if the genome of the organism of interest has been sequenced) and provides a comprehensive catalog of enzymes in the targeted network. Genomic data can be obtained from databases such as NCBI Genome, KEGG (Kyoto Encyclopedia of Genes and Genomes), TIGR-CMR (Comprehensive Microbial Resource), Ensembl (eukaryotic genomes), or directly from the group that sequenced the organism. The minimal information from a genome annotation necessary for the reconstruction includes locus number, gene name, and annotated gene functions. Commonly studied model organisms may have multiple gene indices corresponding to the various strains, and it is important to obtain the correct gene index for the strain of interest. The annotated gene index usually provides a good preview of the overall metabolic characteristics of the organism. Examples of characteristics to consider include preferred oxidative modes (anaerobic, aerobic, or microaerophilic), preferred substrates, and essential and non-essential amino acids.

## 1.2 DNA microarray data

DNA microarray data identify the set of genes that are being transcribed in a particular cell type and condition. As the number of microarray experiments grows, numerous databases have been established to archive and organize this information. Examples of these databases include NCBI Gene Expression Omnibus (GEO), ArrayExpress, Stanford Microarray Database (SMD), and Yale Microarray Database. In

most of these databases, users are allowed to query based on species, experiment type, experimental description, protocol, etc. Studies archived on these databases are also likely to be MIAME compliant[7,40]. The URLs of these databases change frequently, so users are advised to search for the databases by their names or by the research groups maintaining them.

Most microarray experiments contain at least two sets of data, one serving as a control and one subject to stimuli of interest. Since gene expression data are condition dependent, one should obtain the dataset that most resembles the condition to be modeled *in silico*. Given the often non-standardized experimental protocols and statistical difficulty affecting microdata interpretation[41], these data should be used conservatively if they are not generated in-house. For example, most expression studies report expression levels as fold change relative to a baseline at time zero or the control levels. One can use reasonable judgment or statistical analysis, if sufficient replicates are available, to determine whether a gene is present or absent in the condition of interest. This step allows the filtering of genes and enzymes that may not be relevant to the reconstructed network. Though this is a simple concept, determining whether a gene, protein, or reaction should be included in an *in silico* study is tricky for human cells, because the majority of human enzymes do not have a one-to-one association in the chain of information from ORF, transcript, protein, and reaction. Logic statements (section 2.1) are currently the most effective way to describe the relationships among these components. An example of how microarray data are used to identify relevant genes and enzymes for the human fibroblast-specific network is included in the last section of this

chapter. Additional research is underway to develop more sophisticated methods for the application and integration of microarray data for reconstructing purposes.

## 1.3 Proteomics

The proteome of a biological system (an organelle, a cell, or an organism) is characterized by its protein content, localization, and abundance. Proteomic data are thus specific to species, cell type, and experimental condition. Proteomic data are particularly important for studying systems such as mitochondria, where the majority of enzymes are not encoded in the organelle's genome, or for differentiating cells where not all genes are expressed at the same time. These data are typically more difficult to obtain than genomic data, but sub-cellular proteomes and proteomes of small bacteria have been identified[42,43]. Due to the greater complexity of these experiments, databases dedicated to proteomic data are less prevalent and less comprehensive than those for microarray data. Well established databases include Swiss Prot, Open Proteomic Database (OPD), and BIOBASE. Databases specific for subcellular structures are also available such as those for nuclei (the Nuclear Protein Database (NPD), Nuceolar Database) and mitochondria (The Human Mitochondrial Protein Database, MitoProteome, MitoP2, and MITOP). The minimal information needed from proteomic data for reconstructing purposes is gene name, encoded protein, localization, and description of protein function.

## 1.4 Biochemical data

Direct biochemical information is usually the most reliable. Biochemical characterization of an enzyme usually includes its substrate specificity, substrate

stoichiometry, and the physiologically preferred direction of the reaction(s) the enzyme catalyzes. Additional useful information about the enzyme is its gene locus, three-dimensional structure, subcellular localization, catalytic mechanism, kinetic parameters ($V_{max}$, $K_M$, $K_{cat}$, etc) and the change in Gibbs free energy accompanying the conversion of substrates to products. Collections of biochemical data on an organism's metabolism can be found in review articles, biochemistry textbooks, and volumes that focus on the biology of specific organisms. When inconsistencies are found among these sources, one is recommended to refer to the primary literature where the reaction is characterized. Lastly, review articles and textbooks dedicated to a particular organism or cell type are also likely to contain the organism or the cell's preferred substrates, minimal media composition, end products, etc, all of which are valuable for functionally characterizing the metabolic system.

# 2. Define molecular composition and interactions in the network

## 2.1 Define Gene-Protein-Reaction associations

The gene-protein-reaction associations describe the dependency of reactions on proteins and proteins on genes. Such dependency is straightforward if a particular gene encodes only one protein, which then catalyzes a substrate-specific reaction (Figure 2.1, panel 1). A small number of enzymes, particularly those acting on fatty acids or phosphoslipids, can act on a class of substrates that carry similar chemical groups. As each metabolite is a distinguishable compound in the reaction network, separate reactions are written for each of these substrates. Consequently, a separate association between the

enzyme and individual reaction is specified (Figure 2.1, panel 2). Conversely, the same reaction may be catalyzed by more than one isozyme. In these cases, separate associations are used to connect each isozyme to the same reaction.

Additional complexity arises when multiple transcripts can be derived from the same ORF or when the functional enzyme is made up of multiple subunits or proteins. When an ORF gives rise to more than one transcript, each transcript is uniquely defined and associated with the corresponding protein. The same ORF is connected to all of these transcripts. On the other hand, genes encoding protein subunits are all associated with one protein. They are connected by an "AND" association between the mRNA transcript and protein levels (Figure 2.1, panel 3). Protein complexes are different from proteins of multiple subunits in that each protein of the complex has a distinguishable function contributing to the overall catalytic function of the enzyme. As all proteins making up protein complexes are necessary for the reaction to occur, these proteins are connected by an "AND" association between the protein and the reaction levels (Figure 2.1, panel 4).

## 2.2 Define reactions

Accuracy in the biochemical reactions making up the metabolic network is essential, especially if the corresponding stoichiometric matrix is to be used in a predictive model (Chapter 3). Five pieces of information are necessary to describe each reaction: *i)* metabolite specificity, *ii)* metabolite formulas, *iii)* reaction stoichiometry, *iv)* reaction direction, and *v)* metabolite and reaction localization. The first piece of

information defines the reactants and products of the reaction. If an enzyme catalyzes multiple reactions, one should include all reactions that fall within the scope of the metabolic network. Second, each metabolite is to be further defined with the predominant elemental composition and charge at physiological pH. Note that the molecular formula of a metabolite at pH 7.2 (typical physiological pH) may not necessarily be the same as its neutral formula. For example, the neutral formula of lactate is $C_3H_6O_3$, but it is present as $C_3H_5O_2^-$ in most cellular environments. Third, once the elemental and charge formulas of all reaction participants are defined, it should be easy to determine the coefficients of each participant to make the reaction balanced. One must make sure that the reaction is both element and charge balanced; it is likely that $H_2O$ and $H^+$ need to be added to balance the equation. Fourth, thermodynamic properties and results of biochemical studies should be considered in defining the direction of a reaction. This is usually the trickiest step. Some reactions are essentially irreversible due to the large Gibbs free energy difference between reactants and product molecules. However, if this energy difference is sufficiently small, the direction of the reaction can be reversed by fluctuation in concentrations of reactants and products (temperature and pressure are relatively constant in most cellular environments). In these cases, the reaction is considered bidirectional or reversible. The reversibility of a reaction has implications for the metabolic capability of the cell or organism, and therefore reaction direction must be evaluated with care. Lastly, localization of the enzyme gives clues about the subcellular localization of reaction participants. Unless the enzyme is a transporter, reaction participants are assigned to the same location as the enzyme. If a metabolite is present in more than one cellular compartment, they are distinguishable

compounds, which may be interconverted by transporters if such transporters exist. Note that incorrect assignment of a reaction to a cellular compartment may lead to non-functional pathways or inaccurate network properties.

## 2.3 Determine reaction confidence level

Once a reaction is written, it is assigned a *confidence level* to reflect the type of data supporting its existence in the metabolic network of interest. A four-level scoring system for evaluating the confidence level is used in all metabolic networks discussed here (Table 2.2). The criteria for each level are *i*) Biochemical data (score of 4), *ii*) Genetic or proteomic data (score of 3), *iii*) Genomic or physiological data (score of 2); and *iv*) Modeling data (score of 1). Biochemical data, as described in section 1.4, result from experimental characterization of a chemical reaction. The characterization should contain information about the substrates, products, and stoichiometry of the reaction. Direct biochemical data include the identity of the enzyme and indirect data demonstrates the occurrence of the reaction in a cellular extract. Genetic data contain information relating to the experimental identification or genetic characterization of the gene coding for the enzyme, thereby suggesting the existence of the corresponding reaction. Genomic data refer to the putative genes annotated through *in silico* methods and are thus linked to the protein catalyzing the reaction. Physiological data are information about the overall cellular physiology, fitness characteristics, substrate utilization, or phenotypic results that suggest the presence of the biochemical event. Finally, modeling data are purely hypothetical reactions predicted from *in silico* analysis. These are often reactions

required for the model to fulfill certain metabolic demands. The scores assigned to each data type allows the numerically evaluation the overall reliability of a metabolic network.

## 2.4 Assemble metabolic pathways and construct the stoichiometric matrix

Once the list of genes and proteins specified by genomic or proteomic data is translated to a list of reactions, the next step is to assemble them into a complete biochemical pathway. This step is necessary to ensure that the resulting metabolic network is functional. Review articles or books written specifically about the metabolism of the organism or cell of interest can be helpful in determining which metabolic functions the system is capable of. If such reviews or books are not available, the KEGG database can be a useful resource to reference canonical biochemical pathways. Note that not all enzymes or reactions constituting these pathways are present in any single system. One should start with metabolic pathways corresponding to oxidation of preferred substrates of the organism as these are likely the most well studied pathways in the literature. Oxidation of other substrates, followed by biosynthetic pathways, can be built from there. Catabolic pathways allow the metabolic network to produce cellular currency such as ATP, NADH and NADPH, which are necessary for biosynthetic pathways to be functional. In assembling these pathways, one may find gaps, or missing enzymes, compared to canonical biochemical pathways. These gaps may be real, i.e. the organism is incapable of carrying out the corresponding metabolic function, or they may represent missing knowledge about that pathway. Such gaps are ideal candidates for further experimental studies[18].

Once all pathways making up the organism's or the cell's entire metabolic network are assembled, technically one has arrived at a reconstructed metabolic network. However, in order for quantitative results to be extracted from the model, the metabolic network is usually converted to a matrix, which records the reaction participation of every metabolite in every reaction in the network. Specifically a reconstructed network can be represented by a stoichiometric matrix **S** ($m$ x $n$), where $m$ is the number of metabolites and $n$ is the number of reactions[22]. Each element $\mathbf{S}_{ij}$ represents the coefficient of metabolite $i$ in reaction $j$, following the convention that $\mathbf{S}_{ij}$ is positive if the metabolite is the product of the reaction and negative otherwise. A zero entry indicates the metabolite does not participate in the reaction (Chapter 3). Reversible reactions can be written in either direction.

# 3. Metabolic networks studied

## 3.1 The mitochondrial metabolic network

The most complete catalog of the human mitochondrial proteome to date identifies 615 proteins, of which 298 are assigned to 153 unique enzymatic reactions[42] in the human cardiac myocyte. Since the catalogued proteome may be incomplete[44], metabolic reactions catalyzed by enzymes found in the proteome are augmented with biochemical data from the primary literature to form an initial reconstruction of the human mitochondrial metabolic network. This network is comprised of 189 metabolic reactions and 230 metabolites in three cellular compartments. The included reactions

describe the tricarboxylic acid cycle (TCA), oxidative phosphorylation (OxPhos), fatty acid β-oxidation, phospholipid biosynthesis, urea cycle, and reactive oxygen species detoxification (Table 2.3). Glycolytic reactions are also included because they were identified in the mitochondrial proteome and seem to be physically associated with the mitochondria[45]. The complete contents of the metabolic network can be found in Vo *et al.* [46].

Every metabolite and reaction in the network is localized to a cellular compartment. The three compartments included in the model are the mitochondrial matrix, the cytosol, and the extracellular space. Many of the included pathways span more than one compartment (Table 2.3). The mitochondrial inter-membrane compartment is not explicitly accounted for since most metabolites under 10 kD freely travel across the outer mitochondrial membrane, and therefore are considered to be in equilibrium with the cytosol[47]. Metabolites are also characterized by their molecular formulas and predominant charge forms determined at pH 7.2. Each reaction is thus elementally and charge balanced. Figure 2.2 describes the level of detail contained in each reaction. The present reconstruction contains 88 reactions annotated with a score of *4*, 40 with a score of *3*, 31 with a score of *2*, and 30 with a score of *1*. Lumped reactions and exchange reactions between the cytosol and the extracellular space are assigned a score of *1*. Functional characterization of this network can be found in chapter 3 and Vo *et al.* [46].

## 3.2 The hepatocyte metabolic network

HepG2, derived from a liver carcinoma, is an adherent cell line widely used as models for hepatocellular carcinoma[48-50]. Hepatocytes make up 60-80% of the cytoplasmic mass of the liver. These cells are very metabolically active, and participate in many processes such as carbohydrate transformation; synthesis of protein, cholesterol, phospholipids, and bile salts; and detoxification of exogenous and endogenous waste. The major pathways involved in these processes are well studied and reactions making up the pathways are available in the literature. The reconstructed network for HepG2 cells are thus assembled using the annotated human genome and primary literature. The network, containing 254 reactions and 269 metabolites, describes glycolysis, the TCA cycle, the pentose phosphate cycle (PPC), OxPhos, amino acid metabolism, β-oxidation, porphyrin metabolism, the urea cycle, palmitate biosynthesis, and phospholipid synthesis. These pathways are assembled using canonical pathway structures from KEGG and biochemistry textbooks. Every reaction in the network is mass and charge balanced as described above. The complete contents of this network can be found in Vo *et al.* [51].

## 3.2 The fibroblast metabolic network

The complete (general) human metabolic network[16] is used as a basis for the reconstruction of the fibroblast-specific metabolic network. DNA microarray data[52], downloaded from the NCBI GEO database, are used to identify fibroblast-specific genes. The samples for the microarray experiments are dermal fibroblasts collected from apparently normal donors of similar ages. Expression data were measured based on the

33,000 annotated genes on the Affymetrix U133A/B chips using three different normal fibroblast cell lines (GM00038C, GM00316B, and GM08398C)[52]. The published *present* and *absent* calls are analyzed to identify expressed genes, thus "active" enzymes, in the fibroblast network. Using the gene-protein-reaction associations built in the general human metabolic network[16], active enzymes are translated to active reactions to be included in the fibroblast-specific metabolic network. A total of 1690 reactions are identified using this method. This list of reactions is filtered further to reduce the network size and scope so the computation associated with isotopomer analysis is feasible in practical time. The resulting network contains only reactions occurring in the mitochondrial matrix and cytosol and transporters associated with these compartments. The final model network contains 430 metabolites and 508 reactions, describing glycolysis, the TCA cycle, the pentose phosphate cycle (PPC), oxidative phosphorylation, the malate-asparate shuttle, β-oxidation, *de novo* fatty acid synthesis, phospholipid biosynthesis, ketone body metabolism, and amino acid metabolism. Every reaction in the network is mass and charge balanced as described above. The complete contents of this network can be found in the reference [53].

(This chapter, in part, is a reprint of the material appearing in "In silico Analysis of SNPs and Other High Throughput Data", Jamshidi N, Vo TD, and Palsson BO. 2006. *Cardiac gene expression: methods and protocols*, Totowa, NJ: Humana Press, in press; "Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data," Vo TD, Greenberg HJ, and Palsson BO. 2006. *J Biol Chem* 279(38):39532-40; "Isotopomer analysis of cellular metabolism

in tissue culture: a comparative study between the pathway and the network-based methods", Vo TD, Lim SK, Lee PWN, and Palsson BO. 2006. *Metabolomics*, in press; and "Systems analysis of energy metabolism elucidates the affected respiratory chain complex in Leigh's syndrome", Vo TD, Lee PWN, and Palsson BO, *Molecular genetics and metabolism,* submitted. The dissertation author is the primary investigator and author of these publications.)

**Table 2.1: Steps in the reconstruction process**

**1. Data acquisition**

1.1. Acquire an annotated gene index for the organism of interest.
Minimal information necessary for the reconstruction includes locus number, gene name, and annotated gene functions. *Sources*: NCBI Genome, KEGG (Kyoto Encyclopedia of Genes and Genomes), TIGR-CMR (Comprehensive Microbial Resource), Ensembl (eukaryotic genomes), and research groups sequencing the organism.

1.2. Acquire DNA microarray data
Data should be MIAME compliant and relevant to systems and conditions of interest. Gene expression is used to determine whether the reaction catalyzed by the encoded proteins should be included in the reconstructed network. *Sources*: NCBI Gene Expression Omnibus (GEO), ArrayExpress, Stanford Microarray Database (SMD), Yale Microarray Database, and primary literature.

1.3. Acquire proteomic data
Data should be specific to species, cell type, and experimental condition of interest. Minimal information needed includes gene name, encoded protein, localization, and description of protein function. *Sources*: Swiss Prot, Open Proteomic Database (OPD), BIOBASE, the Nuclear Protein Database (NPD), Nuceolar Database, the Human Mitochondrial Protein Database, MitoProteome, MitoP2, and MITOP.

1.4. Review biochemical data
Biochemical characterization of an enzyme usually includes its substrate specificity, substrate stoichiometry, and the physiologically preferred direction of the reaction(s) the enzyme catalyzes. Additional useful information about the enzyme is its gene locus, three dimensional structure, subcellular localization, catalytic mechanism, kinetic parameters ($V_{max}$, $K_M$, $K_{cat}$, etc) and the change in Gibbs free energy accompanying the conversion of substrates to products.

**2. Define molecular composition and interactions in the network**

2.1. Define gene-protein-reaction associations
Use genome annotation to identify gene locus, proteomic data to identify encoded proteins, and biochemical data to identify the biochemical reaction.

2.2. Define reaction by specifying
- 2.2.1. Metabolite specificity
- 2.2.2. Metabolite formulas
- 2.2.3. Reaction stoichiometry
- 2.2.4. Reaction direction
- 2.2.5. Metabolite and reaction localization

2.3. Determine reaction confidence level

2.4. Assemble metabolic pathways and construct the stoichiometric matrix
- 2.4.1. Refer to the KEGG pathway or biochemistry textbooks for canonical pathway structures
- 2.4.2. Assemble reactions identified based on acquired data, start with oxidative pathway of the organism's preferred substrates and extend the network to biosynthetic pathways
- 2.4.3. Resolve gaps: if a reaction is missing from the organism's data as compared to the canonical pathway, investigate if physiological or phenotypic data are available to support the function of such a pathway. If such data are available, include the reactions and assign them a confidence score of 1.

**Table 2.2: Examples and guidelines for evaluating the confidence level of reactions**

| Level | Data | Example | Reason |
|---|---|---|---|
| 4 | Biochemical | Citrate synthase | The enzyme has been purified and characterized, but crystal structure is not necessary. The mechanism and/or kinetics of the enzymatic reaction have been shown. |
| 3 | Genetic or proteomic | Inorganic pyrophosphatase | The gene has been isolated and the corresponding protein was found in mitochondrial proteome, but the mechanism of the catalysis is not well understood. A level of 3 is also assigned to genes/proteins that had biochemical data but were from tissues other than cardiac. |
| 2 | Physiological or sequence homology | Pyruvate mitochondrial transport | Activity of the enzyme has been observed, but no gene has been identified. |
| 1 | *in silico* | Glycerol transport | There has been no documentation on glycerol is transport from the extracellular to the cytosol. It is possible that glycerol is also produced in the cytosol itself, but the biosynthesis of this molecule is outside the scope of this reconstruction. This reaction is included because the presence of glycerol is required for the reconstructed mitochondrial network to produce phospholipids. |

**Table 2.3: Summary of the composition of the mitochondrial metabolic network**

| | Number of reactions | Compartment |
|---|---|---|
| Glycolysis | 12 | cytosol |
| Citric acid cycle | 10 | mitochondrial matrix |
| Oxidative phosphorylation | 6 | mitochondrial matrix |
| ROS detoxification | 9 | cytosol, mitochondrial matrix |
| Fatty acid oxidation | 31 | mitochondrial matrix |
| Phospholipid biosynthesis | 16 | mitochondrial matrix |
| Urea cycle | 8 | cytosol, mitochondrial matrix |
| Porphyrin biosynthesis | 12 | cytosol, mitochondrial matrix |
| Transport and others | 85 | cytosol, mitochondrial matrix, external |
| Total network | 189 | cytosol, mitochondrial matrix, external |

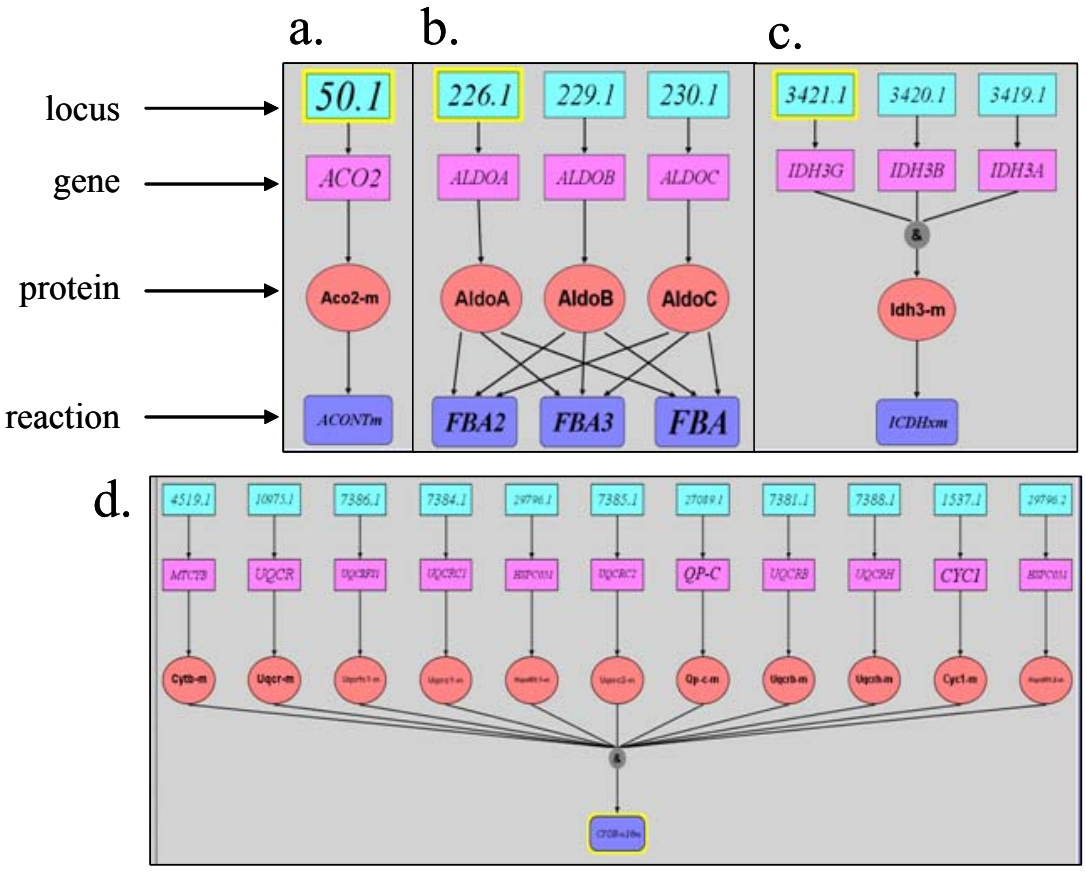



**Figure 2.1: Gene-protein-reaction association**
Panel (a) shows an example where there is a one-to-one relationship among locus, gene, protein, and reaction. Panel (b) shows a case where an enzyme can catalyze multiple substrates and/or reactions. Panel (c) shows the mitochondrial protein Idh3-m is made up of three subunits coded by three separate genes. Panel (d) shows an example where an entire protein complex is necessary to carry out the specified reaction.

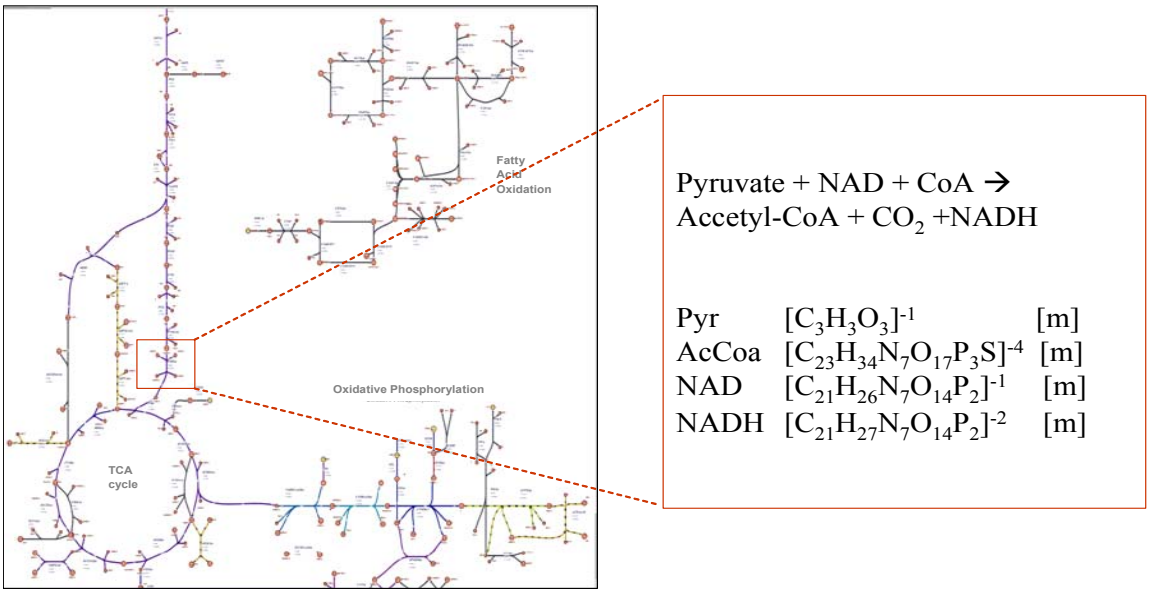



**Figure 2.2: Characterizing reactions in the reconstruction process**
The left panel show a section of the mitochondrial metabolic network. Each reaction, pyruvate dehydrogenase as an example shown to the right, is characterized with metabolite molecular formulas, charge, compartment, and reaction direction.

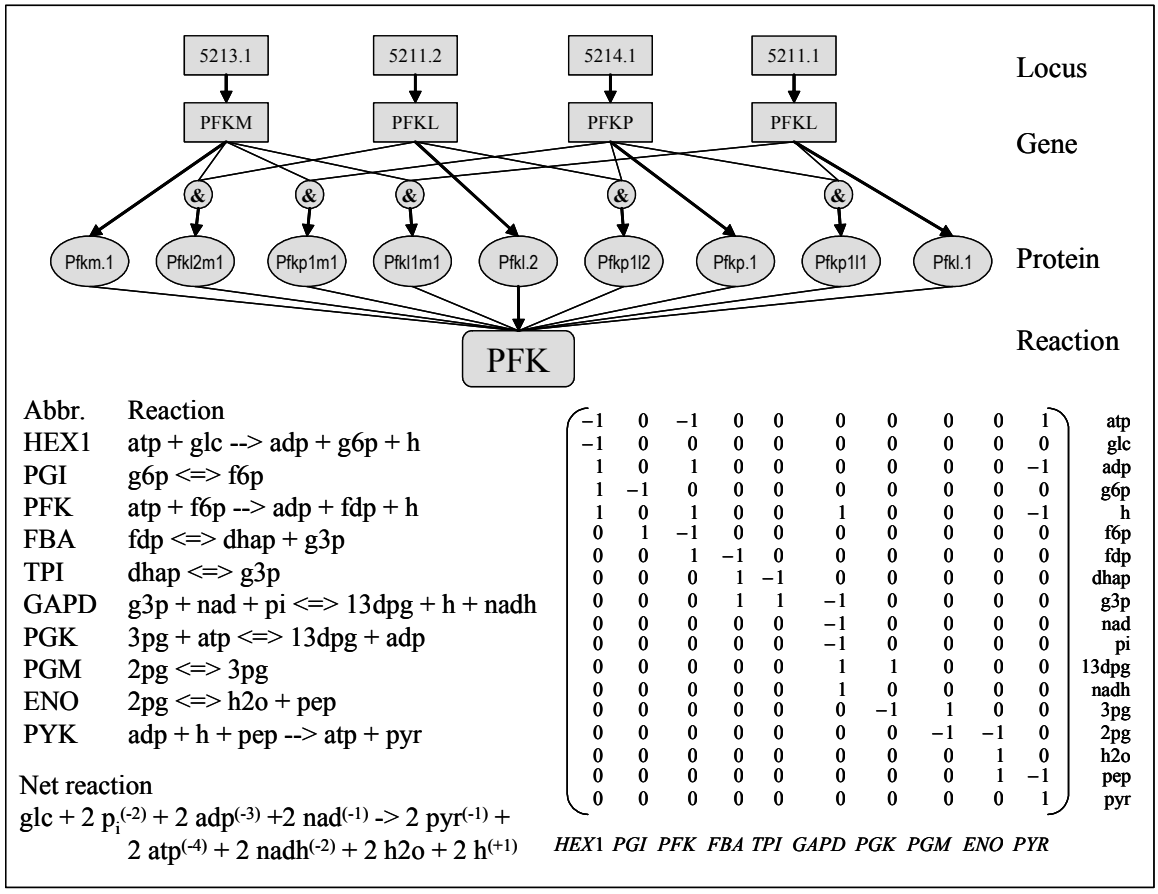



**Figure 2.3: Network reconstruction as applied to the glycolytic pathway**
A reconstruction of a biochemical pathway starts with identification of enzymes, metabolites, and directionality of each reaction. The list of reactions in the glycolytic pathway is shown, where each reaction is labeled with an abbreviation of the corresponding enzyme name. Each reaction is written to balance elements and charges of reactants on both sides. The net reaction shows a gain of two protons, which can possibly contribute to the proton gradient across the mitochondrial membrane. The S matrix on the right hand side is the mathematical representation of the list of reactions on the left. Each element Sij denotes the participation of metabolite i in reaction j. The reconstruction also directly integrates associations of locus, gene, protein, and reaction for each enzyme when possible. An example is shown for the enzymes phosphofructosekinase (PFK). PFK is made up of four subunits and has tissue variations (muscle, liver, platelets) as well as transcript variations (locus 5211 has two alternative splices, 5211.1 and 5211.2). The muscle and liver isoforms are homotetramers of the M (PFKM) and L (PFKL) subunits, respectively, while platelets contain various heterotetramer combinations such as three PFKL and one PFKM or two PFKL and two PFKM[54].

# Chapter 3
# Functional Characterization of the Cardiac Mitochondrial Metabolic Network

In order for a metabolic network to become a predictive model, rules must be established to describe the behaviors of the network. As briefly described in Chapter 1, the constraint-based modeling approach is an effective method to eliminate unattainable metabolic states, thereby narrowing down or identifying theoretically possible and/or physiological network behaviors. This chapter is divided into three main parts. The first part describes the constraint-based modeling formalism, its advantages and limitations (Section 3.1). The second part applies this framework to identify theoretically allowable metabolic state of the cardiac mitochondrial metabolic network constructed using proteomic data (Section 3.2-3.7). The last part discusses the application of Monte-Carlo sampling in identifying feasible flux distributions in the steady-state flux space. Segmentation of the solution space, applied through constraints on extracellular and intracellular reactions, is used to study metabolic steady states in *i)* normal physiological conditions (heart at rest), *ii)* ischemia, and *iii)* diabetes (Section 3.8-3.10).

## 1. Constraint-based modeling formalism

A constraint-based model can serve as a model-centric database that provides quantitative predictions. The constraint-based modeling approach involves the

application of a series of constraints arising from stoichiometry considerations, thermodynamics, enzymatic capacities, and regulatory and kinetic constraints when they are available. This method starts by identifying the molecular composition of the biological reaction network and then defining constraints on their interactions. The key components of a metabolic model are metabolites and enzymes (Chapter 2). Constraints ensure the predicted functional states of the model network adhere to the fundamental laws of physics and chemistry. A biochemical reaction can be viewed as a conversion of substrates into products by the action of an enzyme, whose activity can be described by a rate law. Therefore the change in concentration of a particular metabolite over time is equal to the sum of all fluxes producing the metabolite subtracting the sum of fluxes consuming that metabolite. Under the steady-state assumption, the rate of consumption of every metabolite equals its rate of production. This mass conservation relationship translates to a system of ordinary differential equations:

$$\frac{d[X]}{dt} = \mathbf{S} \cdot v = 0 \qquad \text{(Equation 3.1)}$$

where X is the metabolite vector (length $m$), $v$ is the flux vector (length $n$), and S is the $m \times n$ stoichiometric matrix consisting of the appropriate coefficients for all reactions participating in the network (Figure 3.1). Each element $S_{ij}$ represents the coefficient of metabolite $i$ in reaction $j$. By convention, the coefficient of a metabolite is positive if it is the product of the forward reaction, and negative otherwise. Solutions for $v$ are systematically determined by successive application of additional constraints, such as those representing directional (reversible vs. irreversible) and enzymatic capacity considerations. These constraints have the form:

$$\alpha_i \leq v_i \leq \beta_i \qquad \text{(Equation 3.2)}$$

where $\alpha_i$ and $\beta_i$ represent lower and upper bounds on the steady-state rate of each reaction. Maximum flux values ($\beta$) can be estimated based on enzymatic capacity limitations or maximal measured uptake rates for transport reactions. The lower limit ($\alpha$) is zero for irreversible reactions.

Energy-balance constraints have also been developed to disallow fluxes in thermodynamically infeasible internal reaction cycles[23,24]. These bilinear constraints introduce a second set of variables, stored in the vector $\Delta\mu$, which represent the change in chemical potential associated with reactions in the network:

$$\mathbf{K} \cdot \Delta\mu = 0 \qquad \text{(Equation 3.3)}$$

$$v \cdot \Delta\mu < 0$$

The matrix **K** (Equation 2) stores the null space basis for a matrix **Z**, where **Z** contains rows and columns in **S** corresponding only to internal reactions. The resulting *solution space* (satisfying Equations 3.1-3.3) often contains a range of possible values, rather than a unique number, for each reaction rate $v_i$ that satisfies the stated constraints. The term solution space is thus used synonymously with the term "steady state flux space" in this dissertation.

Equations 3.1 and 3.2 are commonly applied to study metabolic states of cellular systems due to their simple mathematics and guaranteed solutions. Such simplification is based on two fundamental assumptions which must be discussed. First, it is assumed that

all biochemical reactions and physical interactions can be written as equations with known participants. Second, the flux distribution $v$ calculated based on these two equations assumes a steady state where the change in concentration of every chemical species in the network is approximately zero. Rationales and ramifications associated with these two assumptions are as follows. The first assumption is straight-forward for most biochemical reactions where reactants and products are well defined. However, participants and stoichiometry of reactions of interactions involving signaling molecules, activation and inactivation of enzymes, and voltage-gated responses are frequently ill defined, making it difficult to incorporate them into the stoichiometric matrix. Secondly, there are two implications of the steady state assumption: *i)* There must be no internal build-up of metabolites in the cell so that the mass conservation equation (Equation 3.1) holds perfectly and *ii)* Observable phenotypes or biological phenomena of interest occur at a time scale longer than the rate of metabolites being produced and consumed by reactions in the network. Consequently, Equation 3.1 and the mentioned constraints only apply to a subset of biological systems and only at a time scale satisfying assumption (*ii)*. Specifically, these equations are most appropriately used to investigate non-transient phenotypes such as metabolic steady states, gene essentiality[17], and end points of adaptive evolution[55] or time-invariant qualities such as network topology[56] and metabolite pool identification[57]. Successes in these studies have clearly demonstrated the strength of the constraint-based formulation in studying such biological behaviors[58,59]. In contrast, the steady state assumption precludes the use of constraint-based methods to study concentration-dependent behaviors associated with transient or periodic dynamics observed in regulatory (activation, inhibition, feed-back) or signaling responses by

neuronal and muscular cells. For example, even when the components and interactions of the JAK-STAT signaling network are painstakingly identified, only topological characteristics can be satisfactorily analyzed[60]. With respect to mitochondrial systems, the constraint-based approach, in its present definition, can not be used to model oscillatory behaviors resulting from muscle excitation[61,62] but is ideal for studying metabolic disturbance due to enzymatic defects[63]. It should be noted that the limitations discussed here are associated only with the use of Equation 3.1 and are not inherent in the constraint-based methodology. In other words, the constraint-based approach allows the use of Equation 3.1, i.e. the omission of kinetic data, but it does not preclude the use of such data if they are available. In fact, if such data are available, they can be formulated into constraints to further resolve flux calculation. On the other hand, since the constraint-based approach was developed to circumvent the need for kinetic data and the nonlinearity in kinetic rate laws, most researchers only apply the constraint-based method when they do not have kinetic data and/or want to simplify the calculation. Thus, though these assumptions are not a direct result of the constraint-based methodology but have always been associated with it.

## 2. Methods of analysis under the constraint-based framework

The methods of analysis described in this section are applied to analyze functional states of the mitochondrial network. Specifically, methods described in sections 2.1-2.5 are used to study the capabilities of the network (Sections 3-5) and the Artificial Centering Hit-and-Run algorithm[64] is used to sample the entire solution space (Sections 6-9).

## 2.1 Optimization for a metabolic function of interest: linear programming

Linear programming (LP) is frequently used to find a flux distribution that maximizes a particular metabolic objective function[21,46,65]. The underlying assumption of this analysis is that biological systems have evolved the ability to allocate cellular resources to maximize for certain end products (Figure 3.2). The general problem formulation has the form

Maximize $f(v)$

Subject to $\mathbf{S} \cdot v = 0$ (Equation 3.4)

$\alpha_i \leq v_i \leq \beta_i$

The three objective functions considered for the mitochondrial network are

ATP production

$f_1(v)$ = -1 ATP -1 $H_2O$ + 1 ADP + 1 $P_i$ + 1 $H^+$

Phospholipid biosynthesis

$f_2(v)$ = 0.18 cardiolipin + 0.34 phosphatidylethanolamine + 0.43 phosphatidylcholine

Heme biosynthesis

$f_3(v)$ = protoheme

The first objective function is written as an ATP hydrolysis reaction, so that the network can recycle ADP and phosphate. The coefficients in the second objective function are derived from the phospholipid composition of the mitochondria 16-18. The last objective function is simply the production of protoheme. Network reconstruction

and flux balance analysis are done using the software SimPheny (Genomatica, San Diego, CA).

## 2.2 Flux variability analysis

To explore the range of allowable flux of each reaction in the network, one sets the physiological objective functions (ATP production, heme biosynthesis, and phospholipid biosynthesis) to their respective optimal values, then maximizes and minimizes the flux through each reaction[66]. In a particular calculation, one of the three physiological objective functions is set to its optimal value; the other two are set to zero. Two LP problems (maximizing and minimizing) are solved for each reaction in the network. All other constraints regarding mass balance, reversibility, and enzymatic capacity are kept the same.

Thermodynamically infeasible cycles in the network, identified using the extreme pathway algorithm[67], have no net flux[23,66]. Since these cycles are never used *in vivo*, one reversible reaction from the cycle is removed such that it does not disconnect any metabolite from the rest of the network. For this study, two reactions (ORNt3m and CITRtm) are removed for this purpose. All calculations for flux variability analysis are implemented in GAMS (GAMS Development Corporation, Washington, DC) using the CPLEX solver and independently checked with the Mosek solver[68].

## 2.3 Identification of alternate optima

Alternate optima refer to a set of equivalent paths the network may use to optimally satisfy an imposed metabolic demand. They lie at the corners of the intersection between the solution space and the hyperplane created by the objective function. Current linear programming solvers cannot provide alternate optima automatically; this must be done after finding one optimal extreme point and executing a vertex-enumeration algorithm[69,70]. The algorithm to enumerate all alternate optima involves the following three steps:

1) A linear programming formulation of the stoichiometric matrix, constraints, and objective functions as described in the constraint-based modeling section is implemented in MATLAB:

   Maximize $f(v) = c^T v$ (Equation 3.5)

   Subject to $\mathbf{S} \cdot v = 0, \alpha_i \leq v_i \leq \beta_i$

   where **S** is the stoichiometric matrix, $v$ is the flux vector, c corresponds to one of the three objective functions listed in Section 2.1, $\alpha_i$ and $\beta_i$ are the respective upper and lower bounds on the reaction flux $v_i$. Mosek[68], an optimization system available for use with MATLAB interface, is applied to solve this LP problem for an initial optimal extreme point $v^o$.

2) To reduce the dimension of the problem, reactions that must have a fixed flux value (identified from flux variability analysis) are considered to be constants. To enumerate only basic optimal flux distributions, one imposes $f^* = c^T v$, where $f^* = c^T v^o$, as an additional constraint.

3) The last step passes the set of inequalities of the optimal solution space, F* = *v*, to *lrs* to obtain its corresponding vertex representation, and hence its number of extreme points. Since *lrs* requires the origin to be an extreme point, F* is translated to the one computed extreme point, $v^o$, to obtain the set of change-vectors: D = (d: $v^o$+d in F*). Then, if ($d_k$) are the extreme points of D, ($v^o$+$d_k$) are the extreme points of F*.

Details of how *lrs* enumerates the extreme points can be found in the references[69,70]. It should be noted that, while this method works well here, tighter bounds may cause a combinatorial explosion in the number of extreme points, making complete enumeration impractical.

## 2.4 Multiple-objective analysis

The three objective functions can be considered simultaneously by using the concept of *Pareto optimality*. A feasible flux *v* is dominated by another feasible flux *v*' if $f_i(v') \geq f_i(v)$ for all objective functions $f_i$, and $f_i(v') > f_i(v)$ for at least one $f_i$. A feasible flux is Pareto optimal if it is not dominated by any other feasible flux. The set of Pareto optimal fluxes is called the *efficient frontier*.

Two methods to compute Pareto optima are weights and lexicography. The weight formulation allows one to designate a composite objective function that includes all three mitochondrial functions, each weighed by a positive coefficient $w_i$. All

members of the efficient frontier can be found by maximizing an associated weighted sum, $\Sigma_i w_i f_i(v)$, where the weight vector $w > 0$. Alternatively, the objectives can be hierarchically ordered to produce a sequence of optimization problems. The ordering gives a *lexicographic-max* of the vector $f = (f_1, f_2, f_3)$. Here, if the order of the objectives is such that $f_1$ = ATP demand, $f_2$ = heme, and $f_3$ = phospholipids:

$$f_1^* = \max\{c_{atp}^T v : \quad S \cdot v = 0, \ \alpha_i \le v_i \le \beta_i\} \qquad \text{(Equation 3.6)}$$

$$f_2^* = \max\{c_{heme}^T v : \ S \cdot v = 0, \ \alpha_i \le v_i \le \beta_i, \quad c_{atp}^T v = f_1^*\} \qquad \text{(Equation 3.7)}$$

$$f_3^* = \max\{c_{phos}^T v : \ S \cdot v = 0, \ \alpha_i \le v_i \le \beta_i, \quad c_{atp}^T v = f_1^*, \quad c_{heme}^T v = f_2^*\} \qquad \text{(Equation 3.8)}$$

The fluxes in the last optimal set are Pareto optimal.

## 2.5 Sampling of the steady-state flux space

Sampling the steady-state flux space is performed using a random walk algorithm (Artificial Centering Hit-and-Run, ACHR) as described by Kaufman *et al.*[64]. The algorithm involves three steps. The first step identifies an initial point by reducing each of the upper bounds and increasing each of the lower bounds by a small value, and then calculating a solution for $v$ within these new constraints using LP. This procedure ensures that the initial point is chosen within the solution space and avoids computational difficulties that arise when the initial point lies at the extremity of the solution space. The second step of the ACHR algorithm calculates "warm up" points from the initial point using several iterations of a basic Hit-and-Run algorithm[64]. A geometric center, $s$, is calculated from the "warm up" points which are stored as columns of a matrix **W**. The

third step of the ACHR algorithm calculates sample points as follows. The direction for the next iteration from point $x_m$ is chosen by randomly taking one point $y$ out of the matrix **W** and applying a direction vector $p = (y–s)$ to $x_m$. At each iteration, the newly calculated point, $x_{m+1}$, is substituted randomly into **W** in place of a previously calculated point so that the matrix **W** always contains the most recent points. The approximated center is also recalculated after each iteration. This third step is repeated until a desired number of points are reached. This approach allows the distribution of points converges to a uniform distribution at a much faster rate than the standard hit-and-run method. In each sampling calculation, 500,000 randomly distributed points are computed with 100 iterations between each point. The algorithm is implemented in Matlab (MathWorks Inc., Natick, MA) with Lindo (LINDO Systems Inc., Chicago) as the linear programming solver. The same solution space is sampled five times using different randomly chosen initial points to verify that the calculated distribution of points is independent of the starting point. The resulting distributions are compared with one another to ensure that no difference is observed.

## 3. Capabilities of the mitochondrial network

The reconstructed mitochondrial network (Chapter 1) represents the integration and simultaneous analysis of heterogeneous data types. Systemic analysis of the three mitochondrial metabolic objectives highlights the interdependence of various biochemical pathways in this organelle. Previous *in silico* studies[63,71-73] of mitochondrial metabolism primarily focused on energy metabolism, although it is well known that mitochondria play important roles in many cellular processes. The present reconstruction

is more comprehensive, and therefore can be used to assess a broader range of the mitochondria's physiological functions. In particular, three metabolic roles are considered in this study: ATP production, heme biosynthesis, and phospholipid biosynthesis. Oxidative phosphorylation (OxPhos) fulfills the energetic requirement for heme and phospholipid syntheses, whereas the latter processes provide the maintenance for proteins and membrane structure essential for the energy conversion pathway.

## 3.1 Energy conversion

Mitochondria constitute up to 35% of the mammalian cardiomyocyte volume[74]. The primary function of the cardiac mitochondria is to produce energy for the heart's contraction; therefore it is possible that these organelles can operate at approximately theoretical metabolic limits. Normal, well-perfused myocardium generates more than 90% of its ATP by OxPhos and less than 10% by glycolysis[75]. Most of this ATP is generated by aerobic metabolism taking place in mitochondria; so efficiency of ATP production within the mitochondrial network is vitally important. Here the reconstructed network is used to study ATP yield from the two main energy sources of the heart – glucose and fatty acids.

## 3.2 Glucose metabolism

The complete mitochondrial network reconstruction allows for a systemic and accurate calculation of ATP yields and PO ratios from oxidation of various metabolites. The maximal ATP yield per glucose in the reconstructed network is 31.5. This number

has been a topic of discussion for many years with some textbooks reporting a value of 36-38 ATP/glucose[47] while others report a value of 31[76,77]. Since reactions in the present reconstruction account for every proton consumed and produced, it is possible to calculate this ATP yield on a network-wide basis. It is found that after accounting for all protons consumed in glycolysis, the malate-asparate (mal-asp) shuttle, and phosphate transport, there is a net difference of two protons per glucose molecule between our calculations and those previously reported[76,77]:

$$\text{Glucose} + 2\ P_i + 2\ \text{ADP} + 2\ \text{NAD} \rightarrow 2\ \text{pyruvate} + 2\ \text{ATP} + 2\ \text{NADH} + \mathbf{2\ H^+} + 2\ H_2O$$

Previous calculations do not account for the fate of the protons produced in the preceding net reaction when computing the overall ATP yield by glucose. When one balances the entire network; however, these protons must be consumed. These two protons account for the 0.5 ATP discrepancy. Physiologically these protons may or may not contribute to the proton gradient, which is localized to the mitochondrial inter-membrane space, depending on how proximal glycolytic enzymes are to the mitochondria. As the proteomic data from Taylor *et al.*[42] and other studies[45] have shown that glycolytic enzymes are always found in purified mitochondria samples, it is likely that glycolysis takes place near the mitochondria, and thus the additional proton contribution is significant. Considering the high rate of glycolysis in muscle cells, the generated protons probably affect the mitochondrial inter-membrane gradient and contribute to the overall ATP production in the organelle.

## 3.3 Fatty acid metabolism

Cardiac mitochondria derive the majority of their energy from fatty acids[78]. Long-chain fatty acids are transported to the mitochondrial matrix by the carnitine shuttle for β-oxidation. Six types of fatty acids – palmitate (n-C16:0), stearate (n-C18:0), oleate (n-C18:1), octadecynoate (n-C18:2), arachidonate (n-C20:4), docosohexaenoate (n-C24:6) – are considered in the present reconstruction (Chapter 1). These fatty acids make up 90% of the fatty acid composition of human heart phospholipids[79,80]. Only the ATP yield from palmitate is computed here for comparison with existing literature calculations. The remaining long-chain fatty acids are generally not used for energy, but are important in the assembly of mitochondrial phospholipids. The maximal ATP yield for each palmitate in the reconstructed network is 106. Biochemistry textbooks usually report a value of 129 ATP[81] or 106 ATP[77] depending on whether NADH is considered to have a PO ratio of 3 or 2.5 respectively. Here, instead of directly imposing a fixed PO ratio, only the number of protons transported by the electron transport chain and consumed by ATP synthase are specified based on the stoichiometry of these reactions. This result shows that a PO ratio of 2.5 is consistent with the reconstructed network.

## 3.4 Heme biosynthesis

In addition to energy conversion, heme biosynthesis is another function that is common to mitochondria of all tissues[82]. Heme molecules play an essential role in the assembly of many apoproteins including electron transfer complexes and catalase. Cytochromes are examples of biologically active hemes linked to polypeptides. Only the

synthesis of protoheme is considered in the reconstruction, as it has been reported that the synthesis rate of "mature" heme (for example, heme a and b) is controlled by that of protoheme[83]. The carbon backbone of protoheme is derived from succinyl-CoA and glycine. Given the constraints on the network, glycine has a higher shadow price than succinyl-CoA. The shadow price of a metabolite reflects how sensitive the value of the objective function is with respect to that metabolite[84]. This finding confirms experimental results with isolated perfused rat hearts[83], which showed that changes in glycine concentration and uptake rate account for changes in the rate of heme biosynthesis.

## 3.5 Phospholipid biosynthesis

Mitochondria are semi-autonomous organelles. Besides producing energy and various metabolites for the cells, they also perform their own assembly and maintenance. In human heart tissue, the mitochondrial membrane is composed of 18% cardiolipin (CL), 34% phosphaethanolamine (PE), 43% phosphotidylcholine (PC), and 5% phosphatidyinositol (PI)[85]. The mitochondrion can synthesize CL from phosphatidylglycerol and PE from phosphatidylserine decarboxylation[85,86]. Phosphatidylserine is believed to be synthesized on the endoplasmic reticulum membrane closely associated with the mitochondria[86-88]. The objective function representing the mitochondria's phospholipid synthesis activity (Section 2.1) does not include PI because it is primarily produced in the endoplasmic reticulum[85]. This objective function is used for the flux variability computations presented below.

### 3.6 Reactive oxygen species (ROS) detoxification

It has been reported that at least 2-5% of electrons transported on the electron transport chain are *lost,* and therefore not used to reduce oxygen to water[89]. The lost electrons react with oxygen, hydroxide ions, nitrogen containing compounds, and fatty acids to form reactive radicals. These radicals are collectively referred to as ROS. The current network accounts for the generation of superoxides ($O_2$-) but not other species. Superoxide dismutase reduces superoxides to hydrogen peroxide ($H_2O_2$), which is converted to water by either catalase or glutathione peroxidase. The contribution of catalase to this function is believed to be minimal compared to that of glutathione peroxidase in heart mitochondria[74]. The glutathione peroxidase and the glutathione cycle use NADPH to reduce $H_2O_2$, and therefore implicitly imposing an energetic stress on the cell. Consequently metabolism of glucose produces 3% less ATP when 2% of electrons are lost and 7% less when 5% of electrons are lost. Oxidation of palmitate produces 3% and 8% less ATP for these two conditions, respectively. Calculations in the remaining of this chapter assume an overall 2% electron loss.

## 4. Network robustness and flexibility

To better understand the robustness and flexibility of the reconstructed network in carrying out its metabolic functions, the variability of each reaction flux[66] as well as the properties of equivalent optimal flux distributions[90,91] in the network are analyzed. Here, flexibility and robustness are measured by the range of allowable flux values in each

reaction and the number of equivalent pathways (alternate optima) that can lead to the optimal results for the objective function(s), respectively.

## 4.1 Flux variability analysis

The flexibility of a particular reaction in the mitochondrial network depends on the metabolic function under consideration. Flux variability analysis determines the allowable range of flux values while the network optimally satisfies a particular metabolic objective and physiological constraints on enzymatic capacities (Table 3.1). When ATP production is the metabolic objective, the network has the least flexibility: only 8% of the 189 reactions have variable fluxes, 25% always have a constant non-zero flux value, and 67% are never used. When heme is the objective end product, higher flexibility is observed with 60% reactions having variable fluxes, 8% having a constant non-zero flux, and 32% not used. For phospholipid synthesis, these proportions are 60%, 11%, and 29%, respectively (Table 3.2).

Among reactions with variable fluxes, the average flux ranges while maximizing for ATP, heme, and phospholipid, are 60, 23 and 23 μmol/min/g protein, respectively. Heme and phospholipid syntheses have the same set of reactions with variable fluxes. This set contains 114 reactions participating in glycolysis, the TCA cycle, OxPhos, β-oxidation, ROS detoxification, and the urea cycle. Notably, it does not include any reactions in the heme and phospholipid synthesis pathways. Fifteen of these 114 reactions also have variable fluxes under the ATP synthesis condition. These fifteen

reactions are involved in catabolism (ATPS4m, PYK, PCm, PEPCKm, NDPK1), phospholipid metabolism, (DAGKm, PAPAm), and transport (CITtam, CITtbm, GTPtm, HCO3Em, Htm, MALtm, PIt2m, PYRt2m).

The flux range of reactions in the network when no objective function is imposed is also calculated for comparison (Figure 3.3). These flux ranges are only slightly larger than the variation found when heme and phospholipids are optimized, but significantly larger than the variation when ATP is optimized. This implies that a large portion of the solution space is accessible for optimal heme and phospholipid syntheses, but it is much more restricted for ATP production. Overall, in addition to transport reactions, those in glycolysis, the TCA cycle, and the respiratory chain have the highest flux variability.

## 4.2 Properties of alternate optima

The number of different equivalent flux distributions that can optimally satisfy a metabolic function can be used to measure the robustness of a network. Similar to flux variability analysis, the number of alternate network flux distributions is a function of the metabolic objective and metabolite uptake rate ($V_{max}$ of transporters). These constraints are the same as those used for flux variability analysis (Table 3.1).

When glucose, fatty acids, and glutamate are simultaneously available to produce ATP, only four optimal flux distributions are found. The average number of reactions used per optimal solution is 57.5, with a minimum of 55 and a maximum of 62 reactions

(Table 3.2). When the network is optimized for protoheme, a total of 8,288 equivalent flux distributions are found. On average, there are 78.7 reactions used per alternate optima, ranging from 53 to 81 reactions. The network, however, appears to be much more flexible in synthesizing phospholipids as compared to the other two cases. A large number of optimal flux distributions, 21,863, are discovered. An average of 98.1 reactions are used across these flux distributions, with a minimum of 75 and a maximum of 118.

The alternate optima reported in this study are calculated with respect to one objective function at a time, while setting the other two functions to zero. This is done to enumerate equivalent metabolic pathways that can maximally satisfy one metabolic function without being influenced by reactions that are used for a different metabolic objective. It also eliminates pathways that use unnecessarily high numbers of reactions only to produce by-products that are not part of the objective function. The drawback is that extreme points that simultaneously produce all three metabolic products are missed. The effort to enumerate all alternate optima without setting two objective functions to zero has been computational infeasible within practical time. Nevertheless, among the alternate optima found here, those that are highly correlated with each other give clues about the flux distributions that simultaneously satisfy all three metabolic functions.

## 5. Analysis of alternate optima

## 5.1 Optimal flux distributions are correlated with each other

The similarity among flux distributions is evaluated by calculating the pair-wise correlation between each of the four ATP-specific alternate optima and each of heme- and phospholipid-specific optima. Figure 3.4 shows the cumulative distribution of correlation coefficients r between ATP and heme optimal fluxes, and ATP and phospholipid optimal fluxes. Both distributions share the same general characteristics: there are no pairs of fluxes with negative correlation, roughly half of the pairs of fluxes have very low correlation ($r < 0.2$), and the other half has very high correlation ($r > 0.9$). The absence of negative correlation among these fluxes suggests that when the mitochondrion is producing ATP for the cell, it is also likely to produce other compounds important for its own maintenance.

The correlation among the sets of heme and phospholipid flux distributions that have very high correlation with ATP alternate optima ($r > 0.975$) is further investigated. These flux distributions are very well correlated with each other – the majority have correlation value higher than 0.99 (Figure 3.4). This criterion ($r > 0.99$) identifies a subset of 425 flux distributions (out of a total of 30,155 alternate optima calculated) that are highly correlated with each other. One out of the 425 flux distributions is an ATP optimal solution, 157 are heme optimal solutions, and the rest (267) are phospholipid optimal solutions.

In comparing the three flux distributions for ATP, heme, and phospholipid that are most correlated with each other, it is found that only 20 reactions need to be activated

for the network to produce phospholipids simultaneously with ATP. Another seven reactions are needed for simultaneous syntheses of all three products. The high correlation amongst a subset of flux distributions suggests that these are the most probable *in vivo* fluxes. Figure 3.1 shows the flux variations (result of flux variability analysis) found among flux distributions that are optimal for individual metabolic objective functions and among flux distributions that are highly correlated with each other. The variation found in the latter group substantially reduces the flux range seen among heme and phospholipid optimal fluxes.

## 5.2 Correlated reaction sets form network modules

The sets of reactions that are always used together across all optimal flux distributions are referred to as correlated reaction sets[92,93]. These reactions form functional modules that are likely to be co-regulated. Correlated reaction sets among optimal solutions calculated for each objective function and among highly correlated flux distributions are identified here. Reactions that are always active with respect to a specific objective function form the largest correlated reaction set under that objective function. Under the three objective functions - ATP production, heme synthesis, phospholipid synthesis - the largest correlated reaction sets contain 53, 35, and 56 reactions, respectively (Table 3.2).

When ATP production is the objective function, two more correlated reaction sets are found. One set contains DPK1 and GTPtm, which participate in the conversion of

GPT produced by substrate level phosphorylation to ATP. The other reaction set uses phosphoenolpyruvate kinase (PEPCKm) to consume GTP (and oxaloacetate) to produce phosphoenolpyruvate. Two out of four ATP optimal flux distributions use NDPK1 and GTPtm. These two flux distributions are much better correlated with heme and phospholipid optima, confirming that mitochondria usually convert GTP to ATP directly. When heme and phospholipid syntheses are used as the objective functions, 19 and 13 other correlated reaction sets are found, respectively. Some of these correlated sets are present in both cases. Two of these sets contain reactions in the urea cycle and glycolysis. One of these sets appears to form a module for oxygen metabolism. Seven other sets contain reactions involved in the shuttling and oxidation of the seven fatty acids. The remainders are smaller reaction sets with two or three reactions each. Reactions within each of these sets are also known to be involved in the same biochemical pathway.

Correlated reaction sets found among highly correlated flux distributions form functional modules that nicely encompass reactions known to participate in the same or connected biochemical pathways. In particular, reaction sets corresponding to lipid metabolism, heme biosynthesis, glycolysis, the TCA cycle and OxPhos, urea cycle, and mal-asp shuttle are notable examples. These results show that given fluxomic data, this method can be used to identify functional modules that represent physiologically significant components in the network. Whether or not enzymes within these functional modules actually share the same regulatory mechanism is perhaps a more difficult question to answer. Since most of these enzymes are encoded in the nuclear genome, and

many are post-transcriptionally regulated, co-appearance inside the mitochondria may not directly imply co-expression in the nucleus. Nevertheless, knowledge about these functional modules allows a better understanding how the separate components in the network operate to carry out the overall function of the mitochondrion.

## 5.3 Distribution of flux values among the equivalent optimal flux distribution

The distribution of flux values of a reaction across all equivalent optimal solutions can be interpreted as the activity profile of that reaction[46]. These profiles are analogous to gene expression profiles, except that they represent enzyme activity levels rather than mRNA transcript levels. Since most reactions under maximal ATP production conditions do not have much flux variation, only profiles for the heme and phospholipid syntheses are discussed further. Profiles of reactions in glycolysis, the TCA cycle, and OxPhos for heme and phospholipid biosyntheses are similar for both cases. Glycolytic reactions have a bimodal distribution, with high peak values at extreme ends of the flux range; i.e. either these reactions are fully used or not at all in the optimal solutions. The TCA cycle reactions have more diverse distributions. Also, reactions corresponding to complexes I and II seem to have broader flux distribution than the rest of the reactions of the electron transport chain.

Flux distribution profiles of reactions in the urea cycle, ROS detoxification, and fatty acid metabolism are nearly identical for both the heme and phospholipid synthesis

cases. Reactions in the urea cycle and ROS detoxification pathway tend to have one of two dominant flux values, with a few reactions having no variation. Reactions in fatty acid transport and oxidation all have two major peaks at extreme ends of the flux spectra, except for the carnitine carrier (CRNtim). These profiles suggest that only a subset of the 6 fatty acids types is used at a time by the network. However, since the carnitine carrier is shared among all long chain fatty acids, it tends to be active regardless of which fatty acid is used. The rest of the reactions in the network do not have much variation.

## 6. Distribution of fluxes and resources among the three mitochondrial functions

Multiple-objective analysis is used to investigate how the network distributes its resources to simultaneously satisfy all three metabolic functions. The lexicographic method allows one to hierarchically order the three objective functions. All six possible ordering of the three objectives are explored here to study the relative importance of each. Results show that all of the ATP-specific alternative optima have zero flux values for heme and phospholipid production. Note that heme and phospholipids can be simultaneously maximized, but not so with either of them and ATP demand. When the synthesis of either heme or phospholipid is maximized first, near optimal values are obtained for the other two functions. In fact, both heme and phospholipid productions achieve their maximum values simultaneously while ATP demand is strictly less than its maximum value – more so if ATP production is the third objective in the order.

Similar results are observed using the weight formulation. In particular, w is set to be ((g/(2+g), 1/(2+g), 1/(2+g)), where g/(2+g) corresponds to the weight coefficient of the ATP objective function and the other two objective functions have equal weight. As g increases, the weight for ATP demand increases towards one. For $g \geq 0.72$, the weighted solution is the same as maximizing ATP demand (Figure 3.5). The composite objective begins with a simple average of heme and phospholipid, both at their maximum values of 0.125 and 3.704, respectively. For $g = 0.72$, the value of ATP demand reaches its maximum; heme and phospholipid biosyntheses are at zero. The value of the composite objective function, however, is biased by the metabolic function with dominant flux output (ATP production). To remove this bias, the weight coefficient of each of the metabolic function is rescaled by its respective optimal flux value, and g is varied such that $w_1$ goes from zero to approximately one. The value of the new composite objective function increases linearly with g in this case.

The fact that both heme and phospholipid objective functions can be simultaneously satisfied suggests that these two functions of the mitochondria can operate relatively independent from each other without affecting one another's resource pool. However, these two functions require energy in the form of ATP, thus reducing the available ATP for the cell. This, of course, should not be viewed as a compromise to the energy conversion ability of the mitochondria, because ATP is produced precisely for the purpose of fueling cellular processes both outside and inside the mitochondria.

# 7. Candidate steady-state flux distributions in the normal physiological condition

The preceding sections show that are an infinite number of theoretical steady-states $v$ satisfying Equation 3.1-3.2 for the mitochondrial network (this is also true for most metabolic networks). A distribution of possible values for each reaction rate ($v_i$) can be determined by sampling the entire null space of S. Such results allow one to evaluate how changes in steady-state rates of a subset of reactions systemically affect the rates of the remaining reactions in the network. The remaining sections of this chapter apply the ACHR algorithm[64] to evaluate steady-state flux distributions in the cardiac mitochondrion subject to three different metabolic conditions: normal, diabetic (Figure 3.6), and ischemic (Figure 3.7)[94].

## 7.1 Constraints on the reaction network

The normal physiological condition described here represents the metabolism of a human cardiac mitochondrion at rest. Constraints are applied on the uptake and efflux rates of metabolites according to an expected normal state of the mitochondrion. A positive lower bound on the demand for ATP ($DM_{ATP}$) is applied to represent the minimum energy required for normal ion homeostasis. This required ATP level is set at approximately 26% of the total ATP production[95]. An ATP production rate of 30 μmol/min/g protein, taken from measurements in the working dog heart[96], is used here because corresponding data for humans could not be found. The lower bound of $DM_{ATP}$

is thus set at 7.5 μmol/min/g protein. The uptake rate of n-C16:0 has been measured to be 1 μmol/min/g protein[97]. This value is used to approximate the uptake rates of other fatty acids based on their observed distributions in mammalian cardiomyocytes: 53% for n-C18:0, 16% for n-C16:0, and 7% for n-C18:1 and n-C18:2[95]. Since longer chain fatty acids, n-C20:4 and n-C22:6, are not detected in these experiments[95], they are estimated to make up less than 2% of the total fatty acid uptake. The lower bound and upper bound are set by taking a 25% variation around the experimentally measured values (Table 3.3). The uptake rate of lactate is constrained to the same upper bound defined for glucose, as both substrates are consumed at approximately equivalent rates under normal physiological conditions[98]. The uptake of the ketone bodies, acetoacetate and (R)-3-hydroxybutanoate, is allowed but the secretion of ketone bodies is set to zero as the heart does not normally produce and export ketone bodies like the liver does[99]. In addition, the maximum flux rates for these ketone bodies are set to be very small (0.001 μmol/min/g protein) since the plasma concentration of circulating ketone bodies is less than 0.1 mM at normal physiological conditions[99].

## 7.2 Steady state flux distributions

Network flux distributions satisfying the described constraints represent all possible steady states that are consistent with these measurements. Figure 3.8 (black line) shows allowable steady-state flux distributions for selected metabolic reactions under normal conditions. Each histogram corresponds to the range of possible steady-state flux values allowed for a reaction in the network. Peak values represent the most

probable flux values within the distribution. In general, four shapes of flux distribution can be distinguished: *i*) *right peak* ($H^+$ efflux), *ii*) *left peak* (urea excretion), *iii*) *central peak* (phospholipid synthesis), and *iv*) *broad peak* (ATP production). The histogram of heme production shows that the most probable flux values are small (close to zero). The biosynthesis of mixed phospholipids shows a relatively wide range of highly probable flux values. These phospholipids are mainly used for maintenance of the mitochondrial membranes. For a number of reactions, the computed most probable flux values are good estimates of the measured values (NADH2-u10m, complex I of the electron transport chain and ASPGLUm, the aspartate-glutamate shuttle).

# 8. Effects of diabetes

## 8.1 Model of diabetic condition

Diabetic conditions are characterized by a lack of, or insensitivity to, insulin. This condition results in unregulated and increased fatty acid uptake in mitochondria via the carnithine-palmitoyl-transferase (CPT-I) shuttle[100,101]. The mitochondrial fatty acid uptake flux through CPT-I is thus increased in the model to reflect its activity under diabetic conditions. Effects of diabetes on mitochondrial metabolism are assessed by applying constraints reflecting *i*) unregulated fatty acid uptake via carnitine-palmintoyl-transferase-I (CPT-I) shuttles, *ii*) decreased glucose consumption due to the reduced number of glucose transporters in the cell membrane, and *iii*) increased ketone body uptake due to their higher blood concentration.

Though no quantitative experimental values could be found on how much the fatty acid uptake rates increase in diabetic patients, a significant increase has been observed in diabetic mouse myocardium[102]. Therefore, the lower-bounds of the fatty acid uptake rates are increased to 80% of the maximum allowed flux values under the normal condition (Table 3.3). Note that increasing the lower bound on the uptake of n-C20:4 and n-C22:6 results in an empty solution space. Maximum glucose uptake is reduced to 75% of its normal rate[103] due to the decreased number of available glucose transporters in the cardiomycyote membrane under insulin insufficiency or insensitivity[100,101]. Another effect of diabetes is that plasma ketone body concentration may be as high as 25 mM[104,105]. Thus, upper limits on uptake rates of ketone bodies defined under normal conditions are removed. It should be noted, however, that even though the constraints on ketone body uptake are removed in the diabetic condition, the resulting maximum allowable fluxes through the ketone body transporter under diabetic conditions are much lower than that under normal conditions (~1.4%). This reduction in ketone body uptake is due to the increased fatty acid uptake rate in the diabetic condition, which systemically limits ketone body transport. To simulate a high influx of ketone bodies, their minimum uptake rates are set at 10% of the highest possible uptake of each ketone species. These new constraints represent the salient features of changed substrate supply observed in diabetic cardiomyocytes.

## 8.2 Effects of diabetic conditions on mitochondrial metabolism

A possible treatment for diabetes would be to force a normal glucose and ketone

body uptake via therapeutic agents. To investigate these effects, the constraints on glucose and ketone body uptake are set to their values under the normal condition. Figure 3.8 shows the probability distributions for selected reactions in four cases: *i*) normal physiological (black), *ii*) diabetic (red), *iii*) diabetic with normal glucose uptake (green), and *iv*) diabetic with normal glucose and ketone body uptake (blue). Results from these four conditions show that increased mitochondrial fatty acid uptake leads to smaller ranges of allowable flux rates for most reactions and thus reducing network flexibility dramatically. Although normalization of ketone body and/or glucose uptake leads to higher flux ranges, the effect is minimal. The high mitochondrial fatty acid breakdown leads to a higher and narrower distribution of oxygen consumption. Due the compensation of fatty acids for glucose, $DM_{ATP}$ in diabetic cases differs only slightly from that in the normal condition.

Another interesting finding from our results is that the flux through the mitochondrial pyruvate dehydrogenase (PDH) enzyme becomes significantly restricted by network stoichiometry with increased fatty acid consumption (Figure 3.8). Many studies have tried to identify factors that affect the inhibitory mechanism of PDH under conditions such as diabetes[100,101]; this study shows that an increase in cellular fatty uptake flux forces a significantly lower flux through PDH as a direct consequence of overall network stoichiometry. *In silico* predictions of changes in metabolic function in the diabetic condition relative to normal physiological conditions are compared with diabetes-caused changes in metabolism found in the literature (Table 3.4). Results are quite consistent between this study and experimental observations.

# 9. Effects of ischemia

## 9.1 Model of ischemic conditions

Severe ischemia leads to at least 70% reduction in local blood flow and a consequent decrease in oxygen supply. Here effects of mild ischemia (25% oxygen reduction) as well as effects of two therapeutic approaches are studied[106,107]. One therapy involves administration of glucose, insulin, and potassium (GIK)[106]. This therapy seeks to alleviate damage during and after an ischemic event by raising the amount of ATP available for metabolic and non-metabolic tasks through glycolysis. An alternate therapy combines the first approach with increased ketone body uptake. This latter treatment is based on the observation that ketone bodies had a positive effect in the case of acute insulin deficiency[107]. These two approaches are implemented *in silico* by *i*) increasing maximum glucose uptake rate to 2 μmol/min/g protein and *ii*) increasing the lower-bound constraint on uptake rates of both types of ketone bodies.

## 9.2 Effects of ischemic conditions on mitochondrial metabolism

Severe ischemia leads to an undersupply of oxygen, and consequently, of energy for cardiomyocytes. Cell damage during and after an ischemic event is associated with reduced energy production, decreased contractile work, and increasing acidosis due to increased glycolysis. Candidate flux distributions under ischemic conditions are compared to those computed for the normal physiological condition (Figure 3.9).

Interestingly, maximal ATP production under ischemic conditions does not differ significantly from that found for diabetic conditions. However, the shape of the steady-state flux space changes significantly due to the changed maximal oxygen uptake rate. Both the maximally allowed and the most probable flux values are reduced in most reactions since ATP is the most undersupplied metabolite in ischemic conditions. Exceptions are reactions in the phospholipid biosynthesis pathway where the most probable flux values are increased. This increase allows consumption of fatty acids, which can not be oxidized in absence of oxygen. Overall, the changed tendency of the flux distributions of the network is comparable to the reported changes observed in ischemic patients (Table 3.4).

## 9.3 Effects of diet treatments

There have been multiple approaches in ischemia therapy to reduce negative effects of reperfusion; two of these therapies are investigated in this study. The GIK infusion[106] is simulated by increasing the maximal glucose uptake rate. Figure 3.9 (blue line) shows this effect on the metabolic network under oxygen restricted conditions. Flux distributions under GIK treatment differ only slightly from those of the untreated ischemic condition, suggesting that this therapeutic approach may not be effective. Though the goal of this therapy is to increase ATP available for contractile work, the model shows that neither ATP production nor fluxes through ATP consuming reactions increase as a result. However, effluxes of protons and lactate are shifted higher, corresponding to two well known side effects of this therapy. The resulting acidosis

often leads to further damage during reperfusion. The second ischemia therapy is based on the observation that ketone body oxidation produces higher ATP yield per oxygen molecule than glucose oxidation but requires less oxygen than fatty acid breakdown. Effects of GIK in combination with ketone body administration are shown on Figure 3.9 (green line). Results with this second treatment deviate only slightly from those with GIK alone, with the flux distribution of lactate and proton production shifted to the right into a higher flux value range.

An alternative therapeutic approach proposed in the literature to treat ischemia is to stimulate activity of PDH[108]. The rationale for this approach is that higher flux through PDH should result in a higher glytolytic flux and therefore a reduced lactate level[108]. However, our calculations predict that direct stimulation of PDH can only lead to a slightly higher steady-state flux through this reaction due to stoichiometric constraints. The highest possible flux through PDH under the ischemic condition is only 17% of the maximal flux under the normal condition. It is thus concluded that a therapy directly targeting PDH is likely to not have a lasting effect.

(This chapter, in part, is a reprint of the material appearing in "Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data," Vo TD, Greenberg HJ, and Palsson BO. 2004. *J Biol Chem* 279(38):39532-40 and "Candidate metabolic network states in human mitochondria: Impact of diabetes, ischemia, and diet," Thiele I, Price ND, Vo TD, and

Palsson BO. 2005a. *J Biol Chem* 280(12):11683-95. The dissertation author is the primary investigator and an author of the materials used from these papers.)

**Table 3.1: Constraints on reaction fluxes in the mitochondrial network**

These data are collected primarily from experiments with rat heart and liver mitochondria. Please refer to Vo *et al.*[46] for abbreviations of reactions and metabolites. The symbols [m] and [c] at the beginning of a reaction indicate whether the reaction occurs inside the mitochondria or cytosol, respectively. The [m] and [c] following a metabolite denote the compartmental location of that metabolite. LB and UB stand for lower bound and upper bound on reaction fluxes, respectively. All data below have units of μmol/min/g protein, which are converted from reported units as necessary. Most references do not specify whether the reported unit is "g total mitochondrial proteins" or "g isolated protein" based on our interpretation of the reported experimental procedures, it appears that "g total mitochondrial protein" is more appropriate.

| Reaction | Equation | LB | UB | Reference |
|---|---|---|---|---|
| CRNtim | crn[m] → crn[c] | 0 | 200 | [109] |
| C160CPT2 | [m] : coa + pmtcrn → crn + pmtcoa | 0 | 475 | [110] |
| C160CPT1 | [c] : crn + pmtcoa → coa + pmtcrn | 0 | 468 | [110] |
| ASPGLUm | asp-L[m] + glu-L[c] + h[c] ↔ asp-L[c] + glu-L[m] + h[m] | -40 | 93 | [111] |
| PDHm | [m] : coa + nad + pyr → accoa + co2 + nadh | 0 | 32 | [112] |
| MALtm | mal-L[c] + pi[m] ↔ mal-L[m] + pi[c] | -20 | 19 | [113] |
| CITtbm | cit[c] + pep[m] ↔ cit[m] + pep[c] | -104 | 104 | [114] |
| PYRt2m | h[c] + pyr[c] ↔ h[m] + pyr[m] | -110 | 110 | [112] |
| CITtam | cit[c] + mal-L[m] ↔ cit[m] + mal-L[c] | -92 | 113 | [114] |
| SUCCt2m | pi[m] + succ[c] ↔ pi[c] + succ[m] | -13 | 13 | [115] |
| LYStm | h[m] + lys-L[c] ↔ h[c] + lys-L[m] | -120 | 120 | [116] |
| ARGtm | arg-L[c] + h[m] ↔ arg-L[m] + h[c] | -105 | 105 | [116] |
| CITRtm | citr-L[m] ↔ citr-L[c] | -60 | 60 | [116] |
| ATPtm | adp[c] + atp[m] → adp[m] + atp[c] | -32 | 32 | [117] |
| ORNt4m | citr-L[c] + h[c] + orn[m] ↔ citr-L[m] + h[m] + orn[c] | -145 | 145 | [116] |

**Table 3.2: Reaction participation among equivalent optimal flux distributions (alternate optima)**

Each of the three right most columns corresponds to the metabolic objective under which the alternate optima are calculated. Optima refer to the extreme points of the solution space that achieve the optimal value for the objective function, whereas feasible extreme points only satisfy the constraint of the LP problem

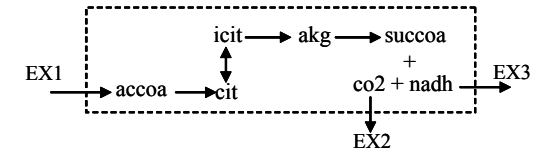
| | ATP | Heme | Phospholipids |
|---|---|---|---|
| Total number of optima | 4 | 8,288 | 21,863 |
| Average reactions used per optimal solution | 57.5 | 78.7 | 98.1 |
| Reactions with variable fluxes | 15 | 114 | 114 |
| Reactions with constant (non-zero) fluxes | 47 | 14 | 21 |
| Reactions always used (largest correlated set) | 53 | 35 | 56 |
| Reactions never used | 127 | 61 | 54 |
| Number of feasible extreme points | 9,160 | 11,000 | 50,012 |

**Table 3.3: Constraints representing diabetic and ischemic conditions**
These constraints represent the key factors that are associated with diabetic or ischemic metabolic conditions. LB and UB stand for lower bound and upper bound, respectively. (*) denotes the effects on the substrate LB and UB are the sole results of network stoichiometry after the constraint in the normal condition is removed. A detailed list of applied constraints can be found in reference [94].

| Condition | Substrate | Uptake rate | Constraint | Uptake rate | Constraint |
|---|---|---|---|---|---|
| Diabetes | Fatty acids | LB | Increased | UB | Increased* |
| | Glucose | LB | Decreased | UB | Decreased |
| | Ketone bodies | LB | Increased | UB | Increased* |
| Ischemia | Oxygen | LB | Decreased | UB | Decreased |

**Table 3.4: Comparison between computed flux values and experimental data**
UR and Expt stand for "uptake rate" and "experimental data", respectively. FAOX stands for fatty acid oxidation. Peak values refer to peaks of the histograms made from uniform random sampling of all the steady state flux space. LB and UB are determined by flux variability analysis (section 2.2). Unit: μmol/min/g protein.

| Reaction | LB | UB | Peak | Expt | References |
|---|---|---|---|---|---|
| $DM_{ATP}$ | 7.50 | 45.99 | 7.86 - 23.89 | 30.00 | Low work dog heart[96] |
| Glucose UR | -0.88 | -0.53 | -0.86 | -0.51 | in vivo human heart[118] |
| Glucose UR | -0.88 | -0.53 | -0.86 | -0.7 | in vivo human heart[119] |
| Glucose UR | -0.88 | -0.53 | -0.86 | -0.57 | in vivo human heart[120] |
| Glucose UR | -0.88 | -0.53 | -0.86 | -0.56 | in vivo human[121] |
| Lactate UR | -0.88 | 24.03 | -0.67 | 1.04 | in vivo working rat heart[122] |
| Oxygen UR | -39.10 | -23.44 | -38.94 | -30.00 | Isolated rat heart[123] |
| Oxygen UR | -39.10 | -23.44 | -38.94 | -28.00 | in vivo human heart[124] |
| CTP-I (C16:0) | 0.10 | 1.25 | 0.11 | 0.54 | in vivo pig skeletal muscle[125] |
| CTP-I (C16:0) | 0.10 | 1.25 | 0.11 | 0.47 | Rat heart mitochondria[110] |
| CTP-I (C16:0) | 0.10 | 1.25 | 0.11 | 0.90 | Isolated rabbit heart[97] |
| CTP-II (C16:0) | 0.10 | 1.25 | 0.11 | 0.48 | Rat heart mitochondria[110] |
| Complex IV | 18.68 | 39.70 | 37.66 | 19.20 | in vitro rat heart[126] |
| FAOX C16:0 | 0.24 | 1.74 | 0.84 | 0.35 | in vitro rat heart[127] |
| FAOX C16:0 | 0.24 | 1.74 | 0.84 | 0.70 | Isolated working rat heart[128] |
| FAOX C18:1 | 0.27 | 1.64 | 0.33 | 1.06 | in vivo working rat heart[122] |
| Complex I | 31.75 | 61.49 | 54.13 | 52.44 | Human muscle[129] |
| 3-oxoacid CoA-transferase | 0.00 | 14.47 | 0.13 | 13.50 | Isolated enzymes from human dermal fibroblast[130] |
| PDH | 0.00 | 8.30 | 0.07 | 6.00 | isolated working rat hearts perfused with 0.4 mM palminate[131] |
| Pyruvate transport | -0.39 | 2.75 | 18.82 | 0.54 | Isolated rat heart[132] |
| Citrate synthase | 2.22 | 11.50 | 14.86 | 11.20 | Isolated rabbit heart[97] |
| Asp/Glu Shuttle | 0.00 | 2.30 | 20.56 | 2.90 | Isolate rabbit heart[97] |




CS accoa + h2o + oaa → cit + coa + h
ACO cit ↔ icit
ICDH icit + nad → akg + co2 + nadh
AKGD akg + coa + nad → co2 + nadh + succoa

**CONSTRAINT-BASED**

$$\mathbf{S} = \begin{array}{c} \\ \text{accoa} \\ \text{h2o} \\ \text{oaa} \\ \text{cit} \\ \text{coa} \\ \text{h} \\ \text{icit} \\ \text{nad} \\ \text{akg} \\ \text{co2} \\ \text{nadh} \\ \text{succoa} \end{array} \begin{array}{c} \begin{array}{ccccccc} \text{CS} & \text{ACO} & \text{ICDH} & \text{AKGD} & EX1 & EX2 & EX3 \end{array} \\ \begin{pmatrix} -1 & 0 & 0 & 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Maximize nadh output

Subject to $\frac{d[X]}{dt} = S \cdot v;\ \alpha_i \le v_i \le \beta_i$

**Parameters:**

Influx of accoa, influx of oaa

**Possible results:**

Steady state reaction rates for maximal production of nadh (or other end products); range of allowable ICDH flux value to maintain steady state; a list of different flux distributions producing the same end products; changes of allowable flux values for AKGD as a function of changes in CS flux; etc

**KINETICS-BASED**

$$\frac{d[cit]}{dt} = v(CS) - v(ACO)$$

$$\frac{d[icit]}{dt} = v(ACO) - v(IDH)$$

$$\frac{d[akg]}{dt} = v(IDH) - v(AKGD)$$

$$\frac{d[succoa]}{dt} = v(AKGD)$$

$$v(CS) = \frac{k_{cat}^{CS} E_T^{CS}}{1 + \frac{K_M^{accoa}}{[accoa]} + \frac{K_M^{oaa}}{[oaa]} + \frac{K_M^{accoa}}{[accoa]}\frac{K_M^{oaa}}{[oaa]}}$$

$$v(ACO) = k_f^{ACO}\left([cit] - \frac{[icit]}{K_E^{ACO}}\right)$$

**Parameters:**

$k_{cat}^{CS}, E_T^{CS}, K_M^{accoa}, K_M^{oaa}, k_f^{ACO}, K_E^{ACO}, etc.$

Initial concentrations for all metabolites included in the model

**Possible results:**

Steady state rate for every reaction; changes of [akg] over time; changes of CS reaction rate as a function of [accoa] or [oaa]; changes of [akg] as a function of changes in [accoa]; etc.

**Figure 3.1: Constraint-based and kinetics-based representations for four reactions in the TCA cycle**
The constraint-based approach requires the presence of a theoretical boundary (dashed line) to distinguish the inside and outside of the system. The **S** matrix is written for the four internal reactions and three exchange reactions (EX1, EX2, EX3), which allow the metabolites (accoa, co2, and nadh) to cross the boundary. Flux balance analysis is applied to find a flux distribution *v* that maximizes 'nadh' output. The mass-balance constraint is represented by $d[X]/dt$, where $[X]$ is a vector of metabolite concentrations. The time derivative $d[X]/dt$ represents a system of 12 differential equations, each similar to the differential equations shown for the kinetics-based model. The system of equation accounts for all 12 metabolites in the **S** matrix. Kinetics-based descriptions for CS and ACO are shown here as examples; similar expressions are required for the rest of the enzymes in the model and can be found in Cortassa *et al.*[71]. There are two notable differences in the two approaches: 1) The relationship between substrate concentrations and reactions is linear in the constraint-based approach, but highly nonlinear in the kinetic counterpart and 2) The requirement of the large number of kinetic parameters is eliminated in the constraint-based model. Consequently, a kinetic model is useful when enzyme mechanisms are known and immediate responses of the system to perturbations are of interest. On the other hand, due to the much simpler mathematics involved, constraint-based models can be applied to large scale networks and are useful when steady-state responses are relevant.

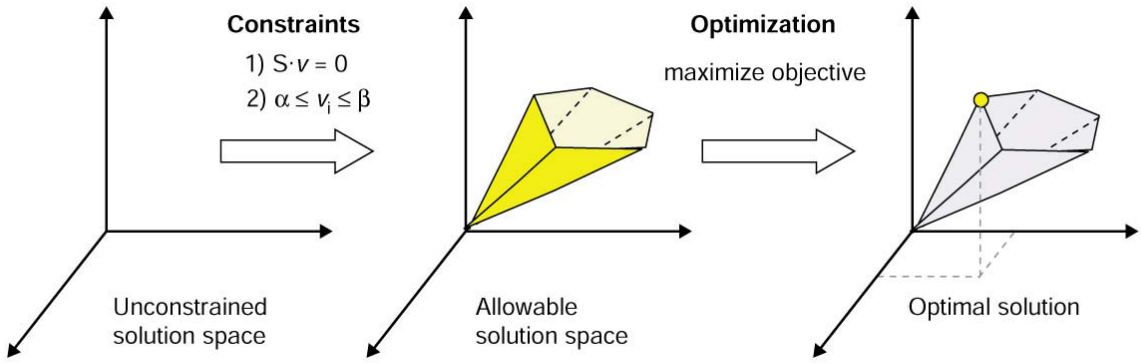



**Figure 3.2: Application of linear programming to search for an optimal flux distribution**
The result of constraints 1-2 is a capped polytope that lies in the null space of S. Any point in this polyhedron is a mathematical valid solution for the vector flux *v*. Since there are an infinite number of points in this space, a hypothesis is needed to identify the most physiologically relevant solutions. The underlying hypothesis of the linear programming method is that the biological system occupies a steady state that is optimal, in terms of resource distribution, for its designated function.

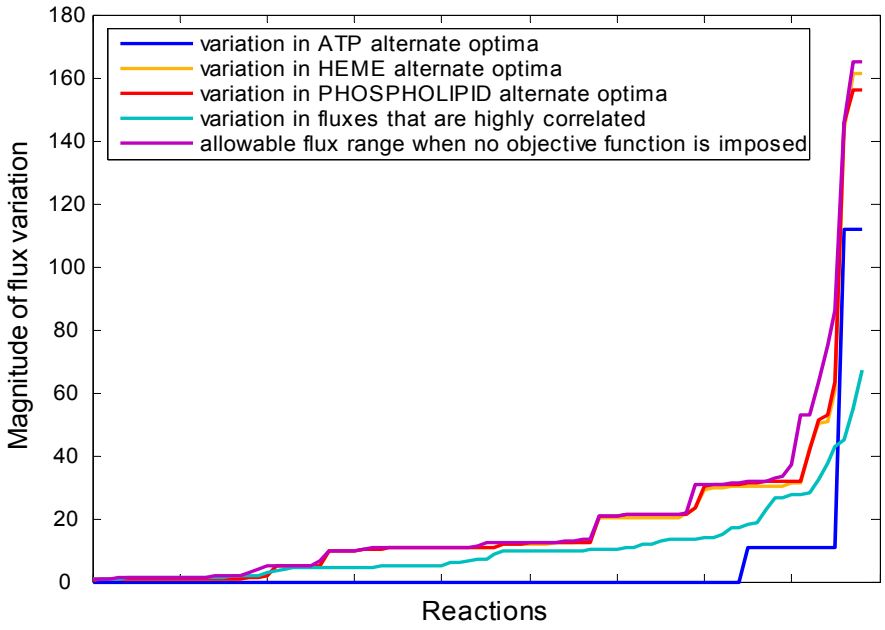



**Figure 3.3: Range of flux variation**
The allowable flux range for each reaction is ordered by magnitude on the x-axis. The flux range is found across flux distributions that are optimal for each individual metabolic objective function and among flux distributions that are highly correlated with each other. The allowable flux ranges when no objective function is imposed is also shown for comparison. Values in the y-axis indicate the magnitude of flux variation of a particular reaction. For clarity, the hundred reactions with the smallest flux variations are omitted here. The mitochondrial proton transport (reaction Htm) with flux variation higher than 400 is also omitted.

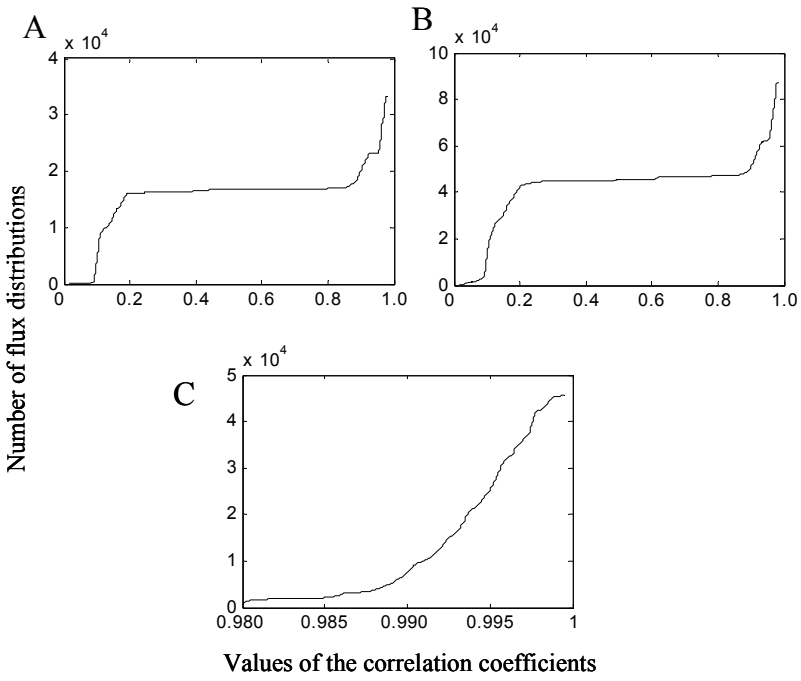



**Figure 3.4: Cumulative distribution of correlation coefficients between pairwise flux distributions**
Pairwise correlations are calculated between A) ATP and heme optimal solutions, B) ATP and phospholipids optimal solutions, and C) highly correlated heme and phospholipid flux distributions. Each point on the plot corresponds to a particular pairing of two flux distributions.

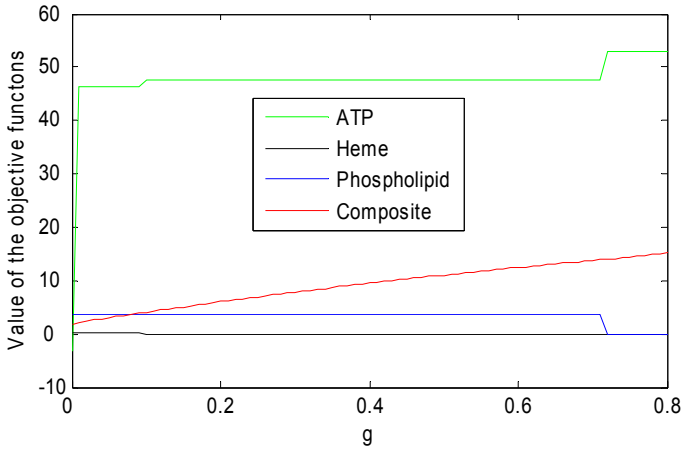



**Figure 3.5: Results of multiple objective analysis.**
Value of composite and individual objective functions are shown as a function of g. For values of g > 0.72, the syntheses of both heme and phospholipids equal zero, and the production of ATP achieves its maximal value.

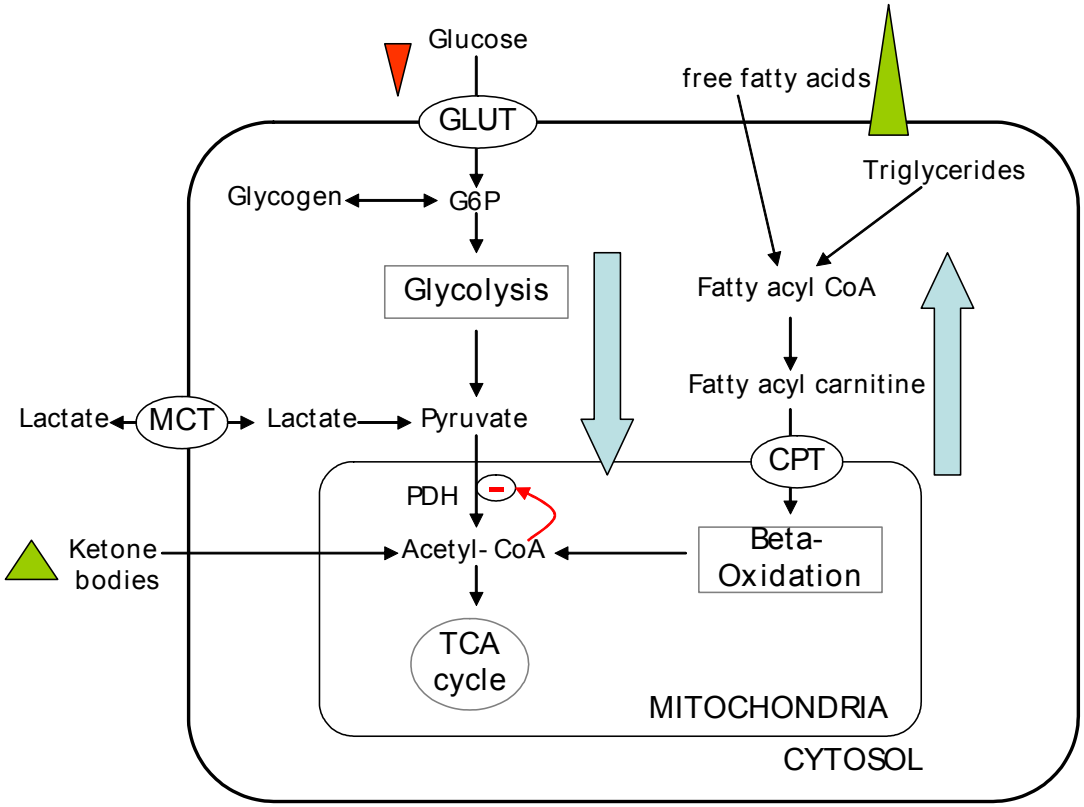



**Figure 3.6: Metabolic disturbance in diabetes**
Diabetic patients have an impaired ability to uptake plasma glucose, which results in increased fatty acid oxidation (either because of the absence of insulin (Type I) or the low level of Acetyl-CoA or Malonyl-CoA). B-oxidation produces Acetyl-CoA, which inhibits PDH further. This diabetic condition is simulated by *i*) reducing maximal glucose uptake, *ii*) increasing the lower bound on fatty acid uptake, and *iii*) removing explicit constraints on ketone body uptake because plasma ketone body levels are very high in these conditions.

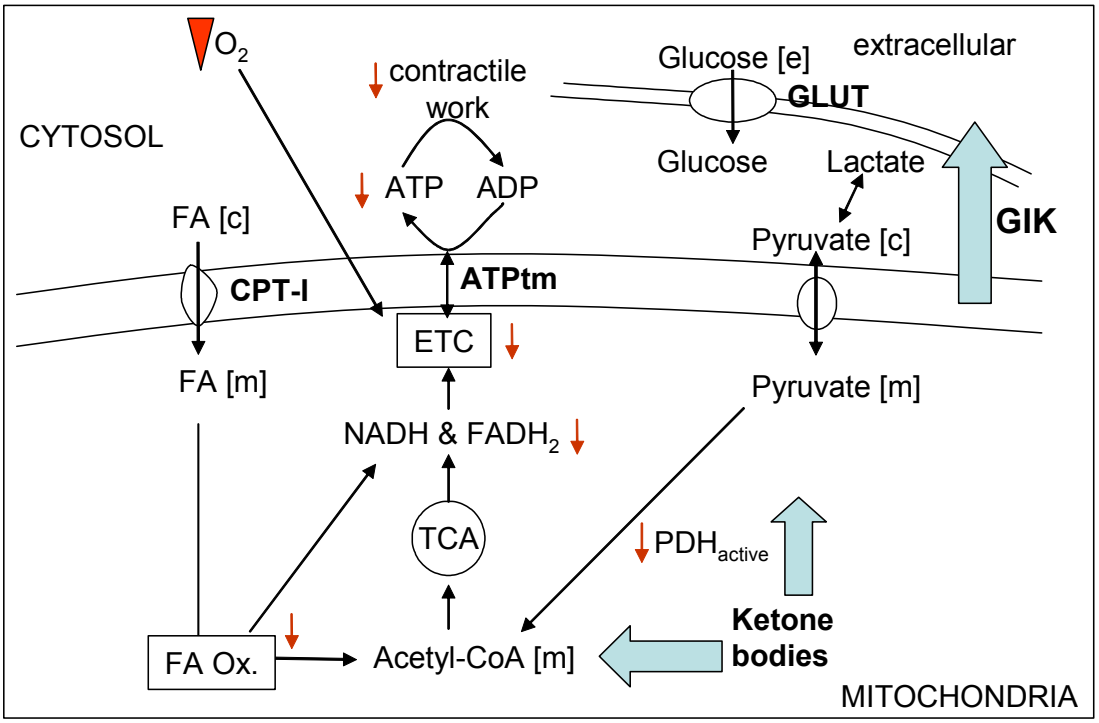
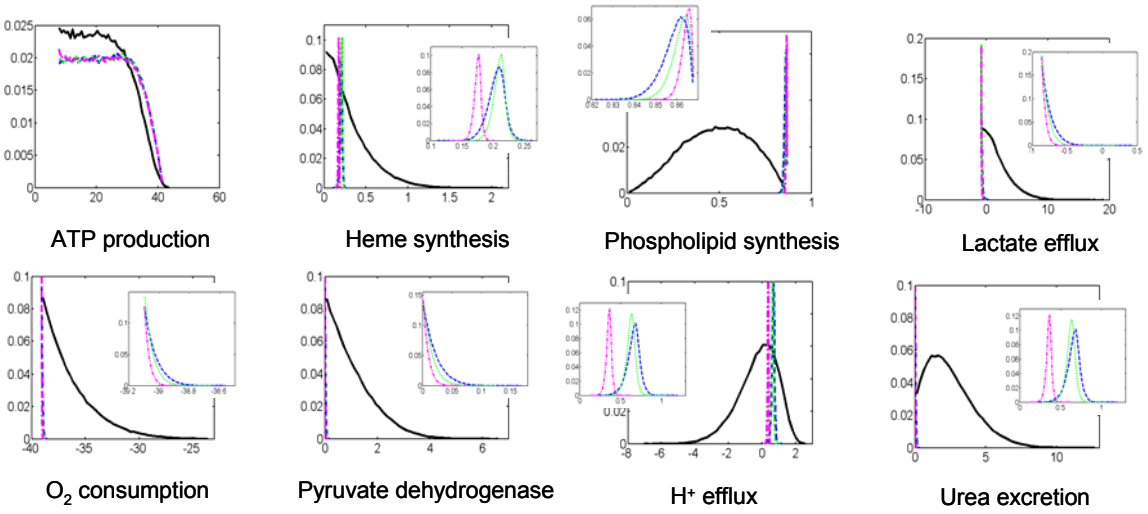



**Figure 3.7: Metabolic disturbance in moderate ischemia**
A decrease in blood flow and $O_2$ supply lead to a transient decrease in aerobic carbohydrate and fatty acid oxidation. This reduces the level of high energy phosphate molecules (both ATP and creatine) available for contractile work. At the same time, glucose influx increases (ischemia, like insulin, leads to increased number of glucose transporter), but the conversion of Acetyl-CoA via the TCA cycle and respiratory chain is reduced. Therapeutic approaches such as GIK (Glucose, Insulin, Potassium infusion) aim to increase immediate ATP level through glycolysis. Side effects include increased lactate and proton production. An alternate approach is to increase ketone body consumption as ketone bodies yield better ratio of ATP produced per O2 consumed than fatty acids. Effects of ischemia and these two therapeutic approaches are implemented *in silico* by *i)* reducing oxygen uptake by 25%, *ii*) increasing maximum glucose uptake rate to 2 µmol/min/g protein, and *iii*) increasing the lower-bound constraint on the uptake rates of both types of ketone bodies.




ATP production slightly reduced
Excess fatty acid increases phospholipid synthesis
Minimal pyruvate dehydrogenase activity
Restricted overall network activity
Effects of normalized glucose and ketone body uptake are positive but small.

Normal condition
Diabetic condition
"Normalized" glucose
"Normalized" glucose + ketone bodies

**Figure 3.8: Metabolic states under normal condition and disease conditions**
Each plot is a histogram of the different flux values a reaction can take while satisfying all constraints in the network. The x-axis represents the allowable flux value for the corresponding reaction. The y-axis shows the likelihood the reaction will assume that particular flux value. The overall result from the four conditions shown above is that the increased mitochondrial fatty acid uptake leads to smaller ranges of allowed flux rates for most reactions, and therefore reducing network flexibility dramatically. The ATP production capability of the mitochondrion in diabetic condition only differs slightly from the normal physiological condition. Due to the high mitochondrial fatty acid breakdown, the flux distribution of oxygen uptake was increased and its flexibility was very restricted. Another interesting finding is that the flux through the mitochondrial pyruvate dehydrogenase enzyme is significantly restricted by network stoichiometry when fatty acid uptake was increased. This behavior is present in the absence of explicit inhibition by pyruvate kinase. Lastly, although changing ketone body and glucose uptake to normal values leads to a higher flexibility of reaction fluxes, the effect is minimal compared to the normal condition.

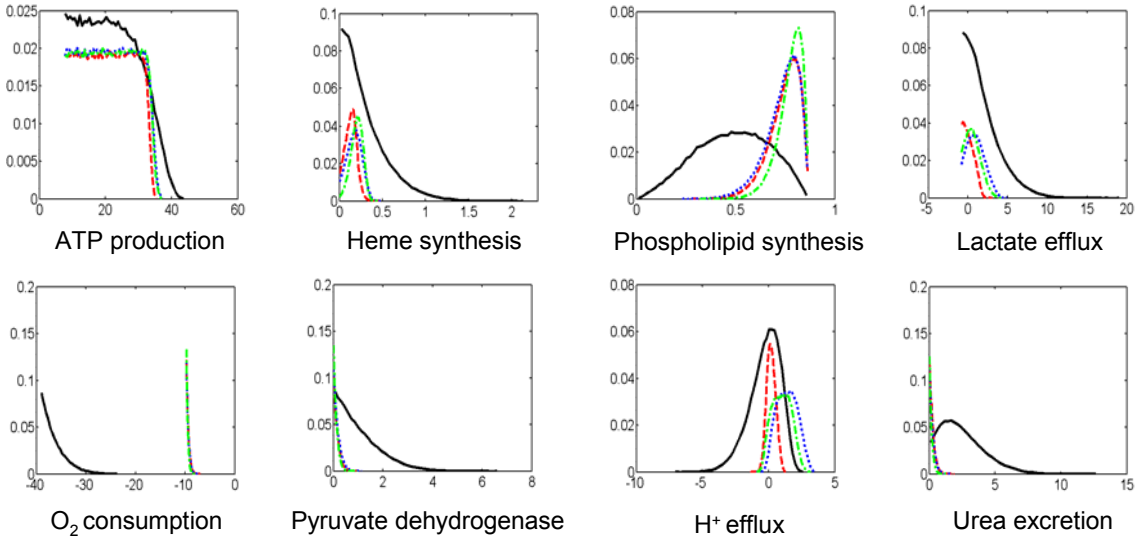



**Figure 3.9: Metabolic states under normal condition and ischemic conditions**
Each plot is a histogram of the different flux values a reaction can take while satisfying all constraints in the network. The x-axis represents the allowable flux value for the corresponding reaction. The y-axis shows the likelihood the reaction will assume that particular flux value. Flux distributions show that both the maximally allowed and the most probable flux values were reduced in most reactions; exceptions were reactions of the phospholipid biosynthesis pathway where the most probable flux values were increased. Such increase is a result of the build-up of non-oxidizable fatty acids in the absence of adequate oxygen. Effects of two therapeutic approaches, GIK and GIK in combination with ketone bodies, are also investigated here (Figure 3.7). Both approaches seem to alleviate damage caused by ischemia: ATP production and the range of allowable flux values for most reaction increase slightly. However, effluxes of protons and lactate are shifted higher, corresponding to two well known side effects of these therapies. These side-effects may lead to further damage during reperfusion.

# Chapter 4
# Integration of Isotopomer Data into the Constraint-Based Framework

The metabolic phenotype of a living cell is the cumulative result of the genomic, transcriptomic, and proteomic responses of the cell to a particular condition. The field of metabolomics seeks to integrate metabolite identification and reaction kinetics to elucidate cellular metabolic phenotypes. Recent efforts to detect and measure concentrations of a large number of low molecular weight compounds have achieved success with gas chromatography-mass spectrometry (GC-MS), liquid chromatography-electrospray ionization-mass spectrometry (LS-ESI-MS), and capillary electrophoresis coupled to mass spectrometry (CE-MS)[133]. Similar to other component-identification research areas, e.g. genomics and proteomics, it is perhaps expected that the comprehensive set of cellular metabolites will eventually be elucidated. The large-scale quantification of reactions kinetics in the cell (fluxomics) is more complex, however, as *in vivo* experiments are not possible and *in vitro* experiments often fail to mimic the molecular crowding environment in the cell[5]. On the other hand, fluxes in exchange with the extracellular environment, i.e. substrate uptake or secretion rates, can be satisfactorily measured and have been applied to microbial organisms[55,134], tissue cultures[135], and organ systems[136]. In addition, the use of tracer-based metabolomics, particularly with $^{13}C$ labeled substrates, has proven successful in inferring intracellular reaction fluxes[137].

Accordingly, the most promising approach for elucidating cellular nutrient-gene interactions will likely require a combination of metabolite profiling, extracellular flux measurements, and an effective method for analyzing isotopomer data.

Analysis of isotopomer data obtained from tracer experiments relies on the relative distribution of $^{13}C$ among the different metabolites and the different mass isotopomers of a metabolite to infer intracellular fluxes in the reaction network. In particular, many existing isotopomer studies follow the $^{13}C$ labels originating from a particular compound to analytically solve for the rates of reactions or pathways associated with that compound[138-144]. These analytical methods, however, have at least two major limitations. First, the quantification of a particular flux or flux ratio requires *a priori* assumptions about which pathways are considered relevant and the direction of flux in such pathways. Such assumptions are difficult to make correctly when biochemical pathways or cycles of interest contain many reversible reactions. Second, conclusions drawn from these analyses provide little insight into the mechanism of how the cellular metabolic system exhibits the observed phenotypes. As a result, this *pathway-based* method[145] does not provide a means to predict changes in intracellular metabolic phenotypes due to changes in uptake (input) and secretion (output) rates of nutrients. As more isotopomer data become available at a cellular scale, a more general and systemic approach for flux analysis is desirable.

More rigorous modeling methods have been proposed, many of which employ an optimization framework to search for globally optimal flux distributions that produce the

observed $^{13}C$ labeling patterns. Concepts of atom mapping matrices (AMMs)[146], isotopomer mapping matrices (IMMs)[147], bondomer mapping matrices[148], isotopomer matrices[149], and T matrices[150] have also been introduced to facilitate bookkeeping of different isotopomer states and formulation of balance equations that are amenable to different types of tracer data. Within these formalism, models of various sizes have been developed to study the metabolism of *Escherichia coli*[151,152], *Bacillus subtilis*[153], *Methylobacterium extorquens*[154], *Saccharomyces cerevisiae*[155,156], *Penicillium chrysogenum*[157,158], and *Corynebacterium glutamicum*[159,160].

In this chapter, isotopomer mapping matrices and isotopomer balance equations are incorporated as constraints in the constraint-based framework for flux analysis. This approach, here termed *network-based*, has two fundamental differences from the earlier pathway-based analysis. First, the method requires a well-defined system boundary encapsulating the biochemical network and permitting the distinction between intracellular fluxes and fluxes in exchange with the environment. Second, the method relies on numerical optimization rather than analytical tools to determine reaction fluxes in the network. The trade-offs between the network-based and the pathway-based methods are the computational complexity and the comprehensiveness of results inferred from observed isotopomers.

## 1. What are isotopomers?

Isotopomers of a metabolite are isomers that are different only in isotope

composition (Figure 4.1). Tracer-based metabolomics uses stable isotopes containing $^{13}C$ as labels to distinguish experimentally supplied carbons from those endogenous in the cell. The $^{13}C$ from substrates are transferred to intracellular metabolites; any metabolites that only differ in their $^{13}C$ labeling patterns are isotopomers of each other. A metabolite of *n* carbons may have up to $2^n$ isotopomers. Since $^{13}C$ has a different mass than natural $^{12}C$, mass isotopomers can also be defined to distinguish isotopomers with different masses. For examples, a pyruvate molecule ($C_3H_3O_3$-) labeled with one $^{13}C$ has a mass of 88 Daltons, whereas a pyruvate molecule labeled $^{13}C$ in all three carbons has a mass of 90. A given metabolite of *n* carbon thus has up to *n*+1 mass isotopomers.

## 2. Constructing a metabolic network

The construction of a model for isotopomer analysis is a four-step process beginning with the definition of a metabolic network (Table 4.1). Chapter 2 describes in detail the steps involved in the reconstruction of a metabolic network. This network should be as complete as possible to obtain the highest resolution on the final flux estimation. The metabolic network is then converted to a stoichiometric matrix S, where S(*i,j*) denotes the participation of metabolite *i* in reaction *j*[22]. Three metabolic networks, cardiomyocyte, HepG2 and fibroblast, are used for isotopomer analysis. The complete studies and results are described in chapters 5 and 6.

## 3. Constructing Atom and Isotopomer Mapping Matrices

## 3.1 Identifying metabolites for which isotopomers are to be tracked in the model

One should track isotopomers for as many metabolites as possible. Every metabolite of *n* carbons adds *n* additional constraints to the network if its isotopomers are tracked (Section 3.3). However, three groups of metabolites should be excluded, as they either make the formulated nonlinear programming problem unsolvable in practical time or their isotopomers are uninformative. Specifically, they are

a) Metabolites which are *dead ends* or *effective dead ends* in the network. Dead ends are compounds that are only produced or only consumed by reactions in the network, and therefore they do not participate in any flux-carrying reactions at steady state. They are usually those that cross the scope of the model or represent a knowledge gap. Effective dead ends are compounds that only participate in reactions that contain dead ends, and hence do not participate in any flux-carrying reactions at steady state, either.
b) Very large metabolites. These are usually fatty acids or phospholipids. Since the number of theoretically possible isotopomers is exponential to the number of carbons in a metabolite, one should exclude all metabolites having more than 10 carbons to keep the size of the problem within the capability of most existing nonlinear programming solvers.
c) Metabolites that do not participate in carbon transferring reactions in the network. Examples of these are ATP, ADP, NADP, etc. If the network does not include

reactions that affect the carbon backbones of these metabolites, their isotopomer balance constraints will be uninformative.

## 3.2 Constructing atom mapping matrix for reactant-product pairs

Atom mapping matrices (AMM) are used to describe the carbon transfer for each pair of reactant and product of relevant reactions in the network. The original convention described by Zupke and Stephanopoulos[146] specifies a matrix **A** ($m$ x $n$), where $m$ is the number of carbons in the product, and $n$ is the number of carbons in the reactant. The element $A_{ij}$ is non-zero when carbon $j$ in the reactant can be transformed to carbon $i$ in the product, and zero otherwise. I found that a value of "1" should always be used instead of fractions for non-zero $A_{ij}$, because the convention of using fractions, as described by Zupke and Stephanopoulos[146], is inconsistent with the algorithm for computing isotopomer mapping matrices (IMMs) from AMMs as developed by Schmidt *et al.*[147].

## 3.3 Constructing isotopomer mapping matrices

Isotopomer mapping matrices (IMMs) describe whether or not a particular isotopic pattern of the product can be derived from a particular isotopic pattern of the reactant. An IMM has size $2^m$ x $2^n$, where $m$ and $n$ are the number of carbons in the corresponding product and reactant, respectively. The original algorithm described by Schmidt *et al.*[147] does not explicitly specify rules for constructing IMMs for reactions with symmetric compounds or reactions with multiple reactants or products of the same species. I extended this algorithm to account for these two cases. For reactions with

symmetric products, the final IMM should be the average of IMMs constructed for each mirror image. For condensation reactions with multiple reactants of the same species, the final IMM is the sum of those created for each of the reactant molecules. Details and illustrative examples for each of these cases can be found at http://systemsbiology.ucsd.edu/organisms/. All atom and isotopomer mapping matrices used in this dissertation are also available for download at that website.

# 4. Formulating variables and constraints on reaction fluxes

There are two types of variables in the model: isotopomer distribution variables and flux variables. Three types of constraints – mass balance, bounds, and isotopomer balance – are used to narrow down and compute values for these variables. The variables and constraints are defined as follows.

## 4.1 Linear constraints

Each isotopomer distribution variable is a vector of size $2^n$, where $n$ is the number of carbon atoms in the corresponding metabolite. Each element of an isotopomer distribution vector (IDVs) describes the relative abundance of a particular isotopomer of the metabolite; these elements sum to one. The flux vector $v$ contains the second set of variables that need to be determined. The mass balance constraint ensures that the time derivative of each metabolite in the network is zero at steady state,

$$\mathbf{S} \cdot v^{net} = 0 \qquad \text{(Equation 4.1)}$$

where **S** is the stoichiometric matrix and $v^{net}$ is a vector of unknown net fluxes. Constraints on reactions rates are linear, $\alpha_i \leq v^{net} \leq \beta_i$, where $\alpha_i$ and $\beta_i$ represent lower and upper bounds on the steady state rate of each reaction in the network.

## 4.2 Nonlinear constraints

Similar to the mass-balance equation described above, the isotopomer balance equations ensure that production and consumption of every isotopomer of every metabolite are equal at steady state. Isotopomer balance constraints can be algorithmically generated from the S matrix and IMMs[147] (Figure 4.2). In addition, as both the forward ($v^{forward}$) and the reverse ($v^{reverse}$) directions of a reversible reaction can affect the observed isotopomer distributions of the reaction's reactants and products, these variables are incorporated in the isotopomer balance equations (Equation 4.2). These variables are also transformed as previously described[154,161].

$$v^{net} = v^{forward} - v^{reverse} \qquad \text{(Equation 4.2)}$$

$$with\ \alpha_i \leq v^{net} \leq \beta_i;\ 0 \leq v^{forward} \leq |\beta_i|;\ 0 \leq v^{reverse} \leq |\alpha_i|$$

It may not be necessary (or possible[162]) to track the rates of both the forward and the reverse directions of every reversible reaction in the network. In a few specific cases, only the net fluxes of these reversible reactions affect the isotopomer distributions of involved metabolites. Criteria for identifying these reactions are available at http://systemsbiology.ucsd.edu/organisms/.

# 5. Solving for optimal flux distributions

The set of isotopomer-balance constraints, in combination with Equations 4.1-4.2, formulates a nonlinear programming problem, where the objective (Equation 4.3) is to minimize the difference between the predicted and observed mass distributions of isolated metabolites:

$$Min\ \ Error = \sum_{i}^{M} \sum_{j}^{N(i)} \left( \frac{MDV_{i,j}^{measured} - MDV_{i,j}^{calculated}}{\sigma_{i,j}} \right)^2 \qquad \text{(Equation 4.3)}$$

In Equation 4.3, MDV is the mass distribution vector of a particular metabolite *i;* each element $MDV_{i,j}$ contains the mole fraction of isotopomers of the same mass, and $\sigma_{i,j}$ is the standard deviation associated with that measurement. The index N(*i*) represents the total number of mass isotopomers of metabolite *i* and the index M represents all metabolites measured in the experiment. Values for measured mass isotopomer distributions are corrected for naturally occurring isotopes[163] before they are substituted in the equation.

The formulated model is implemented in GAMS (GAMS Development Corporation, Washington, DC) and solved using the commercially available solver SNOPT (Stanford Business Software, Inc., CA). Since the model is nonlinear, the predicted flux distributions are locally optimal solutions of the posed optimization problem. The landscape created by the isotopomer balance constraints is generally rough, so convergence to a particular locally optimal solution is dependent on both the

initial starting point and the computer on which the program is run. It is therefore critical that the calculation is repeated for a large number of initial points[164]. Flux results presented in the following studies are the results of 1000 repeated calculations, whose initial points are selected from $10^7$ feasible steady state flux vectors. To limit the analysis to only the most likely flux distributions, flux distributions with *Error* values larger than certain threshold, e.g. more than twice the smallest *Error* value found, should be discarded; the remaining are considered acceptable solutions.

(This chapter, in part, is a preprint of the material appearing in “"Isotopomer analysis of cellular metabolism in tissue culture: A comparative study between the pathway- and network-based methods," Vo TD, Lim SK, Lee PWN, and Palsson BO. 2006. *Metabolomics,* in press and “Isotopomer analysis of myocardial substrate metabolism: A systems biology approach,” Vo TD and Palsson BO. 2006. *Biotechnology and Bioengineering* 95(5):972-83. The dissertation author is the primary investigator and author of these papers.)

**Table 4.1: Procedures to develop a constraint-based model for intracellular flux estimation based on reaction stoichiometry, substrate uptake and efflux rates, and isotopomer data.**
Additional descriptions of each step are provided in the manuscript and cited references

**Step 1**. **Construct a stoichiometric matrix S representing the biochemical reactions in the network**

**Step 2**. **Construct atom and isotopomer mapping matrices**

A. Identify metabolites for which isotopomers are to be tracked in the model, note symmetric metabolites
B. Construct atom mapping matrices (AMMs) for reactant-product pairs of metabolites in step A
C. Compute an isotopomer mapping matrix (IMM) for each AMM

**Step 3**. **Formulate constraints**

A. Linear constraints:

$\mathbf{S} \cdot v^{net} = 0$

**S** is the stoichiometric matrix; $v^{net}$ is a vector of unknown net fluxes

For all irreversible reactions *i*

$\alpha_i \le v^{net} \le \beta_i$

$\alpha_i$ and $\beta_i$ represent the lower and upper bounds on the steady state reaction rates

For all reversible reactions *j*

$$v_j^{net} = v_j^{forward} - v_j^{reverse}$$

$$\alpha_j \le v^{net} \le \beta_j; 0 \le v^{forward} \le \left|\beta_j\right|;\ 0 \le v^{reverse} \le \left|\alpha_j\right|$$

For all metabolites *k* with carbon tracking,
*C* is the number of carbon atoms in metabolite *k*.

$$\sum_{i=1}^{C(k)} IDV_i^k = 1$$

B. Nonlinear constraints: isotopomer balance equations

$$F_k(IDV_i^k, v) = 0$$

where the function $F_k$ for each metabolite *k* is defined in Schmidt *et al*. 1997

**Step 4**. **Solve for optimal flux distributions**

A. Pick an initial starting $v_o$ satisfying $\mathbf{S} \bullet v^{net} = 0$
B. Solve

$$Min \quad Error = \sum_i^M \sum_j^{N(i)} \left( \frac{MDV_{i,j}^{measured} - MDV_{i,j}^{calculated}}{\sigma_{i,j}} \right)^2$$

*Subject to* Constraints 3A-3B.

$MDV_{i,j}$ : mole fraction of mass isotopomer *j* of metabolite *i*
$\sigma_{i,j}$ : standard deviation associated with that measurement $MDV_{i,j}$
N(*i*) : total number of mass isotopomers of metabolite *i*
M : number of metabolites measured in the experiment.

C. Repeat 4A-4B for a sufficiently large number of flux distributions

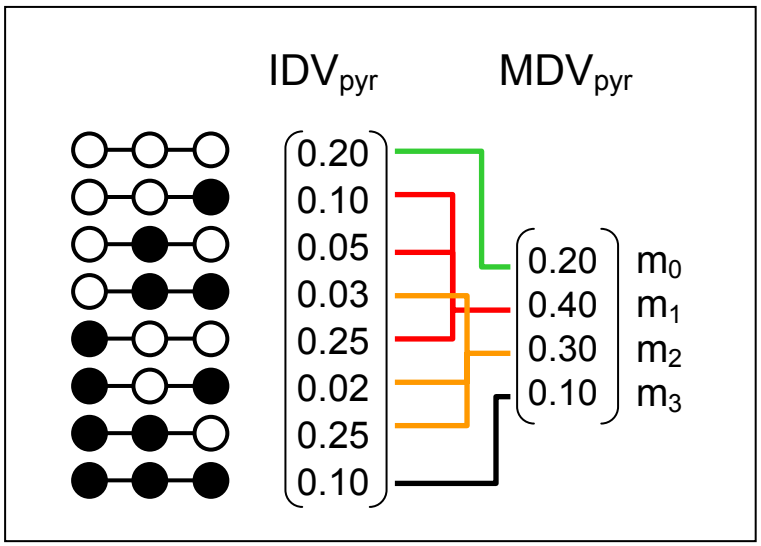
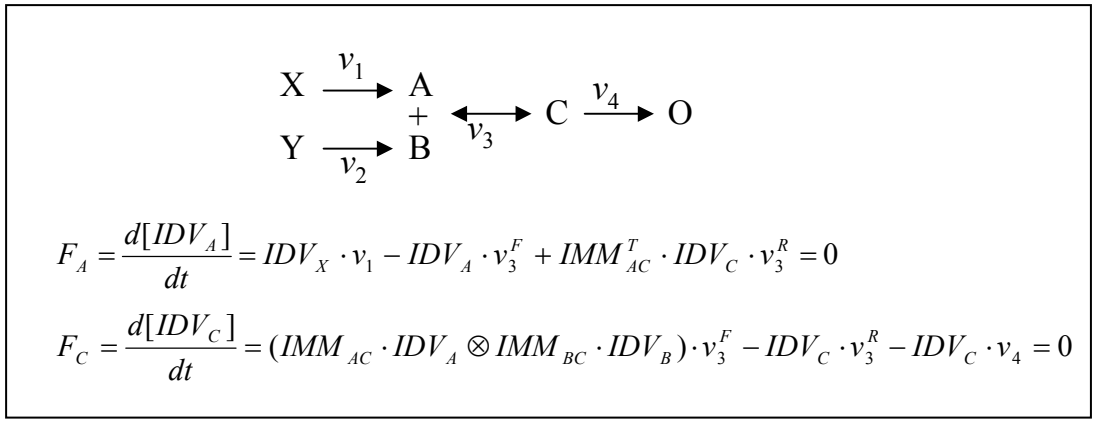



**Figure 4.1: All possible isotopomers of a metabolite with three carbons**
IDVpyr represents a sample isotopomer distribution vector for pyruvate. Each element in IDVpyr indicates the fractional contribution of the isotopomer shown in the left to the total pyruvate pool. MDVpyr shows an example of a mass distribution vector for pyruvate. Elements in MDVpyr are sum of the contributions of isotopomers with the same number of labeled carbons. $m_i$ correspond to the mass isotopomer with *i* labeled cabons. This figure is reproduced with permission from the author of reference 186.




$$F_A = \frac{d[IDV_A]}{dt} = IDV_X \cdot v_1 - IDV_A \cdot v_3^F + IMM_{AC}^T \cdot IDV_C \cdot v_3^R = 0$$

$$F_C = \frac{d[IDV_C]}{dt} = (IMM_{AC} \cdot IDV_A \otimes IMM_{BC} \cdot IDV_B) \cdot v_3^F - IDV_C \cdot v_3^R - IDV_C \cdot v_4 = 0$$

**Figure 4.2: Formulation of isotopomer balance constraints**
The two equations $F_A$ and $F_C$ show the isotopomer balance constraints for these two metabolites, respectively. Such constraints set the time derivative of each isotopomer of each of these metabolites to zero at stead state. $IMM_{XY}$ stands for isotopomer mapping matrix describing the carbon transfer from reactant A to product C. Rules for formulating these constraints can be found in Schmidt *et al.* [147].

# Chapter 5
# Isotopomer analysis of myocardial substrate metabolism: Computational and Experimental Considerations

The quantification of reaction fluxes in cellular metabolism has always been of great interest in physiological and biotechnological research[165,166]. Metabolic flux profiles can uncover details about substrate utilization, substrate redistribution at network branch points, and quantitative information about enzyme activity. As intracellular flux measurements tend to be invasive and difficult, our current ability to profile metabolic flux relies on computational tools to analyze experimental data. Studies have shown that isotopomer data, especially $^{13}C$ tracer data, are useful and effective for estimating intracellular reaction fluxes[167-169]. In particular, mass isotopomer analysis has been extensively applied to study substrate oxidation and anaplerosis in the heart[141,142,170,171], gluconeogenesis[143,172,173] and lipogenesis in the liver[144,174], and activities of the TCA cycle in various tissues[175-177]. Reaction flux estimates in these studies were analytically derived based on observed isotopomer data, assuming a configuration and a direction of flux flow in relevant pathways. Such analytical solutions are restricted to small model networks and are not obtainable for many biochemical pathway structures[178].

In this study, isotopomer mapping matrices and isotopomer balance equations (Chapter 4) are incorporated into the constraint-based framework to analyze isotopomer data obtained from perfused mouse hearts[136]. Advantages of this approach as compared to the analytical expressions originally employed by Khairallah *et al*.[136] are twofold. First, the use of a cohesive model ensures that estimations of intracellular fluxes are consistent with both isotopomer data and flux measurements obtained from experiments. The incorporation of known myocardioal metabolic activities and stoichiometry of the underlying biochemical reactions also provides a more complete picture of how the entire cardiomyocyte metabolic network operates and how fluxes in different pathways fit together. The resulting flux distribution offers a systemic view of cellular metabolism as opposed to glimpses of fluxes or flux ratios calculated separately and possibly under independent assumptions. Second, the models can be used to simulate and analyze experimental scenarios beyond the original experimental conditions. In particular, the model is used to study the effectiveness of 12 different $^{13}C$ glucose substrate mixtures and to identify the most informative metabolites and fluxes to measure in subsequent experiments.

# 1. Cardiomyocyte model

## 1.1 Isotopomer data from the perfused mouse heart

The work of Khairallah *et al*.[136] sought to characterize and trace the origin of pyruvate and citrate carbons in working mouse hearts. Four types of labeled substrates,

perfused at their respective physiological concentrations, were employed in that experiment; here only the isotopomer data from the experiments with uniformly labeled glucose, [U-$^{13}C_6$] glucose, (50% enrichment) are analyzed.

Three types of information are accounted for when integrating these data into the model. First, the reported molar enrichment of perfused glucose is used to set the isotopomer distribution of extracellular glucose. Second, upper and lower bounds on the uptake and efflux rates of lactate, pyruvate, glucose, citrate, succinate, oleate, oxygen are set to two standard errors above and below the mean (Table 4.1). Third, the $^{13}C$ enrichment of the TCA cycle intermediates (citrate, α-ketoglutarate, succinate, fumarate, malate, oxaloacetate moiety of citrate) and their standard errors are used to formulate the objective function. These data were obtained from GC-MS, and the final enrichment, corrected for $^{13}C$ natural abundance, of each mass isotopomer was reported[136]. The average mouse heart wet weight[136] is used to convert reported data into a consistent flux unit (μmol/min/gww).

## 1.2 Size and scope of the model

The present cardiomyocyte metabolic model accounts for 240 metabolites and 257 reactions, of which 39 are exchange reactions[179]. The rank of the corresponding stoichiometric matrix is 221. There are thus 36 reaction fluxes that have to be determined by using isotopomer data in addition to the mass-balance constraint. Reactions in this network describe glycolysis, the TCA cycle, oxidative phosphorylation, ROS

detoxification, anaplerosis, β-oxidation, ketone body metabolism, heme synthesis, and phospholipid synthesis. These reactions are written at the same level of detail as those in previous reconstructed networks[46] (Chapter 2). Five assumptions are made in using this model to analyze isotopomer data from the perfused mouse heart:

1) The 257 reactions included in the model are sufficient to describe the major metabolic activity in the perfused mouse heart.
2) The labeling of substrates with $^{13}C$ does not affect how they participate in a reaction, i.e. a metabolite is not selected against or preferred by an enzyme due to its labeling state.
3) The flux distribution that yields $^{13}C$ labeling patterns most resembling the observed isotopomers (of the isolated metabolites) is the physiological flux distribution in the cell.
4) The perfused mouse heart achieves a steady state during the course of the experiment.
5) There is label scrambling in reactions involving symmetric metabolites.

AMMs[146] and IMMs are employed to track carbon transfer between products and reactants (Chapter 4). Carbon transfer for 121 reactant-product pairs, associated with 79 metabolites, is tracked in the model. This translates to 1700 isotopomer variables. Carbons of the remaining metabolites are not tracked for one or more of the three reasons described in Chapter 4. IMMs for the 121 reaction-product pairs are available for download in MATLAB (The MathWorks, Inc., Natick, MA) matrix format at http://systemsbiology.ucsd.edu/organisms/. A database of such IMMs can be a valuable

repository of unambiguously defined reaction mechanisms. Note that a large number of these IMMs are identity matrices as most biochemical reactions do not involve carbon rearrangement. All identity IMMs are excluded from the isotopomer balance constraints used in this dissertation to avoid unnecessary matrix computation.

The isotopomer balance constraints are algorithmically generated from the stoichiometric matrix and IMMs[147]. Both forward and the reverse directions of a reversible reaction, which affect observed isotopomer distributions of the reaction's reactants and products, are incorporated into isotopomer balance equations. The described constraints and variables are concisely formulated into a nonlinear programming problem, whose objective is to minimize the difference between measured and calculated mass distributions (Chapter 4). Values for measured mass isotopomer distributions are corrected for naturally occurring isotopes before they are used in the model[136,180]. The sequential quadratic programming method[181] implemented in the commercially available solver SNOPT (Stanford Business Software, Inc.) is chosen to solve the formulated nonlinear programming problem. Network contents and computational programs used in this study are available at http://systemsbiology.ucsd.edu/organisms/.

## 2. Sensitivity analysis

The SNOPT solver searches for a locally optimal solution starting from a specified initial point ($v_o$) in the steady state flux space. In searching for a globally

optimal solution, one must sample the entire solution space for locally optimal solutions. The larger the number of locally optimal solutions found, the more likely that one of them is the globally optimal solution. Estimated flux values resulting from these solutions are likely to be sensitive to two parameters: *i)* the user-defined initial points and *ii)* the measured mass distributions ($MDV^{measured}$). The effects of each of these parameters are investigated in this study. The initial values for $v_o$ are generated using two methods. The first method assigns $v_o$ to a flux distribution found by applying a linear programming solver (cplex) with the objective of maximizing or minimizing flux through a chosen reaction in the network. The isotopomer balance constraints (non-linear) are excluded in this step. The second method assigns $v_o$ to a random flux distribution within the convex space using the Hit-and-Run algorithm[94]. These two methods produce 489 and 1000 unique $v_o$, respectively.

To investigate the effect of the uncertainty associated with each isotopomer measurement, 100 random hypothetical measurements (for each mass isotopomer of each metabolite) are drawn from normal distributions simulated with the reported mean and standard error. Randomly selected values from these measurement pools produce 100 *hypothetical* mass distribution data sets based on measurement statistics. The calculation is repeated for these 100 data sets to evaluate how these mass distributions affected the predicted flux distributions.

## 3. Pyruvate branch points and fate

Pyruvate serves as an important branch point of substrate metabolism. It is thus useful to be able to quantify the contribution of various exogenous carbohydrate sources to tissue pyruvate. Cytosolic pyruvate is considered equivalent to tissue pyruvate for this purpose, as the mitochondrial pyruvate pool includes pyruvate produced by mitochondrial lactate dehydrogenase and malic enzyme. Based on the predicted uptake rates for glucose (3.00 ± 0.05 μmol/min/gww), pyruvate (0.43 ± 0.06 μmol/min/gww), lactate (0.30 ± 0.05 μmol/min/gww), and the following equations,

Exogenous glucose contribution = (2*glucose uptake)/total fluxes producing cytosolic pyruvate

Exogenous pyruvate contribution = pyruvate uptake/total fluxes producing cytosolic pyruvate

Exogenous lactate contribution = lactate uptake/total fluxes producing cytosolic pyruvate

Total fluxes producing cytosolic pyruvate = 2*glucose uptake + pyruvate uptake + lactate uptake

(Equations 5.1-5.4)

the estimated fractional contributions of these three exogenous substrates to cytosolic pyruvate are 80 ± 2, 8 ± 2, and 12 ± 2% respectively.

This estimate is rather different compared to the reported estimation by Khairallah *et al*.[136] (Table 4.2) because the latter study used a different method of analysis. Those authors combined data from three experiments, each with a different labeled substrate, and computed the contributions based on enrichment of m3 cytosolic pyruvate in each experiment. However, if they had used their reported flux estimates to compute the fractional contributions, the result would have been much closer to the values presented

here (Table 4.2). The method by Khairallah *et al*.[136] is more experimentally intensive, and may be affected by inconsistency and errors in these experiments. Such inconsistency is likely the reason that those authors could not account for the source that made up 26% of tissue pyruvate. On the other hand, the method described here is more computationally intensive, and does not account for contribution of substrates other than exogenous glucose, lactate, and pyruvate. These two methods are thus complementary for quantifying the contribution of exogenous carbohydrates to tissue pyruvate.

Since tissue pyruvate was only enriched in m3 isotopomer, it is concluded that the pentose phosphate pathway activity was low. The consumed $^{13}C$ glucose thus yields a stoichiometric amount of glycolytic pyruvate at a rate of 3.00 ± 0.05 μmol/min/gww. The model predicts that the majority of cytosolic pyruvate is converted to lactate (90%), which, in turn, is excreted by the cell. This result agrees with the observed m3 lactate efflux rate when mouse hearts were perfused with [U $^{13}C_3$] pyruvate[136]. The remaining pyruvate is transported into the mitochondria for further oxidation. Mitochondrial pyruvate has two major fates: oxidation by pyruvate dehydrogenase and anaplerosis by pyruvate carboxylase. The estimated steady state rate for pyruvate dehydrogenase is 0.25 ± 0.00 μmol/min/gww, while that for pyruvate carboxylase is 0.02 ± 0.00. A small amount of mitochondrial pyruvate (15%) is also inter-converted with lactate via mitochondrial lactate dehydrogenase and the pyruvate-lactate shuttle.

The *ex vivo* perfusion experiment allowed the working mouse heart to take up four substrates ([U $^{13}C$] glucose, pyruvate, lactate, and oleate) from a perfusate that was

optimized to mimic physiological serum[136]. At a respiration rate of 5.49 ± 0.06 μmol/min/gww, the heart takes up significantly more (ten times) exogenous carbohydrates than the fatty acid oleate. After accounting for efflux of pyruvate and lactate, however, oleate is found to have twice the amount of ATP contribution as glucose. Similar results were also found by Khairallah *et al*.[136], where the authors, using analytical expressions, reported a contribution of 62 ± 10% by fatty acids and 34 ± 4% by carbohydrates to the overall ATP production. The TCA cycle is predicted to carry an average net flux of 1.51 ± 0.05 μmol/min/gww, a value very similar to that found in rat hearts, 1.7 ± 0.2 μmol/min/gww[182]. Anaplerosis by pyruvate carboxylation is found to be relatively small, only 1% of the TCA cycle net flux. However, the difference between calculated and predicted isotopomer distributions suggests that anaplerosis by endogenous substrate is about 10% of the TCA cycle net flux. Finally, analysis of exchange fluxes shows that pseudoketogenesis can be a significant source of ketone bodies generated by the heart, amounting up to 50% of the overall ketone body uptake of the organ.

## 4. Activities of the citric acid cycle

Citrate is produced from oxaloacetate and acetyl-CoA in every turn of the TCA cycle. During the time course of the perfusion experiment (20-30 min)[136], most cellular energy is likely to be derived from substrates in the perfusate (glucose, pyruvate, lactate, and oleate). Contributions of amino acids from protein breakdown and lipid from membrane turnover are probably small, and hence not accounted for in the model. The

acetyl-CoA moiety of citrate ($ac^{cit}$) thus mostly comes from fatty acid or pyruvate decarboxylation, while the oxaloacetate moiety ($oaa^{cit}$) mostly comes from pyruvate carboxylation. The origin of each citrate moiety is evaluated using the ratios of pyruvate decarboxylation and pyruvate carboxylation fluxes to that of citrate synthase (CS). Based on calculated fluxes for PDH and pyruvate carboxylase, the pyruvate contribution to $ac^{cit}$ and $oaa^{cit}$ is estimated to be 17% and 1.4% of the TCA cycle flux (1.51 ± 0.05 μmol/min/gww). In comparison, the TCA cycle flux was previously estimated to be 1.88 ± 0.01 μmol/min/gww by Khairallah *et al*.[136] when a linear relationship between oxygen consumption and citrate formation from carbohydrates and fats was assumed.

Recall that since the experimentally measured isotopomer distributions of the six TCA cycle intermediates are used in the objective function, discrepancies between calculated mass distribution vectors (MDVs) and measured MDVs (Figure 5.1) offer a good estimate of the accuracy of the flux calculation. Three key differences are observed between calculated and measured MDVs. First, there is stronger agreement between predicted and experimental averages for m3 and m4 isotopomers, compared to m1 and m2, of the TCA cycle intermediates. The experimental data for m3 and m4 isotopomers have more precise values (smaller standard errors), and thus the model favors flux distributions that had better fit for these isotopomers (see the Error function in Chapter 4). Second, the overall higher $^{13}C$ enrichment predicted for most TCA cycle intermediates indicates that there is a low level of $^{13}C$ dilution (~10%) due to endogenous unlabeled carbon sources (most likely amino acids) not accounted for in the model. Third, the decrease in the total enrichment of α-ketoglutarate and succinate reflects a loss

of $^{13}C$ to $^{13}CO_2$. Lastly, the similar mass distributions calculated for succinate, fumarate, and malate are probably a consequence of *i*) fumarate being produced and consumed in the model only by succinate dehydrogenase and fumarase and *ii*) the assumed rapid isotopomer randomization for symmetric metabolites in the model. Labeling data from Khairallah *et al*.[136] do not have such a homogenous mass distributions for these three metabolites (Figure 5.1).

The TCA cycle, together with OxPhos, produces the majority of the ATP used for contractile function (Myosin ATPase) and various ion pumps in cardiomyocytes. Here all ATP consuming reactions are represented collectively as an ATP demand function, which describes the hydrolysis of the high energy phosphate bond of ATP to ADP and pyrophosphate. This way, ATP produced by metabolism of various substrate sources is coupled with an ATP consuming sink. The amount of ATP produced by anaerobic oxidation is 6.0 ± 0.1 μmol/min/gww, which is approximately 30% of the estimated total ATP production, 16.6 ± 2.3 μmol/min/gww. The total ATP production rate calculated from isotopomer data using this model is less than 40% of the maximal ATP production rate computed based on respiration rate and substrate uptake rates alone[46].

## 5. Bidirectional reaction rates

Many enzymatic reactions are recognized to be bidirectional, i.e. reversible, as they operate near equilibrium in cellular physiological conditions. Even for reactions with low net fluxes, both forward and reverse rates can be quite high, rendering these

rates unobservable during a typical experimental time scale. However, as both directions of the reactions affect the $^{13}C$ labeling patterns of reactants and products of the enzymes, it is possible to estimate these rates based on the isotopomers of these metabolites. In fact, one may incorrectly estimate the net flux of such reactions if forward and reverse directions of the enzymes are ignored.

Of the 95 reversible reactions in the model, 55 reactions are associated with metabolites whose isotopomers are tracked; these are the only reactions whose forward and reverse rates can be reliably estimated. The difference between the forward and the reverse rate, referred to as *exchange fluxes*[183], are predicted with precision for 49 reactions. Note that the term exchange flux used here is not the same as 'exchange reaction'[179], which is used to describe metabolite crossing the system boundary. Histograms of the exchange fluxes (across all the predicted locally optimal solutions) have dominant left peaks. Overall, exchange fluxes are of the same order of magnitude as the net fluxes, but tend to be slightly lower. Average net fluxes for the 49 reactions are found to be 0.46 μmol/min/gww, while average exchange fluxes are 0.41 μmol/min/gww.

Reversibility of reactions also provides information about the dynamics of flux patterns in a pathway. Seven out of ten reactions in the glycolytic pathway are considered reversible as they participate in both glycolysis and gluconeogenesis. However, as the heart does not have a high gluconeogenic activity, these reactions do not have high exchange fluxes; their exchange fluxes make up less than 50% of the net fluxes. The TCA cycle has five reversible reactions; three of them (succinate

dehydrogenase, malate dehydrogenase, and fumarase) have significantly higher exchange fluxes than the other two. The high exchange rates in these enzymes justify the isotopomer scrambling assumption in the model. Lastly, pseudoketogenesis, a process discovered by the label exchange between acetoacetate and acetoacetyl-CoA in rat hearts[184], is also observed in the exchange fluxes predicted by the model. The two reversible enzymes contributing to this pseudoketogenesis have the following net and exchange fluxes: thiolase (0.72 and 0.24 μmol/min/gww) and 3-ketoacyl-CoA transferase (0.41 and 0.13 μmol/min/gww). In comparison, the net uptake to ketone bodies are 0.27 (acetoacetate) and 0.46 (β-hydroxybutyrate) μmol/min/gww. Thus, pseudoketogeneis makes up as much as 50% of net ketone body uptake by the perfused mouse heart.

# 6. Properties of predicted flux distributions

## 6.1 Reducing the solution space

Complete estimates for net and exchange fluxes for reactions in the network can be found in Vo and Palsson[164]. The amount of information gained by the addition of isotopomer data can be estimated by comparing the estimated flux variation computed with and without the isotopomer balance constraints. The 149 reactions having no flux variation are not shown on Figure 5.2. Without isotopomer data, only about 20% of the remaining reactions (reactions with non-zero flux variation) can be predicted with reasonably small flux ranges. The application of isotopomer data, however, brings this number to 90%, a substantial improvement in flux estimation

## 6.2 Sensitivity with respect to user-defined initial points

Locally optimal solutions returned by SNOPT are dependent on the user-defined starting points. Starting points computed by the linear programming method, $v_{LP}^{net}$, and by the Hit-and-Run algorithm[94,185], $v_{rand}^{net}$, produce two sets of solutions, which are evaluated based on four characteristics: *i*) success in finding locally optimal solutions with the starting points; *ii*) values of the objective function *Error*; *iii*) correlation of the best solution (solution yielding the smallest error) with other locally optimal solutions; and *iv*) range of flux variation of each reaction across locally optimal solutions found. First, the SNOPT solver converges to locally optimal solutions for more than 80% of the initial points generated by the LP method, but only 50% with points generated by the ACHR algorithm. Second, the smallest errors found by both methods differ only by 0.1%. Comparing the two best solutions found with $v_{LP}^{net}$ and $v_{rand}^{net}$ respectively, only 14 reactions have flux values differing by more than 5% from each other. Overall, approximately 90% of all locally optimal solutions found by the two methods have very similar error values (less than twice the error of the best solution); the remaining 10% are outliers with significantly higher errors[164].

Third, further analysis is limited to solutions with errors that are within 5% of the smallest error. This way, predicted flux values provide the best estimates of physiological fluxes without over fitting the measured mass distributions. Among this group of flux distributions, it is determined that if two flux distributions are well correlated, then the individual reaction fluxes in the two distributions must be similar. As

expected, SNOPT is more likely to converge to the same optimal solution for $v_{LP}^{net}$ that maximizes or minimizes fluxes through reactions in the same pathways. In contrast, solutions found by $v_{rand}^{net}$ are less correlated with one another; their correlation coefficient values range from 0.5 to 1. Lower correlation among solutions found with the second method implies that there exist a number of reactions whose fluxes cannot be determined precisely. The high correlations seen with the first method are likely a result of incomplete sampling of the solution space.

Lastly, in assessing how much isotopomer data helps in determining reaction fluxes, reaction flux ranges are evaluated for groups of flux distributions computed with $v_{LP}^{net}$ and $v_{rand}^{net}$. Within the first group, 21 reaction fluxes have standard deviations larger than 10% of the mean flux values. Among solutions in the second group, 28 reactions have standard deviations larger than 10% of mean fluxes. The former 21 reactions are a subset of the latter 28 reactions, indicating that the second initialization method provides a more exhaustive list of reactions whose fluxes cannot be precisely determined by isotopomer data. Taken together, these results point to the following conclusions. If one is only interested in the solution with the best objective value, it is possible to find such solution with either method of initialization. However, if one is interested in studying how a set of isotopomer data narrows the range of allowable flux values for each reaction, the second method of initialization provides a more thorough answer.

### 6.3 Sensitivity with respect to experimental error

In order to investigate effects of uncertainty associated with each isotopomer measurement on results of the model, random isotopomer measurements are generated based on measurement statistics (normal distribution having reported mean and standard errors). The $v_{rand}^{net}$ starting point yielding the best error value found in the previous calculations is used as the starting point here. The resulting solutions are also assessed based on the four characteristics mentioned above. A total of 98 out of the 100 sets of isotopomer data produce locally optimal solutions. Similar to the previous results, 90% of these solutions have very similar objective values, while the remaining 10% have significantly higher error values. The best flux distribution found from the previous sensitivity analysis, $v^*$, is correlated with solutions found with these hypothetical isotopomer data as well as it is with solutions found with the original dataset[164]. It is thus concluded that uncertainty associated with isotopomer measurement errors do not significantly change the estimated fluxes, so long as such uncertainty is sufficiently small (having similar relative errors as the data used here).

## 7. Effects of Experimental Design

### 7.1 Choice of labeled carbon sources

In choosing labeled substrates for an experiment, there are two considerations to keep in mind[186]. First, labeled substrates should lead to high total $^{13}C$ enrichment in the cellular system after potential decarboxylation. Second, labeled substrates should result in different mass distributions of isolated metabolites under different flux distributions.

These two qualities are investigated by computing isotopomer distributions of isolated metabolites for a set of 1000 flux distributions. The 1000 flux distributions are calculated using the Hit-and-Run sampling algorithm[94,185]. Twelve commercially available substrate mixtures are studied:

1) 100% [U-$^{13}C_6$] glucose
2) 100% [1 $^{13}C$] glucose
3) 100% [2 $^{13}C$] glucose
4) 100% [4 $^{13}C$] glucose
5) 100% [6 $^{13}C$] glucose
6) 50% [U-$^{13}C_6$] glucose and 50% [1,2 $^{13}C$] glucose
7) 20% [U-$^{13}C_6$] glucose and 80% [1 $^{13}C$] glucose
8) 20% [U-$^{13}C_6$] glucose and 80% [2 $^{13}C$] glucose
9) 20% [U-$^{13}C_6$] glucose and 80% [3 $^{13}C$] glucose
10) 20% [U-$^{13}C_6$] glucose and 80% [4 $^{13}C$] glucose
11) 20% [U-$^{13}C_6$] glucose and 80% [5 $^{13}C$] glucose
12) 20% [U-$^{13}C_6$] glucose and 80% [6 $^{13}C$] glucose

The seven mixtures producing substantially higher total enrichment of the TCA cycle intermediates are 3, 4, 6, 8, 9, 10, and 11. In addition, standard deviations are calculated for values of each mass isotopomer of each metabolite across the 1000 flux distributions. The substrate mixtures producing the largest overall standard deviations are 7, 9 and 10. Considering both qualities, one is recommended to use 20% [U-$^{13}C_6$]

glucose and 80% [3 $^{13}C$] glucose or 20% [U $^{13}C_6$] glucose and 80% [4 $^{13}C$] glucose for studying the TCA cycle dynamics.

An isotopomer model generated under the framework described here is based on the assumption that one can determine the isotopomer distribution of the products of a reaction if the reaction rate and the isotopomer distribution of reactant(s) are known. Therefore, it is essential that the isotopomer distribution of at least one metabolite is always known throughout the time course of the experiment. In practice, this can be done by supplying a tracer that is only taken up and not secreted by the cell at steady state. This way the isotopomer pool for that metabolite does not get "contaminated" by isotopomers produced by the cell. In the study by Khairallah *et al*.[136], even though the perfusate was not recirculated after going through the heart, there is some mixing, in the extracellular space, of labeled pyruvate supplied by the buffer and pyruvate produced by the cell. As a result, the isotopomer distribution of extracellular pyruvate was not known definitively through out the experiment. The same situation occurred in the experiments with labeled lactate. Therefore, of the four experiments by Khairallah *et al*.[136], glucose and oleate are suitable substrates to be analyzed using the method described in this paper, but pyruvate and lactate are not.

## 7.2 Choice of flux or isotopomer measurement

In profiling metabolic fluxes, one can combine mass isotopomer data with flux measurements to accurately determine the intracellular flux distribution in a metabolic

network. As flux measuring experiments tend to be intricate and difficult, it is useful to identify which reaction rates are the most informative for determining the rate of the remaining reactions. The calculated flux distributions yield a set of 28 reactions, whose flux values cannot be precisely determined with the present data. These reactions generally fall into two categories: ketone body and glutamine metabolism. Therefore, reaction fluxes or isotopomers of metabolites in pathways involving ketone bodies and glutamine are good candidates for measurement in future experiments. By iteratively studying the results of previous measurements, each subsequent experiment benefits from the knowledge gained from previous experiments, and together they paint a more complete picture of the metabolic network.

## 8. Computational Considerations

The optimization framework for predicting flux distribution from isotopomer data as described here produces a set of least-squares, best-fit, steady-state flux distributions from a given set of mass distribution data. For common metabolic networks, there is no guarantee of finding the globally optimal solution in polynomial time. As a result, one must sample a sufficiently large group of local solutions, and identify a group of flux distributions that are most likely to be physiological. Alternatively, one can select the solution with the least deviation from experimental data and designate that as the best and most probable flux distribution. The latter approach, however, is likely to over fit the data. Though the process of developing a comprehensive model as presented here is more time consuming than deriving analytical expressions, the benefit of this approach is

that once the network and associated constraints are set up, they can be applied to analyze isotopomer data for various tracers with very little modification. The AMMs and the IMMs are inherently modular; they only have to be constructed once and can be used in any networks that include the associated reactions. In addition, the constraint-based framework ensures that predictions made by the model cannot contradict previously known information about reaction fluxes (those represented by constraints) and therefore the model can serve as a framework to resolve inconsistent data.

In applying this approach for isotopomer analysis, careful consideration must be taken to determine which experimental quantities can be set as constraints and which quantities are used to formulate the objective function. Constraints in the nonlinear programming problem specify mathematical relationships that the network must faithfully obey, while the objective function describes the preferable characteristics of the optimal solutions. Most studies, including this one, have used mass distribution data for the objective function and measurements of substrate uptake and secretion rates as constraints. This practice is usually followed for two reasons. First, setting the mass distribution variables exactly equal to the mean of the observed data often eliminates all feasible steady states. In addition, including the standard deviations as the lower and upper bounds for these variables is cumbersome, and the resulting sum of elements of the isotopomer distribution vectors may not be unity. Second, it is much more straightforward to include flux means and standard deviations as constraints on reaction fluxes. The consequence of these constraints can be quickly determined by solving a linear programming problem that excludes isotopomer balance constraints. Constraints

should be set sufficiently loose to avoid eliminating physiologically relevant flux distributions. Naturally, the decision on constraint formulation should be specific to the system of interest and confidence in the experimental measurements.

(This chapter, in part, is a reprint of the material appearing in "Isotopomer analysis of myocardial substrate metabolism: A systems biology approach," Vo TD and Palsson BO. 2006. *Biotechnology and Bioengineering*, 95(5):972-83. The dissertation author is the primary investigator and author of this paper.)

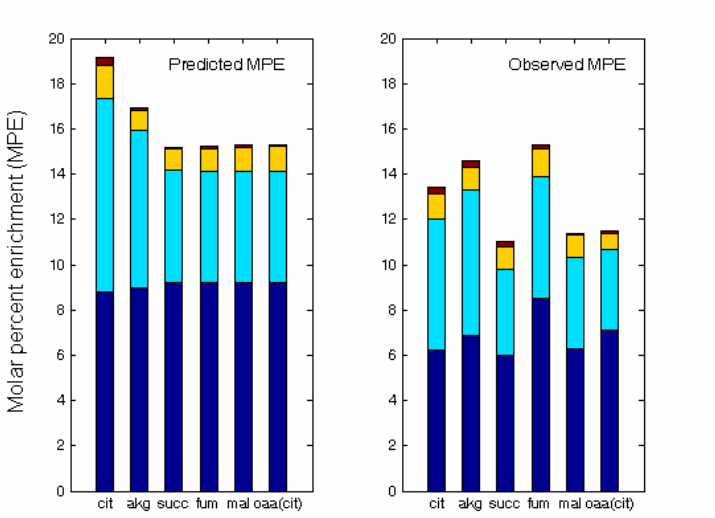
**Table 5.1:Constraints on substrate uptake and efflux**
The constraints for oxygen, pyruvate, lactate, citrate, and succinate are converted directly from flux values reported by Khairallah *et al*.[136] in the experiments with [U-$^{13}C_6$] glucose. The reported unit of μmol/min is interpreted as μmol/min/heart, and is converted to μmol/min/gww using the reported mouse heart wet weight. These constraints are set at two SE around the mean. These values represent only the **net** uptake (negative) or secretion (positive) by the cells. Pyruvate and lactate are allowed to be simultaneously taken up and released by the cells, as observed in the experiment. The positive upper and lower bounds specify a net secretion of these two metabolites. All numbers are derived from Khairallah *et al.* except for oleate which is taken from DeGrella and Light[187] and the lower bound for glucose, which is set arbitrary large.

| Substrates | Lower bound | Upper bound |
|---|---|---|
| Glucose | -10.00 | -1.455 |
| Lactate | 1.33 | 2.17 |
| Pyruvate | 0.125 | 0.625 |
| Oleate | -0.30 | 0.00 |
| Citrate | 0.015 | 0.019 |
| Succinate | 0.005 | 0.017 |
| Oxygen | -7.05 | -5.45 |

**Table 5.2: Fractional contribution of exogenous carbohydrates to cytosolic pyruvate**
Fractional contribution of each exogenous carbohydrate is calculated based on reaction fluxes calculated in the model. Estimated values for Khairallah *et al*.[136] are computed using flux results reported in that paper, with pyruvate uptake = 0.11 ± 0.02 μmol/min, lactate uptake = 0.09 ± 0.03 μmol/min, and glucose uptake rate > 0.51 ± 0.06 μmol/min (sum of lactate and pyruvate efflux when the heart is perfused with [U $^{13}C_6$] glucose). Values are reported as mean ± SD for this study and mean ± SE for Khairallah *et al.*

| | Fractional contribution (%) | |
|---|---|---|
| | This study | Khairallah *et al*. |
| Glucose | 80 ± 2 | > 72 ± 8 |
| Lactate | 8 ± 2 | < 15 ± 4 |
| Pyruvate | 12 ± 2 | < 13 ± 3 |

**A**




**B**

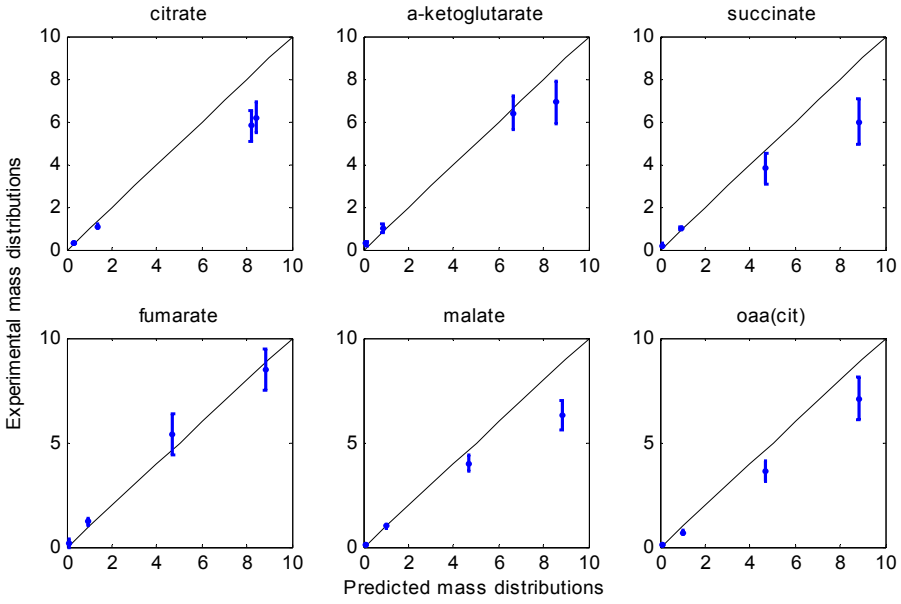
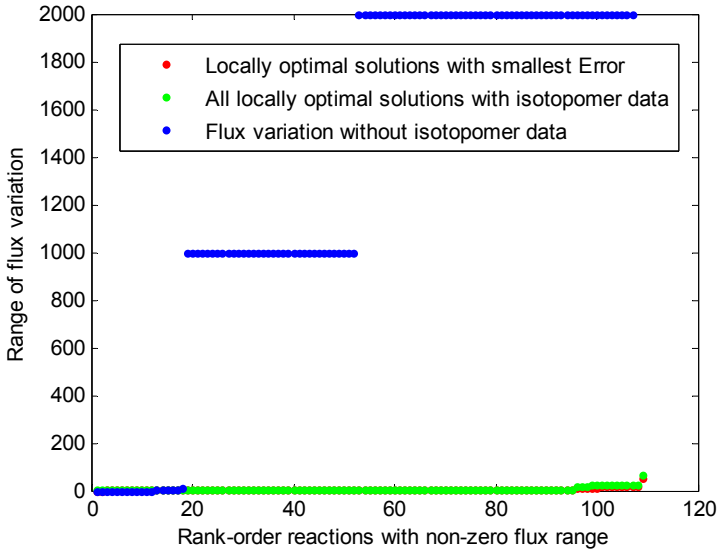



**Figure 5.1: Predicted (calculated) mass distributions for TCA cycle intermediates compared to experimentally measured mass distribution**

In each panel, values from left to right are m4, m3, m2, and m1, reported as average molar percent enrichment. The error bars are SE associated with experimental data reported by Khairallah *et al.* [136]. Predicted mass distributions also have associated standard deviations, but such deviations are very small and are not visible in the figure. Compared to m1 and m2 isotopomers, there is stronger agreement between the predicted and the experimental averages for m3 and m4 isotopomers of the CAC intermediates. The experimental data for m3 and m4 isotopomers have smaller standard errors, and thus the model favors flux distributions that have better fit for these isotopomers. Oaa(cit) refers to the oxaloacetate moiety of citrate.

**A**




**B**

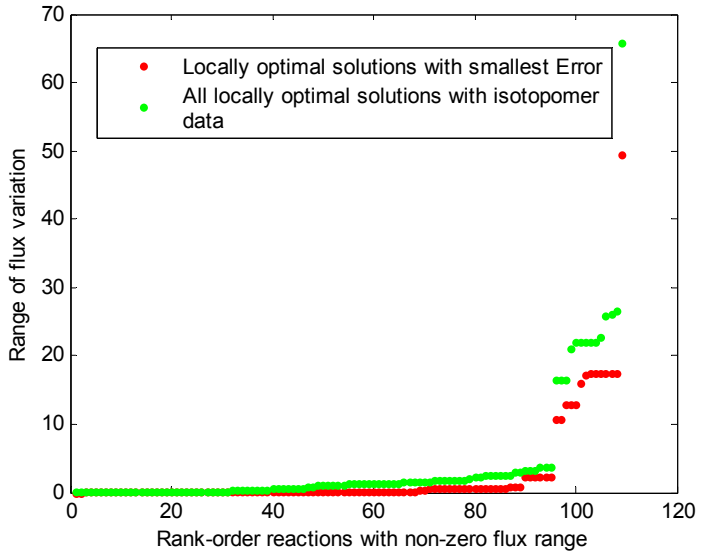



**Figure 5.2: Estimated flux variation from the non-linear model using isotopomer data compared to those from the linear model not using isotopomer data**

"All locally optimal solutions with isotopomer data" refers to all solutions that the SNOPT solver returns with "locally optimal" status. Some of these solutions may have very large error values. "Locally optimal solutions with smallest Error" refers to solutions with objective values no more than 5% of the smallest error found. The precision in flux estimation is at least one order of magnitude better when isotopomer data are used (A). Flux variations have the unit of μmol/min/gww. A total of 149 reactions (not shown) have constant flux values. Flux variation is also reduced if one considers only flux distributions with the smallest errors rather than all locally optimal solutions (B) returned by the solver

# Chapter 6
# Isotopomer Analysis of Cellular Metabolism in Normal and Disease States

As hopefully has been clear in the preceding chapters, experimental studies constitute an essential component of systems biology, without which computational models cannot be built or validated. Among the various data types discussed in this dissertation, staple isotopes as carbon tracers are perhaps the most straightforward to obtain and the most informative in identifying the steady states assumed by the cells. This chapter describes the acquisition of carbon tracer data, i.e. isotopomers and results from their analysis within the constraint-based framework. The chapter is divided into three parts: *i*) experimental studies of HepG2 and dermal fibroblasts in media with labeled substrates, *ii*) analysis of the effects of changed medium composition on internal reaction fluxes of HepG2 cells, and *iii*) characterization of metabolic phenotypes and identification of the affected enzymes in fibroblasts.

## 1. Experimental procedures

### 1.1 Tissue culture

***Experiments with HepG2 cells*:** The human hepatoma HepG2 cell line is obtained from the American Type Culture Collection (ATCC, Manassas, VA), Cat No. HB-8065, and cultured according to published protocols[188]. Briefly, these cells are grown in

DMEM, catalog numbers 11966 (medium A) and 11054 (medium B) (Invitrogen Corp., Carlsbad, CA). The media are supplemented with 10% fetal bovine serum, 1% antibiotic and antimycotic, as well as 25 mM HEPES buffer. Unlabeled and [1,2 $^{13}C_2$] glucose are added to medium A to make a final concentration of 1000 mg/L of each type of glucose. To medium B, [1,2 $^{13}C_2$] glucose is added to make a final concentration of 1000 mg/L (DMEM 11054 already contains unlabeled glucose) and asparagine to make a final concentration of 584 mg/L. The complete contents of the media are shown in Table 6.1. Cells are grown in four flasks of each medium type. When the cells are approximately 50% confluent (~$2.5 \times 10^6$ cells/plate), the media are removed, and the cells are washed with Dulbecco's phosphate-buffered saline (Invitrogen Corp., Carlsbad, CA). Tissue cultures are grown for 72 h, with a medium change every 24 h. The media are saved for lactate and glutamate analysis. After trypsin treatment, cell pellets are collected and kept frozen (-20 $^0$C) until ready for RNA and lipid extraction.

***Experiments with fibroblasts*:** Two fibroblast cell lines, CRL1501 (control) and GM015013 (Leigh's disease) are obtained from ATCC and NIGMS Coriell Institute for Medical Research (Camden, NJ), respectively. Cells are grown in DMEM (Invitrogen Corp., catalog number 11966) supplemented with 10% fetal bovine serum, 1% antibiotic and antimycotic, and 25 mM HEPES buffer[188]. Unlabeled and [1,2 $^{13}C_2$] glucose are added to make a final concentration of 1000 mg/L of each type of glucose. Each cell line is grown in four flasks for 70 h, with a medium change every 24 h. The media are saved for pyruvate, lactate and amino acid analysis. Pyruvate concentrations in media are measured after 2 and 5 h, and lactate and glucose concentrations are measured after 2, 5

and 70 h. Amino acid concentrations in media are determined by comparing their intensities at 70 h with a standard (Sigma, catalog number A6407). Cells are washed with Dulbecco's phosphate-buffered saline (Invitrogen Corp.) and removed from flasks with trypsin. Cell pellets are collected and kept frozen (-20 $^{o}$C) until ready for protein isolation.

## 1.2 Cleanup and derivatization of metabolites

***Lactate*:** A 1-ml aliquot of medium is loaded onto a dowex-1 column and acidified with 2 N acetic acid. The eluting mixture is dried under compressed air. Lactate is derivatized to *n*-propylamide-heptafluorobutyric ester according to the method of Tserng *et al.*[189] for GC-MS analysis.

***Fatty acids***: Lipids are separated from the cell pellet by saponification of triglycerides with 1 ml 30% KOH-ethanol (1:1, vol/vol) at 70°C overnight[190]. Neutral lipids are first removed with petroleum ether extraction. The solution containing the saponified fatty acids is then acidified; palmitate and other fatty acids are recovered with another petroleum ether extraction. Extracted palmitate is dried and then esterified with 0.5 N HCl in methanol (Supelco, Bellfonte, PA) for GC-MS analysis[191].

***Ribose***: RNA is extracted from the cell pellet after treatment with ice-cold 0.5% Tryzol (vol/vol). The purified RNA pellet is hydrolyzed in 2 ml 2 N HCl for 2 h at 100 °C and dried under a flow of nitrogen. Ribose is then derivatized to its aldonitrile acetate form for GC-MS analysis[188].

***Amino acids from cell pellets***: 2 ml of phosphate-buffered saline are added to dissolve a pellet sample of 5-10 million cells. The sample is centrifuged at 2500 rpm for 10 min. The supernatant is removed, followed by an addition of 200 μl of HPLC-grade water. The sample, placed in ice water, is sonicated for 12 min and is transferred to Eppendorf microcentrifuge tubes for 30 min of centrifugation at 10,000 rpm at 4 °C. The supernatant is then pipetted to glass tubes, combined with 300 μl of 12 N HCl and heated over night at 100 $^{o}$C. After being dried under nitrogen, 200 μl $CH_2Cl_2$ and 20 μl of trifluoroacetic anhydride are added to the sample and kept at room temperature for 20 min. After the solvent is dried again with nitrogen, $CH_2Cl_2$ is added to dissolve the sample for GC-MS analysis.

***Amino acids from media*:** 1 ml culture medium is loaded on a dowex-50 cation exchange column to bind positive ions. Columns are washed with 10 ml water to remove neutral and positively charged compounds. The amino acids are eluted with 10 ml 2 N $NH_4OH$ and dried under compressed air. The derivatization procedure is the same as that for amino acids from cell pellets.

***Glutamate from media***: The procedure for isolating glutamate from the media is the same as that described for isolating amino acids from the media. To further separate glutamine from glutamate, samples are dissolved in water and loaded on dowex-1 anion exchange columns. Glutamine is removed with 15 ml water. Glutamate is eluted with 0.5 N acetic acid and dried under compressed air. The dried sample is combined with

200 μl butanolic HCl and heated for at least 1 h at 100 °C. The sample is dried under nitrogen. The derivatization procedure is the same as that for amino acids from cell pellets. The resulting derivative of glutamate is n-trifluoroacetyl-n-butyl-ester.

### 1.3 GC-MS analysis

GC-MS analysis is carried out with a Hewlett Packard Gas Chromatograph (6890 series) connected to a 5973 HP Mass Selective Detector. GC conditions are as follows: injector temperature 250 $^{o}$C and oven temperature 210-220 $^{o}$C at 10 $^{o}$C/min[192]. For lactate, ribose, and amino acid analysis, an Agilent HP-5 capillary column (30m × 250 μm) is used with He as carrier gas at a constant flow rate of 1 ml/min. For palmitate, a SGE bpx70 column (30m × 250 μM) is used, also with a constant He flow of 1 ml/min. Electron impact ionization is used to identify labeled carbons in mass fragments of palmitate (m/z 270), alanine (m/z 140-142), glycine (m/z 126-127), threonine (m/z 153-156), valine (m/z 168-172), leucine (m/z 182-187), isoleucine (m/z 182-187), proline (m/z 166-170), methionine (m/z 153-156), aspartate (m/z 184-187), phenylalanine (m/z 91-98), tyrosine (m/z 203-210), and glutamate (m/z 198-202)[193]. Lactate fragments (m/z 328) and ribose $C_1$-$C_5$ fragment (m/z 256) are monitored under chemical ionization with 20% methane as the reagent gas. Each biological sample is analyzed three times in the GC-MS machine; selected ions are recorded and normalized in abundance.

## 2. Pathway-based analysis

HepG2, derived from a liver carcinoma, is an adherent cell line that has been widely used as a model for hepatocellular carcinoma[48-50]. A variety of Dulbecco's modified Eagle's media (DMEM) have been used to propagate HepG2 cells in tissue cultures[194-196], although it is not clear how closely these media can mimic the physiological extracellular environment or what effects different varieties of DMEM have on the cell's metabolism. Here HepG2 cells are cultured in two DMEM media, which are different in their amino acid and carbohydrate compositions (Table 6.1). Using [1,2 $^{13}C_2$] glucose as the tracer, GC-MS methods are applied to determine the isotopomer mass distributions of glutamate, ribose, lactate, and palmitate. As mentioned, both the pathway-based and network-based methods are applied to calculate reaction fluxes using known HepG2 biochemistry and measured isotopomers. With the pathway-based method, the analysis is limited to 12 reactions involved in the TCA cycle, pentose phosphate cycle (PPC), and palmitate synthesis. On the other hand, a more extensive model of HepG2 metabolism involving 254 reactions and 269 metabolites is used in the network-based analysis.

## 2.1 Anaplerotic and TCA cycle fluxes inferred from glutamate isotopomers

Supplied [1,2 $^{13}C_2$] glucose is metabolized to [2,3 $^{13}C_2$] pyruvate, which, in turn, is primarily either carboxylated to [2,3 $^{13}C_2$] oxaloacetate or decarboxylated to [1,2 $^{13}C_2$] acetyl-CoA. In the first turn of the TCA cycle, these oxaloacetate and acetyl-CoA form [2,3 $^{13}C_2$] citrate (oxaloacetate moiety) or [4,5 $^{13}C_2$] citrate (acetyl-CoA moiety). Resulting α-ketoglutarate molecules are either labeled in $C_2$ and $C_3$ or $C_4$ and $C_5$, both of

which yield m2 glutamate (Figure 6.1). Subsequent turns of the TCA cycle produce m1 α-ketoglutarate, and thus m1 glutamate. Justification and detailed description of these assumptions have been published in the literature[139]. The ratios m1/Σm and m2/Σm, where Σm is defined by Equation 6.1, reflect the activity of the TCA cycle. The different $m_i$ in this equation correspond to the different fractional mass isotopomers of glutamate:

$$\Sigma m = m1 + m2 + m3 + m4 \qquad \text{(Equation 6.1)}$$

If it is further assumed that the two primary fates of oxaloacetate are being converted to citrate and exiting in anaplerosis as phosphoenol pyruvate, then the corresponding fraction of oxaloacetate in each path can be referred to as *r* and (1-*r*), respectively. The effective anaplerosis is simply (1-*r*)/*r*, where the value for *r* can be inferred from the ratio of m1/m2 glutamate (Equation 6.2)[139,197]:

$$\frac{m1}{m2} = \frac{r}{1 - 0.5r} \qquad \text{(Equation 6.2)}$$

## 2.2 Ribose isotopomers and the pentose phosphate cycle fluxes

Riboses isolated from RNA are mostly derived from ribose-5-phosphate (R5P). R5P can either come from the isomerization of ribulose-5-phosphate (produced by 6-phosphogluconate dehydrogenase) or from sedoheptulose-7-phosphate and glycer-aldehyde-3-phosphate (G3P) carbon rearrangement. The former pathway can form m1 and m3 R5P from [1,2 $^{13}C_2$] glucose through the action of 6-phosphogluconate dehydrogenase and transketolase. The latter path can produce either m2 or m4 R5P: m2 R5P can be derived from [1,2 $^{13}C_2$] fructose or [2,3-$^{13}C_2$] G3P, and m4 R5P is produced

when both of these molecules come together to form [1,2,4,5-$^{13}C_4$] xylulose-5-phosphate, which isomerizes to R5P[188]. Therefore, the ratio of the oxidative to non-oxidative branch of the pentose phosphate cycle can be inferred from Equation 6.3:

$$\frac{Oxidative}{Nonoxidative} = \frac{m1 + m3}{m2 + 2 * m4} \qquad \text{(Equation 6.3)}$$

When [1,2 $^{13}C_2$] glucose is used as the tracer, the enzyme 6-phosphogluconate dehydrogenase produces [1 $^{13}C$] R5P, which can be recycled to single $C_1$, $C_2$, or $C_3$ hexose phosphate. These hexoses are eventually converted to [3 $^{13}C$], [2 $^{13}C$], or [1 $^{13}C$] lactate (m1). On the other hand, lactate produced from glycolysis is primarily [2,3 $^{13}C_2$] lactate (m2). Hence, PPC flux can be estimated based on the enrichment in m1 and m2 lactate (Equation 6.4)[188,198]:

$$\text{PPC flux} = \frac{m1}{m1 + 3 * m2} \qquad \text{(Equation 6.4)}$$

## 2.3 Fractional synthesis rate of palmitate

A fraction of acetyl-CoA produced from glucose oxidation is converted to palmitate. In calculating the fraction of palmitate newly synthesized from acetyl-CoA, three assumptions are made: *i*) newly synthesized palmitate is derived entirely from acetyl-CoA; *ii*) this acetyl-CoA pool is made up of m0 and m2 mass isotopomers (m2 acetyl-CoA originates from [1,2 $^{13}C_2$] glucose); *iii*) the fractional enrichment of a particular palmitate isotopomer ($m_k$) can be estimated from the binomial distribution describing eight draws of acetyl-CoA from the mentioned acetyl-CoA pool. The

rationale for these assumptions has been described in the literature[199,200]. Thus, the expected fractional enrichment of palmitate $m_k$ can be calculated with Equation 6.5:

$$m_k = \binom{N}{i} p^i q^{N-i} \qquad \text{(Equation 6.5)}$$

where N is the number of acetyl-CoA per palmitate molecule (eight), *i* is the number of m2 acetyl-CoA incorporated into palmitate, *p* is the fractional enrichment of m2 acetyl-CoA, and *q* is the fractional enrichment of m0 acetyl-CoA. The sum of *p* and *q* is one, based on assumption 2. The expected value for *p* can be computed by taking the ratio of two consecutive mass isotopomers (Equation 6.6),

$$\frac{m_{k+1}}{m_k} = \frac{\binom{N}{j} p^j (1-p)^{N-j}}{\binom{N}{i} p^i (1-p)^{N-i}} = \frac{(N-i)!i!}{(N-j)!j!} p^{j-i} (1-p)^{i-j} \qquad \text{(Equation 6.6)}$$

if $m_k$ represents the fractional enrichment of m2 palmitate (thus $i$=1) and $m_{k+1}$ represents the fractional enrichment of m4 palmitate (thus $j$=2). Equation 6.6 can be simplified to Equation 6.7 to evaluate *p* based on the observed m2 and m4 palmitate:

$$\frac{m4}{m2} = \frac{7}{2} \frac{p}{(1-p)} \qquad \text{(Equation 6.7)}$$

Fractional synthesis rate (FSR) of palmitate derived from [1,2 $^{13}C_2$] acetyl-CoA can be estimated based on the ratio of observed $m_k$ palmitate to the expected $m_k$ if all palmitate molecules are synthesized from the acetyl-CoA pool mentioned above (Equation 6.8)[191].

$$FSR = \frac{m2_{observed}}{\binom{8}{1} p^1 (1-p)^7} \qquad \text{(Equation 6.8)}$$

## 3. Network-based analysis

The network-based analysis requires the use of a reconstructed metabolic network and constraints on reaction fluxes as well as metabolites to identify the set of flux distributions best matching observed isotopomer data. Contents and procedures for constructing the HepG2 and fibroblast metabolic networks are described in Chapter 2. Relevant constraints as well as methods for integrating isotopomer data for reaction flux determination are provided in Chapter 4.

## 4. Effects of altered substrate availability on HepG2 metabolism

Composition of culture media is critical for proper cellular function, growth, and survival. Commercially available media routinely used for established cell lines often include serum, glucose, glutamine, and a few other amino acids. A carbon tracing experiment with HepG2 cells cultured in two different DMEM media is conducted to explore the response of HepG2 cells to variations in medium nutritional composition. These two media (A and B, Table 6.1) are slight modifications of the media obtained from Invitrogen. [1,2 $^{13}C_2$] glucose is chosen as this tracer leads to the same primary and secondary oxaloacetate isotopomers (Figure 6.1), thereby avoiding the complication in isotopomer analysis encountered with [U-$^{13}C_6$] glucose[201]. Secondary oxaloacetate isotopomers are produced through the cycling of pyruvate: pyruvate $\rightarrow$ oxaloacetate $\rightarrow$

PEP → pyruvate. The advantage of [1,2 $^{13}C_2$] glucose over other singly-labeled glucoses in studying the dynamic changes of PPC has also been discussed[188]. Over the course of 72 h, glutamate, ribose, lactate, and palmitate from the cultures are isolated and analyzed with GC-MS methods to investigate any changes in cellular metabolism. Results of this study are discussed in two parts: *i*) changes in cellular energy metabolism as inferred from the pathway-based method (Section 4) and *ii*) additional insights gained from the network-based method about the intracellular flux distribution and nutritional needs of HepG2 cells (Section 5).

## 4.1 Effects of glutamine and asparagine on the TCA cycle

Over the course of 72 h, HepG2 cells grow equally well in both A and B media. Glutamate isolated from cells grown in glutamine-containing culture (medium A) shows a lower relative abundance of labeled glutamate compared to those grown in asparagine-containing culture (medium B) in all 24, 48, and 72 h samples (Table 6.2). This lower enrichment indicates a dilution of cellular glutamate with glutamate produced by glutamine from the culture medium. Since it is assumed that m1 and m2 glutamate make up the major glutamate isotopomers formed from [1,2 $^{13}C_2$] glucose, the relative abundance of these two isotopomers (m1/Σm and m2/Σm, Table 6.2) reflects the flow of glucose carbon through reactions in the TCA cycle[139]. Results show that HepG2 cells grown in glutamine-containing media produce a higher amount of m2 glutamate, but lower m1 glutamate than those grown in the other media. This suggests that [1,2 $^{13}C_2$]

glucose metabolism by the TCA cycle in cultures A and B is not exactly the same, despite the similar observed growth.

Intermediates of the TCA cycle are continually consumed and produced via processes of cataplerosis and anaplerosis. Anaplerosis conserves and refills the TCA cycle intermediate pool which is depleted by biosynthesis processes[202]. Three-carbon anaplerosis at oxaloacetate is maintained by the actions of phosphoenolpyruvate carboxykinase (PEPCK) and CS. Effective anaplerosis is thus defined as the ratio of these two paths representing the ratio of anabolic activity to energy production. Analysis of m1 and m2 glutamate isotopomers indicates that the presence of pyruvate and asparagine in medium B leads to a 30% decrease of effective anaplerosis in HepG2 cells. This suggests that either the flux at CS is increased or that PEPCK flux is decreased with the presence of pyruvate. The former hypothesis is more likely as it is consistent with the presence of pyruvate and the increased m1 glutamate observed above.

## 4.2 Activity of the pentose phosphate cycle (PPC)

In addition to glycolysis, glucose is also metabolized by the PPC to form ribose-5-phosphate. The PPC has two main branches, oxidative and non-oxidative. Fluxes through these two branches are modulated by cellular needs for ribose in nucleic acid synthesis, NADPH in lipogenesis, or both. The oxidative branch produces mostly m1 and m3 riboses, while non-oxidative enzymes produce mostly m2 and m4 riboses. Overall, HepG2 cells grown in the two media exhibit only a small difference in the total

enrichment of ribose from glucose labels; such difference is attributable to random measurement error (Figure 6.2). Accordingly, the ratios of ribose produced by oxidative to non-oxidative branches are essentially the same for HepG2 cells cultured in media A and B. In addition, PPC flux, relative to glycolysis, can also be estimated using isotope yields from triose products[188,198]. Calculations (Equation 6.4) show that HepG2 cells grown in medium A do not have a significant change in relative PPC flux compared to those in medium B (6.5 ± 0.4% and 6.1 ± 0.5%, respectively). The comparable growth observed for cells in both media is also indicative of this similar ribose synthesis rate.

## 4.3 Palmitate synthesis

Palmitate isolated from culture media is used to assess *de novo* lipogenesis as it has been reported that cultured HepG2 cells preferentially and rapidly secrete *de novo* synthesized fatty acids rather than storing them in the cytosol[203]. *De novo* synthesis of cellular palmitate comes entirely from the intracellular acetyl-CoA pool, which itself is mostly produced from oxidation of supplied [1,2 $^{13}C_2$] glucose, unlabeled glucose, and fatty acids. As in previous work, the synthesis of palmitate is modeled by a binomial distribution, with *p* being the probability of drawing a m2 acetyl-CoA and *q* being the probability of drawing a m0 acetyl-CoA from the intracellular acetyl-CoA pool[198,199]. Therefore, the values *p* and *q* represent the fractional enrichment of m2 and m0 acetyl-CoA, respectively. This analysis shows that acetyl-CoA enrichment from [1,2 $^{13}C_2$] glucose is not significantly affected by the composition difference between media A and B (Figure 6.3, Table 6.3). *De novo* palmitate synthesis of cells in culture A and B is

estimated to contribute 47.7% and 34.6%, respectively, to the total intracellular palmitate pool. The estimate for cells in medium B is remarkably similar to the reported 33 ± 2% hepatic *de novo* lipogenesis based on intravenous infusion of stable isotope tracers in healthy volunteers[204]. Since palmitate had the highest synthesis rate among non-essential fatty acids in that study, it is likely that the palmitate synthesis rate estimated here also represents the majority of *de novo* lipogenesis activity in HepG2 cells.

# 5. Insights from the network-based analysis

## 5.1 Computational model used in the network-based analysis

A computational model (Section 3) incorporating the major metabolic pathways in HepG2 cells and measured isotopomer data is used to estimate reaction fluxes in these pathways. Such flux distributions elucidated the mechanism by which cells produce the observed $^{13}C$ isotopomers. Metabolism of HepG2 cells is represented by a biochemical network describing glycolysis, the TCA cycle, the PPC, oxidative phosphorylation, amino acid metabolism, β-oxidation, porphyrin metabolism, the urea cycle, palmitate biosynthesis, and phospholipid synthesis. These pathways make up a total of 254 reactions and 269 metabolites. The model also keeps track of isotopomers of 104 of these metabolites. This is thus the largest metabolic network ever used for isotopomer analysis. In comparison, some of the largest previous models include *Saccharomyces cerevisiae* (54 reactions[155] and 37 reactions[156]), *Corynebacterium glutamicum* (47 reactions[160]), and *Escherichia coli* (32 reactions[152]).

## 5.2 Effects of extracellular environment on intracellular fluxes

A fundamental distinction between the network-based approach and the previous pathway-based method is the presence of a well defined system enclosed by a theoretical boundary[145,179]. This boundary, defined here to be just outside the plasma membrane, differentiates substrates that must be taken up (input) from by-products that must be secreted by the cell (output). In addition, fluxes in exchange with the environment are now distinguishable from internal reaction fluxes, so changes in intracellular flux distribution can be understood in terms of input/output behaviors of the cellular system.

It is well documented that cultured cells utilize both glucose and glutamine as energy sources and precursors for other biosynthetic products[205,206]. When glutamine (in medium A) is replaced with asparagine (in medium B), the cells must be able to substitute their requirement for the former amino acid, at least partly, by asparagine. It has also been reported that while hepatocytes use the same transport systems, systems A and N[207,208], for both glutamine and asparagine, the rate of glutamine uptake is substantially higher than that of asparagine[209]. Results from this study indeed confirm this relationship. Glutamine is estimated to be taken up at the rate of 9.24 nmol/min/mg protein by HepG2 cells in medium A, while those in medium B take up asparagine at a much slower rate, 3.78 nmol/min/mg protein (Figure 6.4). On the other hand, the estimated glucose uptake rates for these two cultures are similar, 1.94 and 1.85 nmol/min/mg protein, in media A and B, respectively. Interestingly, cultures A secrete a net amount of pyruvate, 2.38 nmol/min/mg protein, while cultures B take in a net amount of this compound, 10.34 nmol/min/mg protein. The conversion of pyruvate to lactate, i.e.

activity of lactate dehydrogenase, in cells cultured in medium B is less than half of that in cells cultured in medium A (Figure 6.4). The amount of lactate secreted is similarly decreased, from 1.96 to 0.68 nmol/min/mg protein in cultures A and B, respectively.

Due to the differences in the consumed carbon sources, cells grown in medium A primarily have glutamine entering the TCA cycle as α-ketoglutarate, while those in medium B have asparagine and pyruvate entering the cycle as oxaloacetate and acetyl-CoA. Fluxes through enzymes in the TCA cycle reflect these dynamics: cells in medium A have substantially higher α-ketoglutarate dehydrogenase activity, while the other cell cultures have higher pyruvate dehydrogenase and CS activities (Figure 6.4). As a result, not all enzymes in the TCA cycle functioned at the same steady state reaction rates. Based on the observed glutamate isotopomers and the pathway-based analysis, it was previously suggested, from the pathway-based analysis, that cells grown in medium B have different TCA cycle activity; here the model shows that only isocitrate dehydrogenase and upstream enzymes had higher fluxes. On the other hand, malate dehydrogenase has a higher flux in culture A. In short, the majority of carbon sources enter the TCA cycle at α-ketoglutarate dehydrogenase in cells cultured with medium A, but enter at CS, both by oxaloacetate (product of asparagine) and acetyl-CoA (product of glucose and pyruvate), in those cultured with medium B (Figure 6.4). It is concluded that the observed disparity in glutamate isotopomers as well as the estimated reaction fluxes is a result of the difference in substrate entry points.

## 5.3 End products of cells grown in glutamine and asparagine-containing media

The model predicts that HepG2 cells grown in glutamine-containing media (medium A) secrete various metabolites into the media including lactate (1.96 nmol/min/mg protein), alanine (3.13), proline (0.87), serine (0.15), citrate (0.85), and ketone bodies (0.9). These secretion rates are all within the range reported for thirteen other mammalian cell lines[210]. Of these metabolites, cells grown in medium B are predicted to secrete alanine, serine, and ketone bodies. Perhaps most interesting, cells grown in glutamine-containing media secrete asparagine, 0.83 nmol/min/mg protein, and those in asparagine-containing media secrete glutamine, 1.09 nmol/min/mg protein. Gluconeogenesis is represented in the network as a demand on glucose-6-phosphate, which is exported to the endoplasmic reticulum. The model predicts comparable glucose-6-phosphate export for both cultures: 0.9 and 1.1 nmol/min/mg protein for cells in media A and B, respectively.

Cells use a fraction of acetyl-CoA produced by glucose, pyruvate, and fatty acids to synthesize palmitate. The model predicts a much smaller enrichment of m2 acetyl-CoA (~1%) in both cultures than the value of $p$ calculated from the pathway method (Figure 6.3). The steady state rate of palmitate synthesis is predicted to be 0.06 and 0.04 nmol/min/mg protein for HepG2 grown in medium A and B, respectively. Similarly, the rate of phospholipid synthesis in culture A (0.16 nmol/min/mg protein) is higher than that in culture B cells (0.10 nmol/min/mg protein). On the other hand, the collective β-oxidation of fatty acids in the media (palmitate, stearate, oleate, octadecynoate,

arachidonate, and docosahexaenoate) is 50% lower in the former cultures (0.89 nmol/min/mg protein) compared to the latter (1.29 nmol/min/mg protein). These fatty acids produced acetyl-CoA at the rates of 5.25 and 7.75 nmol/min/mg protein, respectively, making them a major energy source in the cell. The generation of ketone bodies is also indicative of this high fatty acid oxidation. Collectively, the flux in lipid metabolism pointed to the same conclusion that asparagine (medium B) supplied at the same amount as glutamine (medium A) in the media is not able to sustain equivalent cellular metabolism. Cells grown in medium B consequently must take up more of other substrates and produce fewer end products into the media.

Though asparagine and glutamine have the same number of nitrogen atoms per molecule, cells grown with glutamine (medium A) have a higher steady state nitrogen load as they take up glutamine at a faster rate. This effectively increases the amount of ammonia that has to be converted to urea. It is thus not surprising that enzymes in the urea cycle have much higher steady state rates than those in asparagine cells; urea secretion in medium A is 6.14 nmol/min/mg protein and is 1.63 nmol/min/mg protein in medium B (Figure 6.4). Since enzymes in the urea cycle require energy, the energy load in culture A is also likely to be higher. These results suggest that one can improve the growth rate and reduce energy load of cultured HepG2 cells by substituting asparagine for glutamine at a sufficiently high concentration to overcome the slower asparagine transport rate.

## 6. Fibroblasts as a model for Leigh's syndrome

Leigh's syndrome, first described this disease as subacute necrotizing encephalomyelopathy in an infant[211] is a progressive neurodegenerative disorder exhibiting considerably variable clinical signs, symptoms, onset time, and course. Symptoms of LS may include mental retardation, abnormal breathing rhythms, optic atrophy, ataxia, and dystonia[212]. Standard diagnostic tests for LS patients include MRI, lactate and pyruvate levels from plasma and/or cerebro-spinal fluid analysis, and mitochondrial morphology[213]. Studies have associated LS with mutations in complexes of the respiratory chain (commonly in I and IV, and less frequently in II and III), ATPase6, and components of the pyruvate dehydrogenase complex (PDHC)[214] (Table 6.4). However, the low correlation in clinical results with symptoms and inheritance pattern (autosomal dominance, X-linked, and maternal inheritance) of this syndrome indicates that such assays are insufficient to distinguish the different underlying disorders of LS. Therefore, methods to comprehensively elucidate metabolic responses of Leigh-affected cells, thereby identifying affected enzymes or energetic processes, are needed before appropriate treatments can be developed for the disease.

Metabolic profiling is becoming a recognized platform for disease diagnostics[215]. Current MS and NMR-based technologies are capable of identification and quantification of small molecules, both of which are important for metabolic phenotyping of a cell. Tracer-based metabolomics is a branch of metabolomics that focuses on metabolite quantification using isotopes as labeled tracers[137,145]. Relative intensities of different isotopomers of a metabolite reflect the internal flux distribution, hence metabolic phenotype and functional state, of the cell. The large-scale determination of intracellular

fluxes based on isotopomer data requires a computational model that can accommodate flux and carbon tracking for the several hundred reactions making up the major biochemical pathways in a human cell. A constraint-based model, which allows the calculation of reaction fluxes based on network stoichiometry, substrate uptake and secretion rates, and mass balance on $^{13}C$ labels, is the systems biology solution for this problem[216,217]. This systems biology approach is preferred over the traditional pathway-based approach because the resulting flux estimates can be obtained in absolute units, as opposed to flux ratios, and are mutually consistent among the several hundred reactions in the network[51]. Recently, constraint-based models have been successfully used to determine reaction fluxes for *E. coli* [216], cardiomyocytes[217], and hepatocytes[51].

Fibroblasts are chosen as the model system for metabolic profiling of LS metabolism as they are easily and routinely obtained from patients. In addition, fibroblasts are one of the few cell types that have the ability to grow in tissue culture without being transformed. The present studies aim to elucidate substrate utilization, energy production, and deficient enzyme complex(es) in a LS affected cell line through systems analysis of isotopomer data obtained from a $^{13}C$ tracer study.

## 7. Leigh's cells have slower substrate utilization than control cells

The fibroblast metabolic network is used to analyze isotopomer data, thereby characterizing metabolic phenotypes of normal and LS fibroblasts. The slower metabolism of LS affected cells is evidenced by their lower overall substrate utilization

and secretion of end products (Figure 6.5). Based on substrate uptake and secretion rates, the model predicts a respiration rate of 202 ± 50 nmol/hr/mil cells for control fibroblasts and 155 ± 8 for the Leigh's cell line. The normal oxygen consumption of cultured fibroblasts, measured from 158 samples, has the range of 120-240 nmol/hr/mil cells[218]. Glucose taken up by LS affected cells is approximately 35% of the control (1.89 nmol/hr/mil cells vs. 5.31 nmol/hr/mil cells). Pyruvate and lactate secretions in LS cells (0.70 and 2.40 nmol/hr/mil cells) are also lower accordingly (control: 0.96 and 5.46 nmol/hr/mil cells, respectively). Relative to the glucose uptake rate, pyruvate and lactate secretions are higher in LS cells, indicating higher anaerobic respiration in this cell line.

Uptake and secretion rates of the majority of amino acids from and into the media are also slower in affected cells as compared to the controls. Those with the largest differences are alanine secretion (normal vs. LS: 50.5 vs. 15.9 nmol/hr/mil cells), glutamine uptake (16.7 vs. 5.42 nmol/hr/mil cells), histidine uptake (17.7 vs. 7.67 nmol/hr/mil cells), and tyrosine uptake (9.47 vs. 0.74 nmol/hr/mil cells). In contrast, methionine uptake is much higher in LS affected cells (9.34 vs. 13.7 nmol/hr/mil cells) and serine uptake is slightly higher (19.5 vs. 19.7 nmol/hr/mil cells) (Figure 6.6). Predicted flux distributions show that methionine is used for creatine production and serine for succinyl-CoA production. The collective free fatty acid (palmitate, stearate, oleate, octadecynoate, arachidonate, and docosahexaenoate) consumption by LS affected cells is 0.36 nmol/hr/mil cells and of normal fibroblast is 0.46.

Observed growths are comparable between patient and control cell lines

(approximately 0.02 $hr^{-1}$). Interestingly, this growth rate is much slower than what is theoretically permitted based on standard flux balance analysis (linear optimization) [22]. In fact, the model shows that both cell lines are capable of producing ATP at a rate that is an order of magnitude higher than the ATP requirement associated with the biomass function. The biomass function used in the present model is a slight modification of that published for the mouse model[219], as comparable data are not available for human fibroblasts. The lower growth-associated ATP requirement relative to the cell's ATP production capability suggests that either much of the consumed substrates are converted to storage forms or that there are many other energy consumption processes not accounted for in the biomass function. Another implication is that growth prediction made from similar flux balance analyses will vastly overestimate the growth rate of fibroblasts in culture or that substrate consumption predicted based on observed growth using the same procedure will be grossly underestimated.

## 8. Validation of predictions for intracellular reaction fluxes

Flux results described in the following sections are derived from analysis of the 508 reactions in the network, and thus are mutually consistent assuming that the reconstructed biochemical network is correct. Consequently, the validation (through comparison with published results from other studies) of a subset of reaction fluxes in the model also serves as indirect validation for estimated fluxes of the remaining reactions, as all of these reactions are connected through a cohesive model representing the underlying biochemical network.

Flux values predicted by the model are compared to results of *in vitro* enzymatic activity measurements. Except for PDHC whose activity has been reported for the GM01503 (Leigh's) cell line, the remaining comparisons are taken from various cultured fibroblasts that were used as controls in the literature. In particular, the basal activities of PDHC in normal and LS cells were reported to be 1.0-1.5 nmol/min/mg protein[220,221], which is approximately 30-45 nmol/hr/mil cells if one assumes an average of 1 mg cell protein per 2 mil cells[222,223] (Table 6.5). Predicted fluxes for PDHC in control and LS affected fibroblasts in this study are 30.7 ± 2.3 and 31.5 ± 3.6 nmol/hr/mil cells, respectively (Figure 6.6). An enzyme commonly used as an indicator for the TCA cycle activity is citrate synthase (CS). However, reported *in vitro* activities for CS (20-32 nmol/min/mg protein[224] and 53 nmol/min/mg[225]) are significantly higher than that of aconitase (ACONT, 0.4-1 nmol/min/mg protein) and isocitrate dehydrogenase (ICDH, 1-3 nmol/min/mg protein)[224]. Table 1 shows the rates of these enzymes in the flux unit of the model. The model predicts similar fluxes for all three enzymes, ~29 ± 19 nmol/hr/mil cells. This similarity is a result of the steady state constraint which ensures that the majority of products produced by one enzyme are consumed by the next enzyme.

Activity for mitochondrial malic enzyme was reported to be 1.21 ± 0.9 nmol/min/mg protein, which is equivalent to 9-63 nmol/hr/mil cells[221]. The model's predicted flux for this enzyme is 34.6 ± 10.8 nmol/hr/mil cells. Lastly, activity for lactate dehydrogenase has been published to be as high as 699 ± 91 nmol/min/mg protein, equivalent to 9.5-23.7 x $10^3$ nmol/hr/mil cells[221]. This high rate suggests that the enzyme

is highly reversible and sensitive to changes in lactate and pyruvate levels. The model does not predict such a high rate of activity for this enzyme, but it is sufficiently high that intracellular lactate and pyruvate are isotopically equivalent.

# 9. Mutations are predicted to be present in complex II of the respiratory chain

Currently Leigh's syndrome is diagnosed for all patients who have an early-onset progressive neurodegenerative disease that results in cellular necrosis. Thus all patients diagnosed with LS may not have the same underlying disorder[220]. Techniques for characterizing cellular defects at higher resolution are necessary before appropriate treatments can be developed. It is shown in this study that carbon tracer data in combination with a constraint-based model can be used to elucidate the metabolism of the diseased cells, thereby identifying the most likely affected enzyme.

## 9.1 Ruling out PDHC

GM01503 is believed not to contain a deficient PDHC. There are four pieces of evidences supporting this conclusion. First, the model does not predict reduced PDHC activity for LS affected cells as compared to that in the controls. Direct PDHC biochemical assays reported a normal activity range of 30-45 nmol/hr/mil cells[220] which encompasses the predicted PDHC flux values for both fibroblast cell lines used in this study. Second, the lactate levels found in two media (control: 1.20 ± 0.08 mM and Leigh: 1.23 ± 0.01 mM) are within the normal range (< 2.25 mM)[226]. Third, the ratios of

lactate to pyruvate are very similar between the controls (15.6 ± 2.5) and LS cells (11.4 ± 0.4), but are markedly different from that of the known PDH deficient cells (GM03093) also measured in this study (35.0 ± 4.3). The normal range of this ratio is 11.8-24 [214,227]. Lastly, two independent studies have also concluded that this particular Leigh cell line does not have deficient PDHC enzymatic activity. Sorbi *et al.* measured the basal and activated PDH activity and showed that GM01503 had the same basal activity as the normal fibroblasts when (indirect) activators of PDH were not present[220]. Huh *et al.* subsequently reported no abnormality in both subunits of PDH as measured by total enzyme activity, mRNA amounts (Northern blot), and immunoreactive proteins[228].

## 9.2 Ruling out ATP synthase and complexes I, III, and IV

Results from the model also eliminate ATP synthase (complex V) as the deficient complex in the present LS cell line (GM01503). Reasons to support this conclusion are as follows. First, the predicted rate of ATP production by LS cells is very similar to that of the controls. Specifically, the rate of complex V in LS cells is within one standard deviation of that in the controls (686 ± 107 nmol/hr/mil cells), but with a lower mean (601 ± 2). It is important to keep in mind that the lower mean is possibly a result of the lower substrate uptake rates by LS cells rather than a limitation of ATP synthase itself. Secondly, the calculated flux ratio for ATP synthase relative to each of the remaining complexes is essentially the same between LS affected cells and the controls (Table 6.6). This suggests that ATP synthase is not the limiting step in the respiratory chain, and hence is not the complex with the deficiency in this particular patient. As a side note,

mutations causing complex V deficiency in LS have been located only on mtDNA genes (inherited maternally), so patients with these defects can also be confirmed with pedigree study[229-231].

What remains from the list of possible causes for LS are complexes I-IV. These enzymes convert NADH and $FADH_2$ produced by the TCA cycle to the proton-motive force used to generate ATP. Reasons supporting a defect in complexes I-IV are discussed below.

First, LS-affected cells generate NADPH to supplement NADH. Absolute flux values for oxidative PPC (producing NADPH) are higher in LS-affected cells compared to the controls (0.54 nmol/hr/mil cells in LS and 0.68 in controls). Relative to normal fibroblasts, the ratio of oxidative PPC/glycolysis in LS affected cells favors the oxidative branch of the PPC more than the usual glucose oxidation by glycolysis (PPC/glycolysis = 0.6 in LS and 0.1 in controls). This result is also reflected in lactate isotopomers (products of [1,2 $^{13}C$] glucose), which are predominantly labeled in only one carbon. High activity in the oxidative branch of the PPC is indicative of the high NADPH requirement by the cells[232]. NADPH is used as an alternative cofactor for NADH in many enzymes (glutamate dehydrogenase, ICDH, malic enzyme, etc) in the model network.

Second, cofactor consumption and regeneration by complexes I-IV are slower in LS cells. The overall activity of the malate-aspartate shuttle, whose function is to transfer

electrons from cytosolic NADH to mitochondrial NADH, is lower in LS-affected cells than in the controls. The actual net activity of this shuttle is somewhat difficult to assess because reactions in the shuttle also participate in other biochemical processes. However, the model predicts lower fluxes for all six reactions of the shuttle in LS fibroblasts. Furthermore, despite the first three reactions in the TCA cycle (CS, ACONT, ICDH) having similar fluxes in both cell lines, the three reactions that require $NAD^+$ and FAD (regenerated by complexes I-IV) to act as electron acceptors operate at slower rates in LS cells (Figure 6.6). The only exception is ICDH, which carries the same flux as ACONT mainly because isocitrate is not consumed anywhere else in the model.

Third, comparison of activity ratios among complexes I-IV shows that complex II is the most deficient (Table 6.6). Although many LS case reports have provided activities for the various respiratory chain complexes, and combinations thereof, direct comparison between the model's predicted fluxes and these data is difficult because *i)* spectrophotometric and polarographic measurements of enzyme activities *in vitro* only reflect enzyme capacity (i.e., $V_{max}$) and not their physiological activities[233] and *ii)* reported values vary significantly from patient to patient and study to study[233-235]. On the other hand, Rustin and colleagues suggested that ratios of complex activities, particularly COX/SCCR (cytochrome c oxidase to succinate cytochrome c reductase), are less sensitive to patients' age, physical training, and tissue type, and are thus more reliable parameters for diagnosis[233,236-238]. Evaluating the activities of ATP synthase and COX relative to the remaining complexes shows that complex II (SCCR) is the most likely the deficient complex in the present Leigh-affected cell (Table 6.6).

## 9.3 Treatment options

Though there is no recognized treatment for Leigh and Leigh-like syndromes, there have been a number of nutritional treatments and interventions with limited successes reported[213,239]. In particular, LS with PDHC deficiency has been successfully treated with thiamin, a cofactor of the PDH enzyme, and alternative energy sources such as carnitine and/or ketogenic diets[240,241]. Patients with defective ATP synthase have also benefited from creatine monohydrate supplement[242,243]. On the other hand, due to the similar roles complexes I, II, and IV have on oxidative phosphorylation, LS patients with these deficiencies were given similar vitamin cocktails containing coenzyme $Q_{10}$, riboflavin, and vitamins C, $K_3$, and E[213,239,244].

In summary, the present study describes the application of tracer-based metabolomics in characterizing physiological steady states assumed by fibroblasts obtained from normal and LS patient. Though the experiments and analysis presented here are more time consuming than current clinical tests, they provide a more complete picture of the metabolic state of the cells. In fact, just the elimination of PDHC and ATP synthase as the deficient complexes is sufficient to identify the most appropriate vitamin treatment for this patient. The model's prediction of a defective complex II also narrows the choices of candidate genes for sequencing. The current study thus serves as an important first step towards a comprehensive characterization of the different disorders presently classified as Leigh or Leigh-like syndromes.

(This chapter, in part, is a preprint of the material appearing in "Systems analysis of energy metabolism elucidates the affected respiratory chain complex in Leigh's syndrome, Vo TD, Lee PWN, and Palsson BO, *Molecular genetics and metabolism,* submitted. The dissertation author is the primary investigator and author of this paper.)

**Table 6.1: Composition of media A and B**
Concentrations shown are provided by Invitrogen catalog. In addition to the tabulated components, the two media also contain vitamins (choline chloride, D-calcium pantothenate, folic acid, i-inositol, niacinamide, pyridoxine hydrochloride, riboflavin, and thiamine hydrochloride) and inorganic salts ($CaCl_2$, $Fe(NO_3)_3 \cdot 9H_2O$, $MgSO_4$, KCl, $NaHCO_3$, NaCl, and $NaH_2PO_4 \cdot H_2O$). Equal weight (584 mg/L) instead of equimolar concentrations of the two amino acids is used to partially compensate for the potentially lower energy capacity in asparagine (fewer carbons per asparagine molecule).

| Medium A | | Medium B | |
|---|---|---|---|
| **Components** | **Conc (mM)** | **Components** | **Conc (mM)** |
| Glucose | 11.12 | Glucose | 11.12 |
| Sodium pyruvate | - | Sodium pyruvate | 1.0 |
| Glycine | 0.4 | Glycine | 0.4 |
| L-Arginine hydrochloride | 0.398 | L-Arginine hydrochloride | 0.398 |
| L-Cystine 2HCl | 0.201 | L-Cystine 2HCl | 0.201 |
| L-Glutamine | 4.0 | L-Asparagine | 4.4 |
| L-Histidine hydrochloride-$H_2O$ | 0.2 | L-Histidine hydrochloride-H2O | 0.2 |
| L-Isoleucine | 0.802 | L-Isoleucine | 0.802 |
| L-Leucine | 0.802 | L-Leucine | 0.802 |
| L-Lysine hydrochloride | 0.798 | L-Lysine hydrochloride | 0.798 |
| L-Methionine | 0.201 | L-Methionine | 0.201 |
| L-Phenylalanine | 0.4 | L-Phenylalanine | 0.4 |
| L-Serine | 0.4 | L-Serine | 0.4 |
| L-Threonine | 0.798 | L-Threonine | 0.798 |
| L-Tryptophan | 0.0784 | L-Tryptophan | 0.0784 |
| L-Tyrosine disodium salt dihydrate | 0.398 | L-Tyrosine disodium salt dihydrate | 0.398 |
| L-Valine | 0.803 | L-Valine | 0.803 |

**Table 6.2: Mass distribution of glutamate $C_2$-$C_5$ fragment**
This fragment does not include the backbone carboxylic carbon of the amino acid. Glutamate samples are isolated from cells at 24, 48, and 72 h. m0, m1, m2, m3, and m4 represent the molar fraction of each of these five mass isotopomers. m4 isotopomers are not detectable in a number of samples. The $^{13}C$ enrichment of these isotopomers increase over time. HepG2 cells cultured in medium B have higher overall $^{13}C$ enrichment in glutamate, although m4 glutamate is not detectable in these cells. Σm represents the sum of all mass isotopomers enriched with $^{13}C$. Columns m1/Σm and m2/Σm show the normalized fractional enrichment of singly and doubly labeled glutamate isotopomers, respectively, excluding unlabeled glutamate. Reported values are mean and standard deviation (shown under mean).

| | A | B | A | B | A | B |
|---|---|---|---|---|---|---|
| | 24 h | | 48 h | | 72 h | |
| **m0** | 0.9353<br>0.0027 | 0.9200<br>0.0176 | 0.9170<br>0.0036 | 0.8744<br>0.0095 | 0.9013<br>0.0047 | 0.8243<br>0.0049 |
| **m1** | 0.0262<br>0.0014 | 0.0471<br>0.0009 | 0.0363<br>0.0005 | 0.0619<br>0.0021 | 0.0433<br>0.0024 | 0.0880<br>0.0032 |
| **m2** | 0.0356<br>0.0010 | 0.0471<br>0.0016 | 0.0449<br>0.0005 | 0.0634<br>0.0015 | 0.0483<br>0.0020 | 0.0860<br>0.0027 |
| **m3** | 0.0017<br>0.0001 | 0.0007<br>0.0015 | 0.0026<br>0.0003 | 0.0037<br>0.0011 | 0.0035<br>0.0003 | 0.0020<br>0.0028 |
| **m4** | 0.0012<br>0.0005 | - | - | - | 0.0038<br>0.0005 | - |
| **Σm** | 0.0647<br>0.0027 | 0.0953<br>0.0012 | 0.0830<br>0.0036 | 0.1312<br>0.0038 | 0.0987<br>0.0047 | 0.1761<br>0.0045 |
| **m1/ Σm** | 0.4049<br>0.0053 | 0.4943<br>0.0136 | 0.4379<br>0.0142 | 0.4716<br>0.0119 | 0.4381<br>0.0067 | 0.4996<br>0.0114 |
| **m2/ Σm** | 0.5500<br>0.0092 | 0.4942<br>0.0117 | 0.5421<br>0.0207 | 0.4838<br>0.0134 | 0.4890<br>0.0033 | 0.4883<br>0.0104 |

**Table 6.3: Mass distributions of palmitate**
Twelve ions (m/z 169-280) are monitored and the most abundant ions are shown here. The remaining ions make up about 2% of the total. The m4/m2 ratios are used to estimate the fractional enrichment of intracellular acetyl-CoA and fractional synthesis rate of palmitate (Equations 7-8) using the pathway-based method. Reported values are means ± standard deviations.

| | Media A | Media B |
|---|---|---|
| **m0** | 0.7372 ± 0.0243 | 0.8100 ± 0.0431 |
| **m1** | 0.0318 ± 0.0037 | 0.0243 ± 0.0052 |
| **m2** | 0.1614 ± 0.0153 | 0.1163 ± 0.0266 |
| **m3** | 0.0158 ± 0.0017 | 0.0115 ± 0.0026 |
| **m4** | 0.0430 ± 0.0037 | 0.0305 ± 0.0068 |
| **m5** | 0.0041 ± 0.0006 | 0.0029 ± 0.0007 |
| **m6** | 0.0067 ± 0.0006 | 0.0045 ± 0.0015 |
| **m4/m2** | 0.2665 ± 0.0571 | 0.2627 ± 0.1047 |

**Table 6.4: Causes of Leigh or Leigh-like disease**
Thus, there are at least 3 major causes of Leigh syndrome, each transmitted by a different mode of inheritance: X-linked recessive, mitochondrial, and autosomal recessive. The incidence of strictly defined LS is 1 per 77,000 and of LS and Leigh-like disease is 1 per 44,000 [214].

| Defects | Frequency | Components of affected genes |
|---|---|---|
| PDHC | 10 | E1α subunit (X-linked) |
| Complex I | 19 | mitochondrial-encoded MTND3, MTND5, and MTND6<br>nuclear-encoded NDUFV1, NDUFS1, NDUFS3, NDUFS4, NDUFS7, and NDUFS8 |
| Complex II | ~39 | Fp subunit of SDHA |
| Complex III | | BCS1L |
| Complex IV (COX) | 14 | mitochondrial-encoded MTCO3<br>nuclear-encoded COX10, COX15, SCO2, and SURF1 |
| Complex V | 18 | ATPase6 encoded by mtDNA (maternally inherited) |

**Table 6.5: Comparison between predicted and published reaction rates**
Predicted fluxes are showed for control fibroblasts. Model-predicted flux values are constrained by both enzymatic capacity (Vmax) and network stoichiometry. Published values are taken from *in vitro* experiments, where reaction rates are not limited by substrate limitation. Unit: nmol/hr/mil cells.

| Enzyme | Predicted flux | Published value | References |
|---|---|---|---|
| PDH | 30.7 ± 2.3 | 30-45 | [222,223] |
| CS | 29 ± 19 | 600-900, 1600 | [224,225] |
| ACONT | 29 ± 19 | 12-30 | [224] |
| ICDH | 28 ± 20 | 36-90 | [224] |
| ME | 34.6 ± 10.8 | 9-63 | [221] |

**Table 6.6: Absolute and relative activities of different complexes in the respiratory chain.**
**A)** Though the absolute activities of all five respiratory chain complexes are lower for LS cells compared to the control, they are all within on SD of the mean fluxes predicted for the normal fibroblasts. Flux ratio of ATP synthase relative to each of the remaining complexes suggests that ATP synthase is not the limiting step in the respiratory chain. The high activity ratio of ATPS/CII for the LS cells implies that CII is limiting factor. **B)** Ratios of complex IV (COX) relative to the remaining complexes also indicate that complex II (SCCR) is likely to be the most defective complex.

**A)**

| | Mean flux ± SD | | ATPS/$C_i$ ($i$ = I .. IV) | |
|---|---|---|---|---|
| | Control | Leigh | Control | Leigh |
| CI | 223 ± 41 | 195 ± 2 | 3.1 | 3.1 |
| CII | 87 ± 19 | 65 ± 4 | **7.9** | **9.4** |
| CIII | 310 ± 58 | 259 ± 4 | 2.2 | 2.3 |
| CIV | 160 ± 31 | 136 ± 1 | 4.3 | 4.4 |
| CV | 686 ± 107 | 601 ± 2 | | |

**B)**

| | Control | Leigh |
|---|---|---|
| CIV/CI | 1.43 ± 0.03 | 1.40 ± 0.02 |
| CIV/CII | **3.72 ± 0.33** | **4.28 ± 0.36** |
| CIV/CIII | 1.03 ± 0.01 | 1.05 ± 0.01 |
| CIV/CV | 0.47 ± 0.04 | 0.45 ± 0.004 |

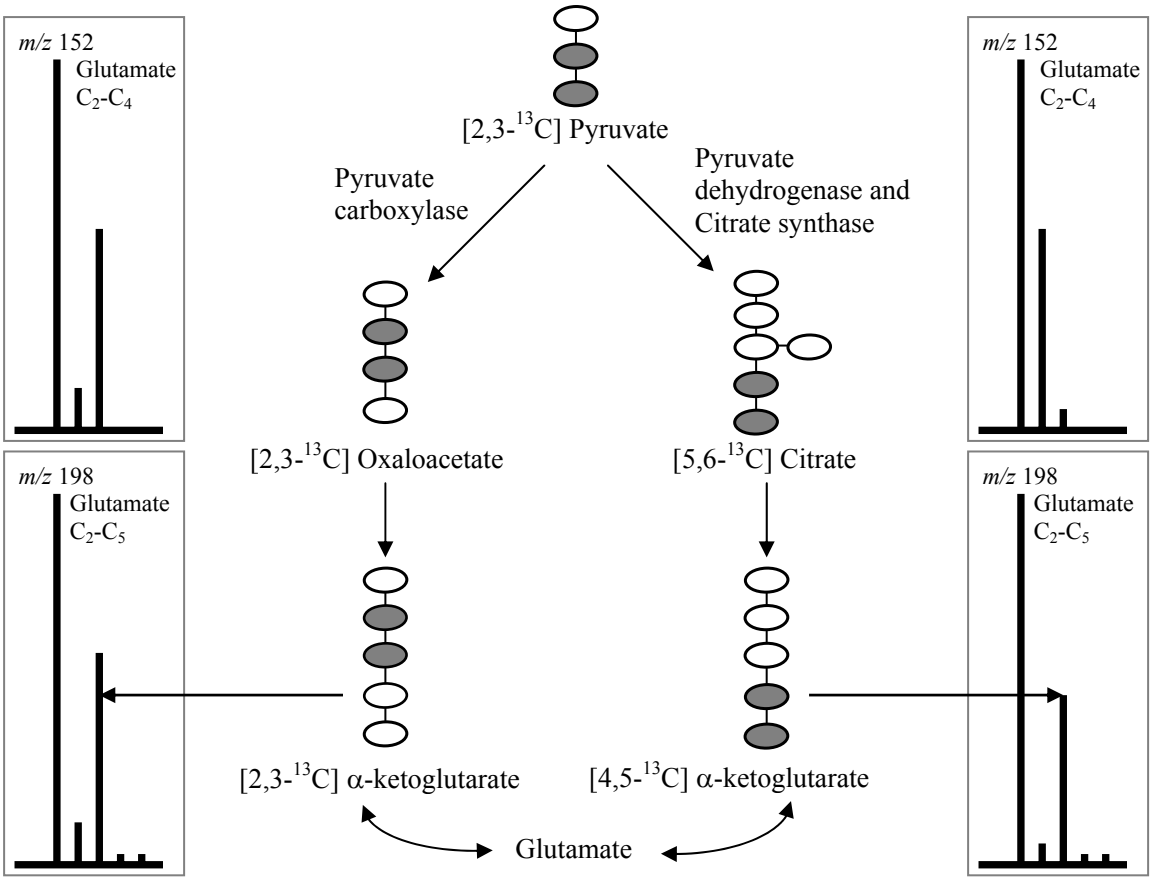



**Figure 6.1: Schematic carbon transmission from [1,2 $^{13}C_2$] glucose to glutamate**
The following carbon transfer scheme is assumed to infer reaction fluxes in the pathway-based method. Labeled carbons (shaded) in m2 pyruvate yield labeled oxaloacetate and labeled citrate through processes of carboxylation and decarboxylation, respectively. The first turn of the TCA cycle turns these compounds into m2 glutamates (m/z 198 fragments).

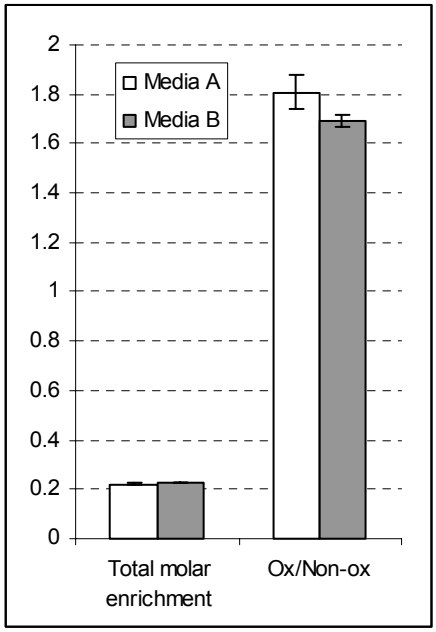



**Figure 6.2: Ribose enrichment and the inferred oxidative to non-oxidative ratio (Ox/Non-ox) of fluxes in the pentose phosphate cycle**
Total molar enrichment is the sum of the fractional molar enrichment of m1 to m5 riboses. This enrichment does not change significantly between cells grown in media A and B, suggesting the similar RNA synthesis and growth rates between the two cell cultures. Error bars reflect values of standard deviations over four biological replicates and three analytical measurements per replicate; some are too small to be seen. The ratios of oxidative to non-oxidative branches in the pentose phosphate cycle are also similar for HepG2 cells in two media.

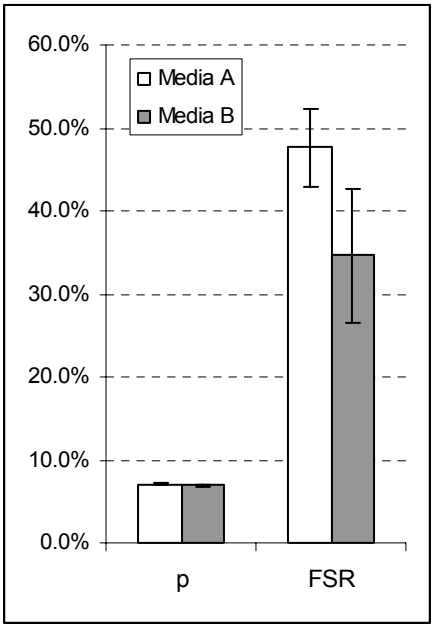



**Figure 6.3: Fractional enrichment of acetyl-CoA and the inferred fractional synthesis rate of palmitate**
The enrichment of $^{13}C$ from glucose to acetyl-CoA is not different between HepG2 cells grown in media A (glutamine containing) and B (asparagine containing). The palmitate fractional synthesis rate, i.e. the rate of *de novo* palmitate synthesis, is decreased when cells are cultured with asparagine. These results are consistent with reaction fluxes estimated by the network-based analysis.

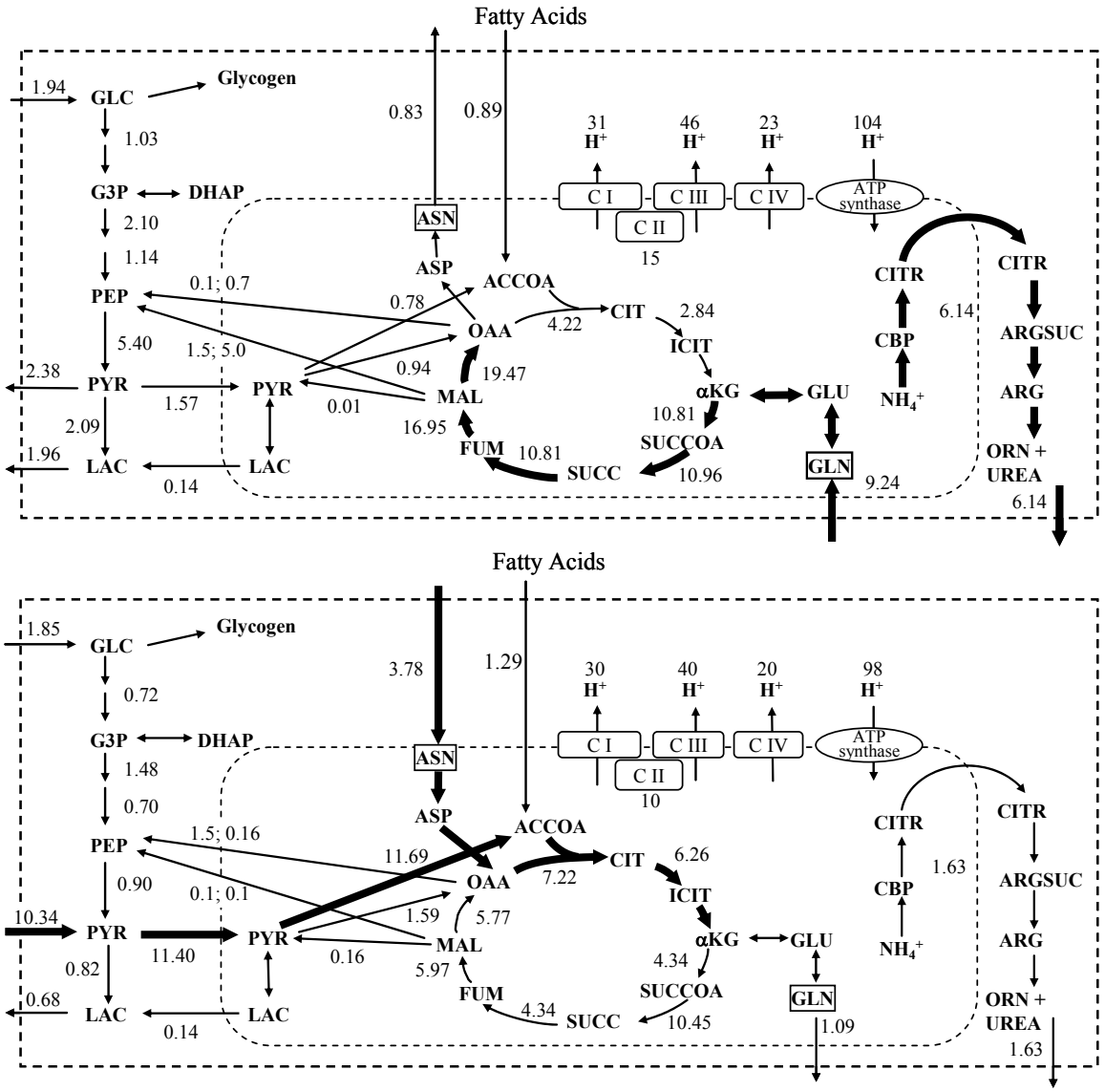



**Figure 6.4: Estimated reaction fluxes calculated using the network-based method for the metabolic network of HepG2 cells cultured in media A and B**

Only a selected group of reactions and their flux values are shown here to illustrate the different flux pattern displayed by cells in the two cultures. Reference 50 contains estimated reaction fluxes for all 254 reactions in the model. The first flux values of malic enzyme and PEPCK are of their cytosolic isoforms; the second correspond to the mitochondrial isoforms. Within the same panel, the widths of arrows do not correspond to the relative magnitude of reaction fluxes, but are used to compare flux values of the same reactions between the two panels. All reactions of the urea cycle have the same flux in each panel. All flux values have the same unit, nmol/min/mg protein.

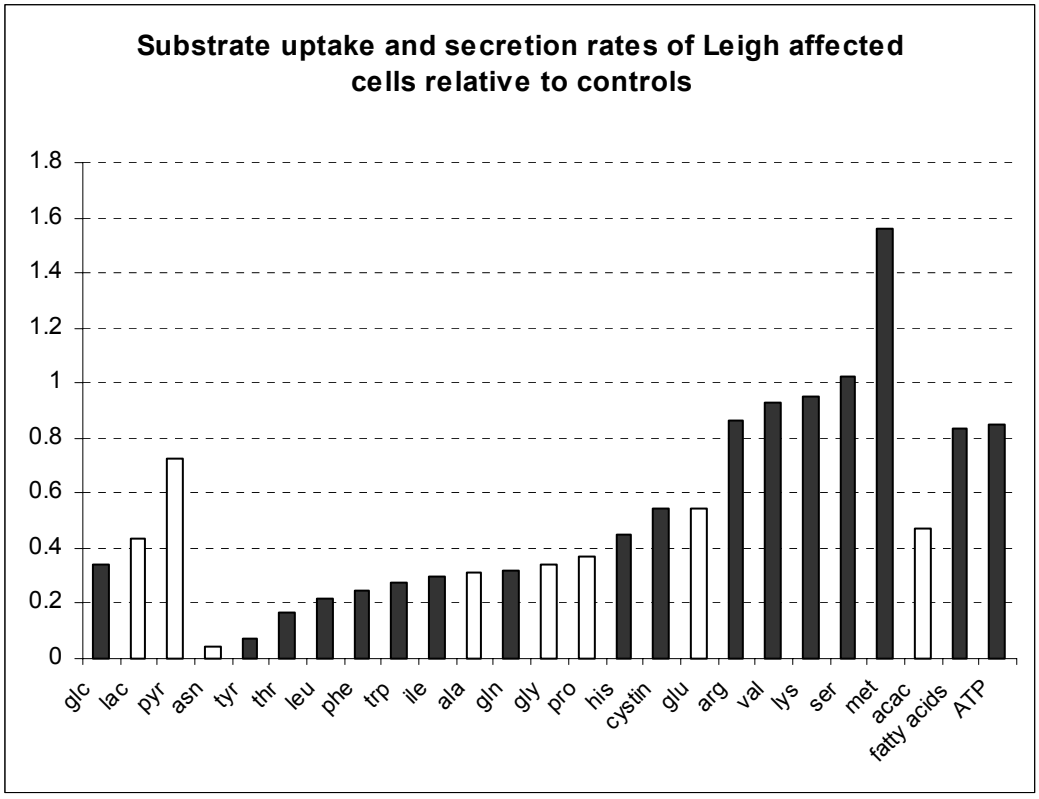
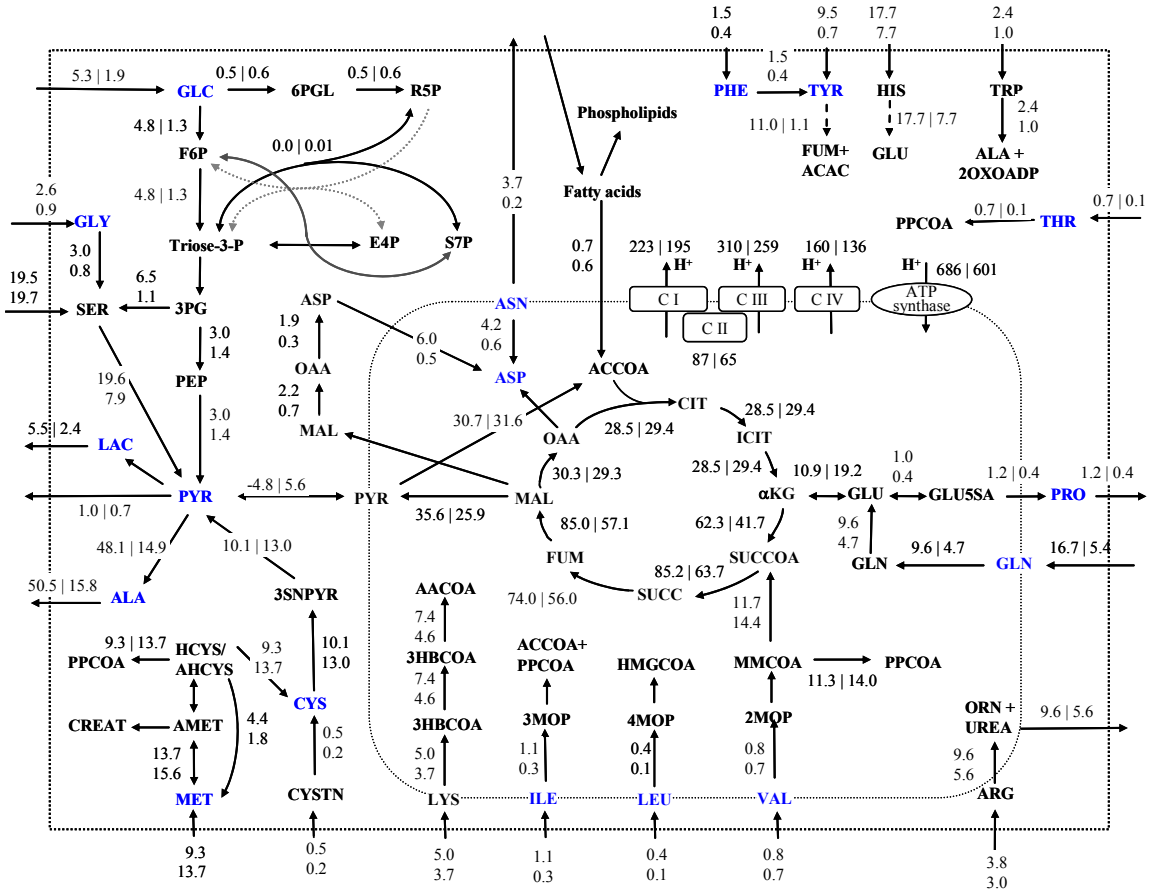



**Figure 6.5: Substrate uptake and secretions rates of Leigh affected cells relative to controls**
Substrate uptake (black) and secretion (white) rates that are below the line y=1 are lower in LS cells as compared to normal fibroblasts. For the majority of these substrates, the lower uptake and secretion rates reflect the slower metabolism of LS cells. Specifically, LS cells produce only about 80% of the amount of ATP produced by normal cells.

**Figure 6.6: Schematic map and estimated fluxes of the reactions in the fibroblast metabolic network**
In each reaction, the first number is the flux values calculated from the controls followed by that of LS cells. Each arrow may represent multiple reactions in the model. Secondary metabolites (proton or water) are not shown, but are included in the stoichiometric matrix. Metabolites in blue are those isolated and measured in this experiment. Definitions for metabolite abbreviations can be found in reference . Flux unit is nmol/hr/mil cells



# Chapter 7
# Conclusions and Future Outlook

Biology is a dynamic and expanding discipline. It is perhaps the one field where technology advances sometimes precede philosophical development. There is no doubt that the emergence of high-throughput technology has given rise to bioinformatics and systems biology, and it not until now that we are starting to define and differentiate these two fields from traditional biology. Perhaps what sets these two fields apart is the employment of mathematics and a desire to predict systemic behaviors of organisms from knowledge of their constituent processes. Towards this goal, the work in this dissertation involves the design of novel methods of analysis as well as their applications to uncover new discoveries using the constraint-based modeling approach as the focal point.

## 1. Contributions to the field

The work described in this dissertation and accompanying peer-reviewed manuscripts covers the development of new analytical methods as well as novel discoveries resulting from application of such methodologies. Chapter 1 provides an overview of some of the latest technology, biological data they generate, and the opportunities and challenges they present. Chapter 2 introduces metabolic network reconstruction as a tool for effective data integration and analysis. Though metabolic network reconstruction is not a new concept, the types of data previously used for such

purposes had been largely textbook-based canonical pathways and genome annotations. In addition, existing networks mostly describe bacterial systems such as *E. coli*, where the entire intracellular reaction network is localized in the same compartment and enzymatic and molecular components of most biochemical pathways are well studied. Expanding the network construction formalism to account for other data types and more complex cellular systems requires the introduction of new concepts. Three new data types, proteomics, microarray, and transcript variants, can now be incorporated into a reconstructed network based on the rules described in chapter 2. In addition, two new types of metabolic systems, organelles (cardiac mitochondria) and human cells (hepatocytes and fibroblasts), can now be represented with reconstructed networks. Defining the contents of these networks is not as straight-forward as for bacterial systems for two reasons. First, the system boundary is not obvious because many biochemical pathways span both the mitochondrial and cytosolic compartments. In fact accounting only for matrix reactions would result in a "bag" of unconnected reactions and a functionless network. Thus, the scope of a metabolic network should be defined in terms of metabolic functions it describes rather than by its physical boundary. Second, expanding the reconstruction formalism from single to multi-compartmental systems requires connections, i.e. transporters, that allow metabolites to move from one compartment to another. However, as transporters are the least characterized set of enzymes[16], the decision to include or exclude a particular transporter relies on physiological data. In short, these challenges and solutions serve as guidelines for future size and scope expansion of network reconstruction. Indeed, the transition from

microbial systems to human systems is a stepping stone in the advancement of systemic study of human health and disease.

Chapter 3 details the functional analysis of the mitochondrial metabolic network using the constraint-based framework and flux balance analysis. Neither constraint-based modeling nor flux balance analysis is a novel concept; however, their application in studying mitochondrial metabolism is unique because unlike microbials whose primary metabolic objective is growth, the metabolic objective(s) of mitochondria are not obvious. Novel techniques such as multiple objective analysis and alternate pathway identification are designed to accommodate the multiple functions of these organelles. The second part of this chapter introduces random sampling of the steady state flux space as a way to evaluate effects of metabolic disturbance on the entire network. The novelty of this study is twofold. First, the method makes it possible to access the entire distribution of allowable flux values for each reaction. Second, results show that the number of accessible steady states is drastically reduced in diabetic and ischemic conditions.

Chapter 4 describes step-by-step procedures to incorporate isotopomer or tracer data into the constraint-based framework for reaction flux estimation. This type of data has been used to infer flux ratios at network branch points (pathway-based method), but its use for computing absolute flux values in a large scale network is a novel development. The integration of isotopomer data is instrumental in determining the physiological steady states in the cells without imposing an assumed metabolic objective.

Chapter 5 applies this method to elucidate flux distributions in a perfused mouse heart. A further advantage of model-centric rather than pathway-based analysis of isotopomer data is that the model can be used to simulate and analyze experimental scenarios beyond the original experimental conditions.

Lastly, chapter 6 applies all the tools developed in the previous chapters to study cellular metabolic phenotypes of tissue culture in response to substrate availability and disease conditions. This is an advancement in two ways. First, these studies show that constraint-based modeling, which previously has only been applied to microbial systems, can now be applied to elucidate metabolic phenotypes of human cells with the addition of tracer data. Second, the ability to characterize intracellular resource distribution as a function of medium composition as well as genetic limitations is an improvement made in both nutrigenomics (deciphering the nutrient-gene interaction) and disease diagnosis and characterization (Leigh's disease).

# 2. Lessons learned

## 2.1 Stoichiometric constraints alone are not sufficient to predict mammalian cellular phenotypes

Every modeling approach has its triumphs and failures. A good scientist understands the limitations of his tools so as not to use them inappropriately. The first lesson I have learned is that stoichiometric constraints and the flux balance analysis formalism alone are insufficient to identify metabolic phenotypes in mammalian cells.

Certainly this is no reason to abandon constraint-based modeling as a whole, but is a reason to search for additional constraints. Figure 7.1 shows that even if the functions of all *n* genes in the genome (which translate to stoichiometric constraints on the corresponding reactions) are known, that is only a fraction of the phenotypes the encoded protein can exhibit, giving rise to a much large possibilities of metabolic phenotypes. Clearly the ability to determine the unique metabolic state the cell occupies in this vast phenotypic space requires knowledge of the intricate regulation and coordination of different pathways and functions in the cellular system. In the absence of this knowledge, an alternative approach would be to simulate cellular activities based on observed input and output of the system. The difference between simulation and prediction is that the former does not claim to represent what governs the behavior of the systems, but simply to emulate the observable phenotypes to infer unobservable activities.

The addition of isotopomer data and isotopomer balance constraints is essential for the present model to simulate the steady state assumed by the cell. In fact, searching for a solution (i.e. flux distribution) that best fits observed $^{13}C$ labels is an empirical approach, and thus the more isotopomer data obtained, the more reliable the results. Though this seems like a "black-box" approach, it is currently the best solution for shedding light on the intracellular reaction fluxes. In particular, since most metabolic networks have a high level of "theoretical" redundancy (i.e. high number of alternate optima), isotopomer data allows us to identify which pathways are actually taken by the cell. Knowledge about *the paths not traveled* also provides clues about the missing

constraints, or regulatory rules, of the system. Accordingly, the lesson about the deficiency of stoichiometric constraints is also a lesson about the usefulness of isotopomer data in filling in missing constraints.

## 2.2 Biological models and model-centric discoveries

The second lesson is that models will hold an instrumental/integral role in biological discoveries in the future. It is perhaps necessary to redefine the term *model*. A model is an abstraction of the salient characteristics of a system or process. It is thus necessary that a model is not as complex as the real system or it loses its purpose. A common criticism of the constraint-based modeling approach is that it is restricted to steady state behaviors. Section 3.1 discusses this subject more extensively; however, it is worthwhile to mention here that *i)* metabolic steady states are frequently achieved during the time scale of involved experiments (hours to days) and *ii)* dynamic behaviors obtained by integrating metabolite and enzyme concentrations (traditional kinetic approaches) are reliable only when initial concentrations of these components are known (which is usually not the case). What this all means is that models are a delicate balance between abstraction and realism, and effective models capture the relevant features of the biological systems without being unnecessarily complex.

With the risk of being biased, I believe that model-centric discoveries will lead the way to comprehensive understanding of biological systems. Quantitative models provide a framework to structure data in a concise and coherent organization which will

necessarily replace the arrow-based diagrams that have dominated molecular biology for years. The difference between these two model systems is that computational models can identify missing information as well as confirm known behaviors. The ability to simulate known behaviors through a quantitative model is necessary to be sure that we really understand the system.

### 2.3 Personal lesson

It is sometimes discouraging to think that the number of lessons I learned upon completing my graduate research is minuscule compared to the vast body of knowledge I acquired though my undergraduate education. This is of course an unfair comparison because most of the lessons I learned throughout graduate school were either self-taught or through researchers in very similar fields, while what I was taught at Berkeley is knowledge compiled from work of an uncountable number of people. The most valuable lesson I have learned, however, is that graduate school is about learning *how to learn*. The amount of existing knowledge is vast and more are accumulated everyday; it is not possible for anyone to know everything. However, the ability to figure out new concepts as well as critically assess and interpret new information is instrumental in the building of a scientist. To this end, I now comfortably know that graduate school has equipped me with an ability to learn.

## 3. Looking forward

## 3.1 The future of constraint-based modeling

The development of the constraint-based modeling framework is effectively the quest for the missing constraints. There are three areas that can be improved by formulating appropriate constraints in the new generation of constraint-based models. First, current proteomic and metabolomics data are represented in the model framework indirectly through the presence of the reactions in the biochemical network. Such representation does not account for the dynamical nature of these data in response to post-translational modification, allosteric regulation, ubiquitination, and substrate availability. The ability to integrate these effects in the modeling framework will lead to a big improvement in phenotypic prediction in future models. Second, biochemical processes such as voltage-gated ion channeling and osmo-regulation are concentration-based transient behaviors that can only be effectively modeled with kinetic expressions. An effective method to integrate kinetic expressions without requiring kinetic parameters for the entire system would be highly desirable. Lastly, though whole-cell models are a major advance over previous generations of biological models, complex diseases often involve processes that span multiple cell types and organs. Accordingly new formalisms and methods are needed to model inter-cellular communication as well as inter-organ interactions.

## 3.2 The future of systems biology

In order for systems biology to significantly impact biological research in the future, collaborations among scientist of different biological sub-disciplines are necessary

to develop a formal language and standards for biology. The need for such standards is reflected in the paradox that while the field is flooded with high-throughput and legacy data[245-248], every modeler, including those whose work is reviewed here, is keenly aware of the absence of data needed to complete his study. Responsibilities for realizing such standards should be accepted by both experimental and computational biologists. As an experimental biologist, such responsibilities may include standardizing criteria for model systems, experimental protocols, and data presentation as well as making samples (cell lines, vector constructs, proteins, etc) accessible to those who are interested. Such availability would not only facilitate verification and reproducibly of new findings but will also form new collaborations among researchers. Responsibilities of computational biologists, on the other hand, may include outlining the specificity of needed data, providing testable hypothesis, suggesting experimental designs, and making models available for others. Currently many computational studies only briefly describe the underlying algorithms and do not publish associated computational programs as it is not required by most peer-reviewed journals. However, in order for results to be reproducible, documented source code of such programs should accompany every study.

Perhaps one way to accelerate the realization of such standards and driving systems biology forward is to have experimental and computational researchers meet each other half way. It is probably no longer realistic for each group to stay within the boundary of their respective disciplines. In order to transform biology into a more quantitative discipline, efforts should be contributed from both groups of researchers. Experimentalists should understand the basics of computer-aided analytical tools, and

computational biologists ought to understand the systems beyond the mathematical representation by experiencing experimental biology first-hand at the bench. Through interdisciplinary training, perhaps the next generation of researchers will no longer have to identify themselves as either experimentalist or computational biologist, but rather biologists of a new era, the systems biology era.

(This chapter, in part, is a preprint of the material appearing in "Building the power house: Recent advances in mitochondrial studies through proteomics and systems biology", Vo TD and Palsson BO. *Am J Physiol Cell Physiol*. 2006. in press. The dissertation author is the primary investigator and author of this paper.)

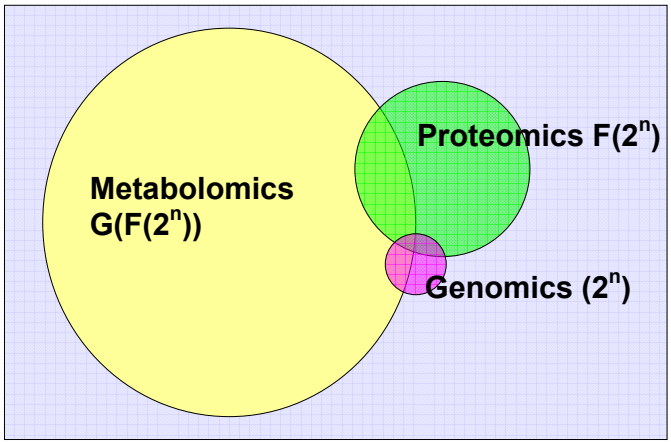



**Figure 7.1: Relationships among genotypes (genomics) and phenotypes exhibited by proteomics and metabolomics**
If there are n genes in the genome and each can be either expressed or not expressed, then the number of possible genotypes is $2^n$. The number of proteomic phenotypes is a function of these genotypes, expressing post-translational and allosteric regulation effects. The number of metabolomic phenotypes, thus metabolic phenotypes, in turn, is a compound function of the number of proteomic phenotype, as the activities of the encoded proteins are subject to constraints other than enzymatic regulation.

# Bibliography

**1.** Nowak R. Entering the postgenome era. *Science* 1995;270:368-9, 371.

**2.** McPherson JD, The International Human Genome Sequencing Consortium. A physical map of the human genome. *Nature* 2001;409:934-41.

**3.** Venter JC, Celera Genomics. The sequence of the human genome. *Science* 2001;291:1304-51.

**4.** Brett D, Pospisil H, Valcarcel J, Reich J, Bork P. Alternative splicing and genome complexity. *Nat Genet* 2002;30:29-30.

**5.** Sanford K, Soucaille P, Whited G, Chotani G. Genomics to fluxomics and physiomics - pathway engineering. *Curr Opin Microbiol* 2002;5:318-22.

**6.** Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995;270:467-70.

**7.** Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001;29:365-71.

**8.** Hack CJ. Integrated transcriptome and proteome data: the challenges ahead. *Brief Funct Genomic Proteomic* 2004;3:212-9.

**9.** Yates JR, 3rd, Gilchrist A, Howell KE, Bergeron JJ. Proteomics of organelles and large cellular structures. *Nat Rev Mol Cell Biol* 2005;6:702-14.

**10.** Wu Q, Van Orden S, Cheng X, Bakhtiar R, Smith RD. Characterization of cytochrome c variants with high-resolution FTICR mass spectrometry: correlation of fragmentation and structure. *Anal Chem* 1995;67:2498-509.

**11.** Solouki T, Emmett MR, Guan S, Marshall AG. Detection, number, and sequence location of sulfur-containing amino acids and disulfide bridges in peptides by ultrahigh-resolution MALDI FTICR mass spectrometry. *Anal Chem* 1997;69:1163-8.

**12.** Leavell MD, Leary JA, Yamasaki R. Mass spectrometric strategy for the characterization of lipooligosaccharides from Neisseria gonorrhoeae 302 using FTICR. *J Am Soc Mass Spectrom* 2002;13:571-6.

**13.** Soga T, Ohashi Y, Ueno Y, Naraoka H, Tomita M, Nishioka T. Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J Proteome Res* 2003;2:488-94.

**14.** Schiffer E, Mischak H, Novak J. High resolution proteome/peptidome analysis of body fluids by capillary electrophoresis coupled with MS. *Proteomics* 2006.

**15.** Eddy SR. "Antedisciplinary" science. *PLoS Comput Biol* 2005;1:e6.

**16.** Duarte N, Becker S, Jamshidi N, Thiele I, Mo M, Vo T, Srivas R, Palsson B. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. U.S.A.* 2006; accepted.

**17.** Duarte NC, Palsson BO, Fu P. Integrated analysis of metabolic phenotypes in *Saccharomyces cerevisiae*. *BMC Genomics* 2004;5:63.

**18.** Reed JL, Vo TD, Schilling CH, Palsson BO. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol* 2003;4:R54.

**19.** Lazebnik Y. Can a biologist fix a radio? -- Or, what I learned while studying apoptosis, (Cancer Cell. 2002 Sep;2(3):179-82). *Biochemistry (Mosc)* 2004;69:1403-6.

**20.** Covert MW, Schilling CH, Famili I, Edwards JS, Goryanin, II, Selkov E, Palsson BO. Metabolic modeling of microbial strains in silico. *Trends Biochem Sci* 2001;26:179-86.

**21.** Price ND, Reed JL, Palsson BO. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol* 2004;2:886-97.

**22.** Reed JL, Famili I, Thiele I, Palsson BO. Towards multidimensional genome annotation. *Nat Rev Genet* 2006;7:130-41.

**23.** Price ND, Famili I, Beard DA, Palsson BO. Extreme pathways and Kirchhoff's second law. *Biophys J* 2002;83:2879-82.

**24.** Beard DA, Liang SD, Qian H. Energy balance for analysis of complex metabolic networks. *Biophys J* 2002;83:79-86.

**25.** Paphrag. DNA microarray: wikipedia, 2006:http://en.wikipedia.org/wiki/DNA_microarray.

**26.** Ge H, Walhout AJ, Vidal M. Integrating 'omic' information: a bridge between genomics and systems biology. *Trends Genet* 2003;19:551-60.

**27.** Covert MW, Schilling CH, Famili I, Edwards JS, Goryanin II, Selkov E, Palsson BO. Metabolic modeling of microbial strains in silico. *Trends Biochem. Sci.* 2001;26:179-186.

**28.** Salzberg SL, Delcher AL, Kasif S, White O. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res* 1998;26:544-8.

**29.** Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* 1999;27:4636-41.

**30.** Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 1997;268:78-94.

**31.** Burge CB. Modeling dependencies in pre-mRNA splicing signals. In: Salzberg S, Searls D, Kasif S, eds. Computational Methods in Molecular Biology. Amsterdam: Elsevier Science, 1998:127-163.

**32.** Henikoff S. Scores for sequence searches and alignments. *Curr Opin Struct Biol* 1996;6:353-60.

**33.** Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389-402.

**34.** Altschul SF. Amino acid substitution matrices from an information theoretic perspective. *J Mol Biol* 1991;219:555-65.

**35.** Issac B, Raghava GP. GWFASTA: server for FASTA search in eukaryotic and microbial genomes. *Biotechniques* 2002;33:548-50, 552, 554-6.

**36.** Pearson WR. Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol* 1990;183:63-98.

**37.** Pearson WR. Using the FASTA program to search protein and DNA sequence databases. *Methods Mol Biol* 1994;25:365-89.

**38.** Durbin R. Biological sequence analysis : probabalistic models of proteins and nucleic acids. Cambridge, UK New York: Cambridge University Press, 1998:xi, 356 p.

**39.** Eddy SR. Profile hidden Markov models. *Bioinformatics* 1998;14:755-63.

**40.** Knudsen TB, Daston GP. MIAME guidelines. *Reprod Toxicol* 2005;19:263.

**41.** Shields R. MIAME, we have a problem. *Trends Genet* 2006;22:65-6.

**42.** Taylor SW, Fahy E, Zhang B, Glenn GM, Warnock DE, Wiley S, Murphy AN, Gaucher SP, Capaldi RA, Gibson BW, Ghosh SS. Characterization of the human heart mitochondrial proteome. *Nat Biotechnol* 2003;21:281-6.

**43.** Jiang XS, Zhou H, Zhang L, Sheng QH, Li SJ, Li L, Hao P, Li YX, Xia QC, Wu JR, Zeng R. A High-throughput approach for subcellular proteome: Identification of rat liver proteins using subcellular fractionation coupled with two-dimensional liquid chromatography tandem mass spectrometry and bioinformatic analysis. *Mol Cell Proteomics* 2004.

**44.** Westermann B, Neupert W. '*Omics*' of the mitochondrion. *Nat Biotechnol* 2003;21:239-40.

**45.** Danial NN, Gramm CF, Scorrano L, Zhang CY, Krauss S, Ranger AM, Datta SR, Greenberg ME, Licklider LJ, Lowell BB, Gygi SP, Korsmeyer SJ. BAD and glucokinase reside in a mitochondrial complex that integrates glycolysis and apoptosis. *Nature* 2003;424:952-6.

**46.** Vo TD, Greenberg HJ, Palsson BO. Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data. *J Biol Chem* 2004;279:39532-40.

**47.** Voet D, Voet JG, Pratt CW. Fundamentals of Biochemistry. New York: Wiley, 1999.

**48.** Arai M, Yokosuka O, Hirasawa Y, Fukai K, Chiba T, Imazeki F, Kanda T, Yatomi M, Takiguchi Y, Seki N, Saisho H, Ochiai T. Sequential gene expression changes in cancer cell lines after treatment with the demethylation agent 5-Aza-2'-deoxycytidine. *Cancer* 2006;106:2514-2525.

**49.** Wu M, Chen S, Wu X. Differences in cytochrome P450 2C19 (CYP2C19) expression in adjacent normal and tumor tissues in chinese cancer patients. *Med Sci Monit* 2006;12:BR174-178.

**50.** Fukuda H, Ebara M, Okuyama M, Sugiura N, Yoshikawa M, Saisho H, Shimizu R, Motoji N, Shigematsu A, Watayo T. Increased metabolizing activities of the tricarboxylic acid cycle and decreased drug metabolism in hepatocellular carcinoma. *Carcinogenesis* 2002;23:2019-23.

**51.** Vo T, Lim S, Lee P, Palsson B. Isotopomer Analysis of Cellular Metabolism in Tissue Culture: A Comparative study between the pathway and network-based methods. *Metabolomics* 2006;in press.

**52.** Csoka AB, English SB, Simkevich CP, Ginzinger DG, Butte AJ, Schatten GP, Rothman FG, Sedivy JM. Genome-scale expression profiling of Hutchinson-Gilford progeria syndrome reveals

widespread transcriptional misregulation leading to mesodermal/mesenchymal defects and accelerated atherosclerosis. *Aging Cell* 2004;3:235-43.

**53.** Vo T, Lee P, Palsson B. Systems analysis of energy metabolism elucidates the affected respiratory chain complex in Leigh's syndrome. *Molecular Genetics and Metabolism* 2006; submitted.

**54.** Eto K, Sakura H, Yasuda K, Hayakawa T, Kawasaki E, Moriuchi R, Nagataki S, Yazaki Y, Kadowaki T. Cloning of a complete protein-coding sequence of human platelet-type phosphofructokinase isozyme from pancreatic islet. *Biochem Biophys Res Commun* 1994;198:990-8.

**55.** Ibarra RU, Edwards JS, Palsson BO. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* 2002;420:186-9.

**56.** Price ND, Papin JA, Palsson BO. Determination of redundancy and systems properties of the metabolic network of Helicobacter pylori using genome-scale extreme pathway analysis. *Genome Res* 2002;12:760-9.

**57.** Famili I, Palsson BO. The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools. *Biophys J* 2003;85:16-26.

**58.** Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO. Integrating high-throughput and computational data elucidates bacterial networks. *Nature* 2004;429:92-6.

**59.** Barrett CL, Herring CD, Reed JL, Palsson BO. The global transcriptional regulatory network for metabolism in *Escherichia coli* exhibits few dominant functional states. *Proc Natl Acad Sci U S A* 2005;102:19103-8.

**60.** Papin JA, Palsson BO. The JAK-STAT signaling network in the human B-cell: an extreme signaling pathway analysis. *Biophys J* 2004;87:37-46.

**61.** Cortassa S, Aon MA, Winslow RL, O'Rourke B. A mitochondrial oscillator dependent on reactive oxygen species. *Biophys J* 2004;87:2060-73.

**62.** Aon MA, Cortassa S, Marban E, O'Rourke B. Synchronized whole cell oscillations in mitochondrial metabolism triggered by a local release of reactive oxygen species in cardiac myocytes. *J Biol Chem* 2003;278:44735-44.

**63.** Ramakrishna R, Edwards JS, McCulloch A, Palsson BO. Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. *Am J Physiol Regul Integr Comp Physiol* 2001;280:R695-704.

**64.** Kaufman DE, Smith RL. Direction choice for accelerated convergence in hit-and-run sampling. *Operations Research* 1998;46:84-95.

**65.** Bonarius HPJ, Schmid G, Tramper J. Flux analysis of underdetermined metabolic networks: The quest for the missing constraints. *Trends in Biotechnology* 1997;15:308-314.

**66.** Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng* 2003;5:264-76.

**67.** Schilling CH, Letscher D, Palsson BO. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of Theoretical Biology* 2000;203:229-48.

**68.** Anderson ED. The MOSEK Optimization Toolbox for MATLAB, Version 2.5, User's Guide and Reference Manual, http://www.mosek.com. Denmark, 2002.

**69.** Avis D. lrs: A Revised Implementation of the Reverse Search Vertex Enumeration Algorithm. In: Ziegler GKG, ed. Polytopes - Combinatorics and Computation: Birkhauser-Verlag, 2000:177-198.

**70.** Avis D. lrs homepage, http://cgm.cs.mcgill.ca/~avis/C/lrs.html, 2003.

**71.** Aimar-Beurton M, Korzeniewski B, Letellier T, Ludinard S, Mazat JP, Nazaret C. Virtual mitochondria: metabolic modelling and control. *Mol Biol Rep* 2002;29:227-32.

**72.** Cortassa S, Aon MA, Marban E, Winslow RL, O'Rourke B. An integrated model of cardiac mitochondrial energy metabolism and calcium dynamics. *Biophys J* 2003;84:2734-55.

**73.** Korzeniewski B, Mazat JP. Theoretical studies on the control of oxidative phosphorylation in muscle mitochondria: application to mitochondrial deficiencies. *Biochem J* 1996;319 ( Pt 1):143-8.

**74.** Antunes F, Han D, Cadenas E. Relative contributions of heart mitochondria glutathione peroxidase and catalase to $H_2O_2$ detoxification in in vivo conditions. *Free Radic Biol Med* 2002;33:1260-7.

**75.** Searcy DG. Metabolic integration during the evolutionary origin of mitochondria. *Cell Res* 2003;13:229-38.

**76.** Salway JG. Metabolism at a glance. Oxford ; Malden, MA: Blackwell Science, 1999:111.

**77.** Stryer L. Biochemistry. New York: W.H. Freeman, 1995:xxxii, 1089.

**78.** Grynberg A, Demaison L. Fatty acid oxidation in the heart. *J Cardiovasc Pharmacol* 1996;28 Suppl 1:S11-7.

**79.** Rocquelin G, Guenot L, Justrabo E, Grynberg A, David M. Fatty acid composition of human heart phospholipids: data from 53 biopsy specimens. *J Mol Cell Cardiol* 1985;17:769-73.

**80.** Rocquelin G, Guenot L, Astorg PO, David M. Phospholipid content and fatty acid composition of human heart. *Lipids* 1989;24:775-80.

**81.** Lehninger AL, Cox MM, Nelson DL. Principles of biochemistry. New York, NY: Worth Publishers, 1993.

**82.** Atamna H, Walter PB, Ames BN. The role of heme and iron-sulfur clusters in mitochondrial biogenesis, maintenance, and decay with age. *Arch Biochem Biophys* 2002;397:345-53.

**83.** Sedman R, Ingall G, Rios G, Tephly TR. Heme biosynthesis in the heart. *Biochem Pharmacol* 1982;31:761-6.

**84.** Edwards JS, Ramakrishna R, Palsson BO. Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnol Bioeng* 2002;77:27-36.

**85.** Daum G. Lipids of mitochondria. *Biochim Biophys Acta* 1985;822:1-42.

**86.** Shiao YJ, Lupo G, Vance JE. Evidence that phosphatidylserine is imported into mitochondria via a mitochondria-associated membrane and that the majority of mitochondrial phosphatidylethanolamine is derived from decarboxylation of phosphatidylserine. *J Biol Chem* 1995;270:11190-8.

**87.** Stone SJ, Vance JE. Phosphatidylserine synthase-1 and -2 are localized to mitochondria-associated membranes. *J Biol Chem* 2000;275:34534-40.

**88.** Vance JE. Phospholipid synthesis in a membrane fraction associated with mitochondria. *J Biol Chem* 1990;265:7248-56.

**89.** Brown GC, Nicholls DG, Cooper CE. Mitochondria and cell death. Princeton University Press, 1999.

**90.** Lee S, Phalakornkule C, Domach MM, Grossmann IE. Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Comp Chem Eng* 2000;24:711-16.

**91.** Phalakornkule C, Lee S, Zhu T, Koepsel R, Ataai MM, Grossmann IE, Domach MM. A MILP-based flux alternative generation and NMR experimental design strategy for metabolic engineering. *Metab Eng* 2001;3:124-37.

**92.** Papin JA, Price ND, Palsson BO. Extreme pathway lengths and reaction participation in genome-scale metabolic networks. *Genome Research* 2002;12:1889-900.

**93.** Burgard AP, Nikolaev EV, Schilling CH, Maranas CD. Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res* 2004;14:301-12.

**94.** Thiele I, Price ND, Vo TD, Palsson BO. Candidate metabolic network states in human mitochondria. Impact of diabetes, ischemia, and diet. *J Biol Chem* 2005;280:11683-95.

**95.** Opie L. The heart physiology and metabolism. Raven Press, 1991.

**96.** Williamson JR, Ford C, Illingworth J, Safer B. Coordination of citric acid cycle activity with electron transport flux. *Circ Res* 1976;38:I39-51.

**97.** O'Donnell JM, Alpert NM, White LT, Lewandowski ED. Coupling of mitochondrial fatty acid uptake to oxidative flux in the intact heart. *Biophys J* 2002;82:11-8.

**98.** Berne RM, Levy, M.N., Koeppen, B.M., Stanton, B.A. Physiology. 2003.

**99.** Fukao T, Lopaschuk GD, Mitchell GA. Pathways and control of ketone body metabolism: on the fringe of lipid biochemistry. *Prostaglandins, Leukotrienes and Essential Fatty Acids* 2004;70:243-251.

**100.** Taegtmeyer H, McNulty P, Young ME. Adaptation and Maladaptation of the Heart in Diabetes: Part I: General Concepts. *Circulation* 2002;105:1727-1733.

**101.** Stanley WC, Lopaschuk GD, McCormack JG. Regulation of energy substrate metabolism in the diabetic heart. *Cardiovascular Research* 1997;34:25-33.

**102.** Aasum E, Hafstad AD, Severson DL, Larsen TS. Age-Dependent Changes in Metabolism, Contractile Function, and Ischemic Sensitivity in Hearts From db/db Mice. *Diabetes* 2003;52:434-441.

**103.** King LM, Sidell RJ, Wilding JR, Radda GK, Clarke K. Free fatty acids, but not ketone bodies, protect diabetic rat hearts during low-flow ischemia. *Am J Physiol Heart Circ Physiol* 2001;280:H1173-1181.

**104.** Veech RL, Chance, B., Kashiwaya, Y., Lardy, H.A., Cahill, G.F. Jr. Ketone bodies, potential therapeutic uses. *IUBMB Life* 2001;51:241-247.

**105.** Casteels K, Mathieu C. Diabetic ketoacidosis. *Rev Endocr Metab Disord* 2003;4:159-66.

**106.** Apstein CS. Glucose-Insulin-Potassium for Acute Myocardial Infarction : Remarkable Results From a New Prospective, Randomized Tria. *Circulation* 1998;98:2223-2226.

**107.** Sato K, Kashiwaya Y, Keon C, Tsuchiya N, King M, Radda G, Chance B, Clarke K, Veech R. Insulin, ketone bodies, and mitochondrial energy transduction. *FASEB J.* 1995;9:651-658.

**108.** Stanley WC, Lopaschuk GD, Hall JL, McCormack JG. Regulation of myocardial carbohydrate metabolism under normal and ischaemic conditions: Potential for pharmacological interventions. *Cardiovascular Research* 1997;33:243-257.

**109.** Indiveri C, Tonazzi A, Palmieri F. Characterization of the unidirectional transport of carnitine catalyzed by the reconstituted carnitine carrier from rat liver mitochondria. *Biochim Biophys Acta* 1991;1069:110-6.

**110.** Murthy MS, Pande SV. Mechanism of carnitine acylcarnitine translocase-catalyzed import of acylcarnitines into mitochondria. *J Biol Chem* 1984;259:9082-9.

**111.** Williamson JR, Hoek JB, Murphy E, Coll KE, Njogu RM. Kinetics and mechanisms of glutamate transport across the mitochondrial membrane. *Ann N Y Acad Sci* 1980;341:593-608.

**112.** Bunger R, Mallet RT. Mitochondrial pyruvate transport in working guinea-pig heart. Work-related vs. carrier-mediated control of pyruvate oxidation. *Biochim Biophys Acta* 1993;1151:223-36.

**113.** Palmieri L, Pardo B, Lasorsa FM, del Arco A, Kobayashi K, Iijima M, Runswick MJ, Walker JE, Saheki T, Satrustegui J, Palmieri F. Citrin and aralar1 are $Ca^{2+}$-stimulated aspartate/glutamate transporters in mitochondria. *Embo J* 2001;20:5060-9.

**114.** Claeys D, Azzi A. Tricarboxylate carrier of bovine liver mitochondria. Purification and reconstitution. *J Biol Chem* 1989;264:14627-30.

**115.** Fiermonte G, Dolce V, Arrigoni R, Runswick MJ, Walker JE, Palmieri F. Organization and sequence of the gene for the human mitochondrial dicarboxylate carrier: evolution of the carrier family. *Biochem J* 1999;344 Pt 3:953-60.

**116.** Fiermonte G, Dolce V, David L, Santorelli FM, Dionisi-Vici C, Palmieri F, Walker JE. The mitochondrial ornithine transporter. Bacterial expression, reconstitution, functional characterization, and tissue distribution of two human isoforms. *J Biol Chem* 2003;278:32778-83.

**117.** De Marcos Lousa C, Trezeguet V, Dianoux AC, Brandolin G, Lauquin GJ. The human mitochondrial ADP/ATP carriers: kinetic properties and biogenesis of wild-type and mutant proteins in the yeast *S. cerevisiae*. *Biochemistry* 2002;41:14412-20.

**118.** Kofoed KF, Carstensen S, Hove JD, Freiberg J, Bangsgaard R, Holm S, Rabol A, Hesse B, Arendrup H, Kelbaek H. Low whole-body insulin sensitivity in patients with ischaemic heart disease is associated with impaired myocardial glucose uptake predictive of poor outcome after revascularisation. *Eur J Nucl Med Mol Imaging* 2002;29:991-8.

**119.** Paternostro G, Pagano D, Gnecchi-Ruscone T, Bonser RS, Camici PG. Insulin resistance in patients with cardiac hypertrophy. *Cardiovasc Res* 1999;42:246-53.

**120.** Knuuti MJ, Maki M, Yki-Jarvinen H, Voipio-Pulkki LM, Harkonen R, Haaparanta M, Nuutila P. The effect of insulin and FFA on myocardial glucose uptake. *J Mol Cell Cardiol* 1995;27:1359-67.

**121.** Iozzo P, Chareonthaitawee P, Rimoldi O, Betteridge DJ, Camici PG, Ferrannini E. Mismatch between insulin-mediated glucose uptake and blood flow in the heart of patients with Type II diabetes. *Diabetologia* 2002;45:1404-9.

**122.** Goodwin GW, Ahmad F, Doenst T, Taegtmeyer H. Energy provision from glycogen, glucose, and fatty acids on adrenergic stimulation of isolated working rat hearts. *Am J Physiol* 1998;274:H1239-47.

**123.** Bittl JA, Weisfeldt ML, Jacobus WE. Creatine kinase of heart mitochondria. The progressive loss of enzyme activity during in vivo ischemia and its correlation to depressed myocardial function. *J Biol Chem* 1985;260:208-14.

**124.** Sharov VG, Todor AV, Silverman N, Goldstein S, Sabbah HN. Abnormal mitochondrial respiration in failed human myocardium. *J Mol Cell Cardiol* 2000;32:2361-7.

**125.** Heo K, Lin X, Odle J, Han IK. Kinetics of carnitine palmitoyltransferase-I are altered by dietary variables and suggest a metabolic need for supplemental carnitine in young pigs. *J Nutr* 2000;130:2467-70.

**126.** Osborn BA, Daar JT, Laddaga RA, Romano FD, Paulson DJ. Exercise training increases sarcolemmal GLUT-4 protein and mRNA content in diabetic heart. *J Appl Physiol* 1997;82:828-34.

**127.** Belke DD, Larsen TS, Gibbs EM, Severson DL. Altered metabolism causes cardiac dysfunction in perfused hearts from diabetic (db/db) mice. *Am J Physiol Endocrinol Metab* 2000;279:E1104-13.

**128.** Schonekess BO. Competition between lactate and fatty acids as sources of ATP in the isolated working rat heart. *J Mol Cell Cardiol* 1997;29:2725-33.

**129.** Pallotti F, Lenaz G. Isolation and subfractionation of mitochondria from animal cells and tissue culture lines. *Methods Cell Biol* 2001;65:1-35.

**130.** Niezen-Koning KE, Wanders RJ, Ruiter JP, Ijlst L, Visser G, Reitsma-Bierens WC, Heymans HS, Reijngoud DJ, Smit GP. Succinyl-CoA:acetoacetate transferase deficiency: identification of a new patient with a neonatal onset and review of the literature. *Eur J Pediatr* 1997;156:870-3.

**131.** Kantor PF, Lucien A, Kozak R, Lopaschuk GD. The antianginal drug trimetazidine shifts cardiac energy metabolism from fatty acid oxidation to glucose oxidation by inhibiting mitochondrial long-chain 3-ketoacyl coenzyme A thiolase. *Circ Res* 2000;86:580-8.

**132.** Halestrap AP. The mitochondrial pyruvate carrier. Kinetics and specificity for substrates and inhibitors. *Biochem J* 1975;148:85-96.

**133.** Birkemeyer C, Luedemann A, Wagner C, Erban A, Kopka J. Metabolome analysis: the potential of *in vivo* labeling with stable isotopes for metabolite profiling. *Trends Biotechnol* 2005;23:28-33.

**134.** Ibarra RU, Fu P, Palsson BO, DiTonno JR, Edwards JS. Quantitative analysis of *Escherichia coli* metabolic phenotypes within the context of phenotypic phase planes. *J Mol Microbiol Biotechnol* 2003;6:101-8.

**135.** Winkler BS, Sauer MW, Starnes CA. Effects of L-glutamate/D-aspartate and monensin on lactic acid production in retina and cultured retinal Muller cells. *J Neurochem* 2004;89:514-25.

**136.** Khairallah M, Labarthe F, Bouchard B, Danialou G, Petrof BJ, Des Rosiers C. Profiling substrate fluxes in the isolated working mouse heart using $^{13}$C-labeled substrates: focusing on the origin and fate of pyruvate and citrate carbons. *Am J Physiol Heart Circ Physiol* 2004;286:H1461-70.

**137.** Lee WN, Go VL. Nutrient-gene interaction: tracer-based metabolomics. *J Nutr* 2005;135:3027S-3032S.

**138.** Radziuk J, Lee WP. Measurement of gluconeogenesis and mass isotopomer analysis based on [U-$^{13}$C]glucose. *Am J Physiol* 1999;277:E199-207.

**139.** Lee WN, Edmond J, Bassilian S, Morrow JW. Mass isotopomer study of glutamine oxidation and synthesis in primary culture of astrocytes. *Dev Neurosci* 1996;18:469-77.

**140.** Cline GW, Lepine RL, Papas KK, Kibbey RG, Shulman GI. 13C NMR isotopomer analysis of anaplerotic pathways in INS-1 cells. *J Biol Chem* 2004;279:44370-5.

**141.** Malloy CR, Jones JG, Jeffrey FM, Jessen ME, Sherry AD. Contribution of various substrates to total citric acid cycle flux and anaplerosis as determined by 13C isotopomer analysis and $O_2$ consumption in the heart. *Magma* 1996;4:35-46.

**142.** Cohen DM, Bergman RN. Improved estimation of anaplerosis in heart using $^{13}C$ NMR. *Am J Physiol* 1997;273:E1228-42.

**143.** Katz J, Tayek JA. Recycling of glucose and determination of the Cori Cycle and gluconeogenesis. *Am J Physiol* 1999;277:E401-7.

**144.** Bederman IR, Reszko AE, Kasumov T, David F, Wasserman DH, Kelleher JK, Brunengraber H. Zonation of labeling of lipogenic acetyl-CoA across the liver: implications for studies of lipogenesis by mass isotopomer analysis. *J Biol Chem* 2004;279:43207-16.

**145.** Lee WNP. Characterizing phenotype with tracer based metabolomics. *Metabolomics* 2006;1.

**146.** Zupke C, Stephanopoulos G. Modeling of Isotope Distributions and Intracellular Fluxes in Metabolic Networks Using Atom Mapping Matrices. *Biotechnology Progress* 1994;10:489-498.

**147.** Schmidt K, Carlsen M, Nielsen J, Villadsen J. Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices. *Biotechnology and Bioengineering* 1997;55:831-840.

**148.** van Winden WA, Heijnen JJ, Verheijen PJ. Cumulative bondomers: a new concept in flux analysis from 2D [$^{13}C$,$^{1}H$] COSY NMR data. *Biotechnol Bioeng* 2002;80:731-45.

**149.** Forbes NS, Clark DS, Blanch HW. Using isotopomer path tracing to quantify metabolic fluxes in pathway models containing reversible reactions. *Biotechnol Bioeng* 2001;74:196-211.

**150.** Wiechert W, Mollney M, Isermann N, Wurzel M, de Graaf AA. Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. *Biotechnol Bioeng* 1999;66:69-85.

**151.** Fischer E, Sauer U. Metabolic flux profiling of Escherichia coli mutants in central carbon metabolism using GC-MS. *Eur J Biochem* 2003;270:880-91.

**152.** Schmidt K, Nielsen J, Villadsen J. Quantitative analysis of metabolic fluxes in *Escherichia coli*, using two-dimensional NMR spectroscopy and complete isotopomer models. *J Biotechnol* 1999;71:175-89.

**153.** Dauner M, Bailey JE, Sauer U. Metabolic flux analysis with a comprehensive isotopomer model in *Bacillus subtilis*. *Biotechnol Bioeng* 2001;76:144-56.

**154.** Van Dien SJ, Strovas T, Lidstrom ME. Quantification of central metabolic fluxes in the facultative methylotroph *Methylobacterium extorquens* AM1 using $^{13}C$-label tracing and mass spectrometry. *Biotechnol Bioeng* 2003;84:45-55.

**155.** Gombert AK, Moreira dos Santos M, Christensen B, Nielsen J. Network identification and flux quantification in the central metabolism of *Saccharomyces cerevisiae* under different conditions of glucose repression. *Journal of Bacteriology* 2001;183:1441-51.

**156.** Christensen B, Gombert AK, Nielsen J. Analysis of flux estimates based on $^{13}$C-labelling experiments. *Eur J Biochem* 2002;269:2795-800.

**157.** Christensen B, Nielsen J. Metabolic network analysis of *Penicillium chrysogenum* using $^{13}$C-labeled glucose. *Biotechnol Bioeng* 2000;68:652-9.

**158.** van Winden WA, van Gulik WM, Schipper D, Verheijen PJ, Krabben P, Vinke JL, Heijnen JJ. Metabolic flux and metabolic network analysis of *Penicillium chrysogenum* using 2D [$^{13}$C, $^{1}$H] COSY NMR measurements and cumulative bondomer simulation. *Biotechnol Bioeng* 2003;83:75-92.

**159.** Marx A, deGraaf AA, Wiechert W, Eggeling L, Sahm H. Determination of the fluxes in the central metabolism of *Corynebacterium glutamicum* by nuclear magnetic resonance spectroscopy combined with metabolite balancing. *Biotechnology and Bioengineering* 1996;49:111-129.

**160.** Klapa MI, Aon JC, Stephanopoulos G. Systematic quantification of complex metabolic flux networks using stable isotopes and mass spectrometry. *European Journal of Biochemistry* 2003;270:3525-3542.

**161.** Wiechert W, de Graaf AA. Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments. *Biotechnol Bioeng* 1997;55:101-117.

**162.** van Winden WA, Heijnen JJ, Verheijen PJT, Grievink J. A priori analysis of metabolic flux identifiability from $^{13}$C-labeling data. *Biotechnology and Bioengineering* 2001;74:505-516.

**163.** Lee WNP, Byerley LO, Bergner EA, Edmond J. Mass Isotopomer Analysis - Theoretical and Practical Considerations. *Biological Mass Spectrometry* 1991;20:451-458.

**164.** Vo TD, Palsson BO. Isotopomer analysis of myocardial substrate metabolism: A systems biology approach. *Biotechnol Bioeng* 2006.

**165.** Yarmush M, Berthiaume F. Metabolic engineering and human disease. *Nature Biotechnology* 1997;15:525-528.

**166.** Nielsen J. It is all about metabolic fluxes. *J Bacteriol* 2003;185:7031-5.

**167.** Wittmann C, Heinzle E. Mass spectrometry for metabolic flux analysis. *Biotechnology and Bioengineering* 1999;62:739-750.

**168.** Wiechert W, Siefke C, deGraaf AA, Marx A. Bidirectional reaction steps in metabolic networks .2. Flux estimation and statistical analysis. *Biotechnology and Bioengineering* 1997;55:118-135.

**169.** Schmidt K, Marx A, de Graaf AA, Wiechert W, Sahm H, Nielsen J, Villadsen J. $^{13}C$ tracer experiments and metabolite balancing for metabolic flux analysis: Comparing two approaches. *Biotechnology and Bioengineering* 1998;58:254-257.

**170.** Comte B, Vincent G, Bouchard B, Jette M, Cordeau S, Rosiers CD. A $^{13}C$ mass isotopomer study of anaplerotic pyruvate carboxylation in perfused rat hearts. *J Biol Chem* 1997;272:26125-31.

**171.** Panchal AR, Comte B, Huang H, Kerwin T, Darvish A, des Rosiers C, Brunengraber H, Stanley WC. Partitioning of pyruvate between oxidation and anaplerosis in swine hearts. *Am J Physiol Heart Circ Physiol* 2000;279:H2390-8.

**172.** Sherry AD, Jeffrey FM, Malloy CR. Analytical solutions for $^{13}C$ isotopomer analysis of complex metabolic conditions: substrate oxidation, multiple pyruvate cycles, and gluconeogenesis. *Metab Eng* 2004;6:12-24.

**173.** Haymond MW, Sunehag AL. The reciprocal pool model for the measurement of gluconeogenesis by use of [U-$^{13}C$]glucose. *Am J Physiol Endocrinol Metab* 2000;278:E140-5.

**174.** Puchowicz MA, Bederman IR, Comte B, Yang D, David F, Stone E, Jabbour K, Wasserman DH, Brunengraber H. Zonation of acetate labeling across the liver: implications for studies of lipogenesis by MIDA. *Am J Physiol* 1999;277:E1022-7.

**175.** Comte B, Vincent G, Bouchard B, Des Rosiers C. Probing the origin of acetyl-CoA and oxaloacetate entering the citric acid cycle from the $^{13}C$ labeling of citrate released by perfused rat hearts. *J Biol Chem* 1997;272:26117-24.

**176.** Fernandez CA, Des Rosiers C. Modeling of liver citric acid cycle and gluconeogenesis based on $^{13}C$ mass isotopomer distribution analysis of intermediates. *J Biol Chem* 1995;270:10037-42.

**177.** Katz J, Lee WN, Wals PA, Bergner EA. Studies of glycogen synthesis and the Krebs cycle by mass isotopomer analysis with [U-$^{13}C$]glucose in rats. *J Biol Chem* 1989;264:12994-3004.

**178.** Klapa MI, Park SM, Sinskey AJ, Stephanopoulos G. Metabolite and isotopomer balancing in the analysis of metabolic cycles: I. Theory. *Biotechnol Bioeng* 1999;62:375-391.

**179.** Schilling CH, Letscher D, Palsson BO. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol* 2000;203:229-48.

**180.** Fernandez CA, Des Rosiers C, Previs SF, David F, Brunengraber H. Correction of $^{13}C$ mass isotopomer distributions for natural stable isotope abundance. *J Mass Spectrom* 1996;31:255-62.

**181.** Gill PE, Murray W, Saunders Ma. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM J. Optim* 2002;12:979-1006.

**182.** Vincent G, Bouchard B, Khairallah M, Des Rosiers C. Differential modulation of citrate synthesis and release by fatty acids in perfused working rat hearts. *Am J Physiol Heart Circ Physiol* 2004;286:H257-66.

**183.** Wiechert W, deGraaf AA. Bidirectional reaction steps in metabolic networks .1. Modeling and simulation of carbon isotope labeling experiments. *Biotechnology and Bioengineering* 1997;55:101-117.

**184.** Fink G, Desrochers S, Des Rosiers C, Garneau M, David F, Daloze T, Landau BR, Brunengraber H. Pseudoketogenesis in the perfused rat heart. *J Biol Chem* 1988;263:18036-42.

**185.** Kaufmann DE, Smith, R.L. Direction Choice for Accelerated Convergence in Hit-and-Run Sampling. *Operations Research* 1998;46:84-95.

**186.** Reed J. Model driven analysis of *Escherichia coli* metabolism. University of California, San Diego: Ph.D. dissertation, 2005:189.

**187.** DeGrella RF, Light RJ. Uptake and metabolism of fatty acids by dispersed adult rat heart myocytes. I. Kinetics of homologous fatty acids. *J Biol Chem* 1980;255:9731-8.

**188.** Lee WN, Boros LG, Puigjaner J, Bassilian S, Lim S, Cascante M. Mass isotopomer study of the nonoxidative pathways of the pentose cycle with [1,2-$^{13}C_2$]glucose. *Am J Physiol* 1998;274:E843-51.

**189.** Tserng KY, Gilfillan CA, Kalhan SC. Determination of $^{13}C$ labeled lactate in blood by gas chromatography/mass spectrometry. *Anal Chem* 1984;56:517-23.

**190.** Lowenstein JM, Brunengraber H, Wadke M. Measurement of rates of lipogenesis with deuterated and tritiated water. *Methods Enzymol* 1975;35:279-87.

**191.** Lee WN, Bassilian S, Guo Z, Schoeller D, Edmond J, Bergner EA, Byerley LO. Measurement of fractional lipid synthesis using deuterated water ($^2H_2O$) and mass isotopomer analysis. *Am J Physiol* 1994;266:E372-83.

**192.** Marin S, Lee WN, Bassilian S, Lim S, Boros LG, Centelles JJ, FernAndez-Novell JM, Guinovart JJ, Cascante M. Dynamic profiling of the glucose metabolic network in fasted rat hepatocytes using [1,2-$^{13}C_2$]glucose. *Biochem J* 2004;381:287-94.

**193.** Leimer KR, Rice RH, Gehrke CW. Complete mass spectra of N-trifluoroacetyl-n-butyl esters of amino acids. *J Chromatogr* 1977;141:121-44.

**194.** Dehn PF, White CM, Conners DE, Shipkey G, Cumbo TA. Characterization of the human hepatocellular carcinoma (hepg2) cell line as an in vitro model for cadmium toxicity studies. *In Vitro Cell Dev Biol Anim* 2004;40:172-82.

**195.** Du QY, Wang XB, Chen XJ, Zheng W, Wang SQ. Antitumor mechanism of antisense cantide targeting human telomerase reverse transcriptase. *World J Gastroenterol* 2003;9:2030-5.

**196.** Lee WN, Lim S, Bassilian S, Bergner EA, Edmond J. Fatty acid cycling in human hepatoma cells and the effects of troglitazone. *J Biol Chem* 1998;273:20929-34.

**197.** Kelleher JK. Gluconeogenesis from labeled carbon: estimating isotope dilution. *Am J Physiol* 1986;250:E296-305.

**198.** Lee WN, Bergner EA, Guo ZK. Mass isotopomer pattern and precursor-product relationship. *Biol Mass Spectrom* 1992;21:114-22.

**199.** Lee WN, Byerley LO, Bassilian S, Ajie HO, Clark I, Edmond J, Bergner EA. Isotopomer study of lipogenesis in human hepatoma cells in culture: contribution of carbon and hydrogen atoms from glucose. *Anal Biochem* 1995;226:100-12.

**200.** Lee WN, Bassilian S, Ajie HO, Schoeller DA, Edmond J, Bergner EA, Byerley LO. In vivo measurement of fatty acids and cholesterol synthesis using $D_2O$ and mass isotopomer analysis. *Am J Physiol* 1994;266:E699-708.

**201.** Brunengraber H, Kelleher JK, Des Rosiers C. Applications of mass isotopomer analysis to nutrition research. *Annu Rev Nutr* 1997;17:559-96.

**202.** Owen OE, Kalhan SC, Hanson RW. The key role of anaplerosis and cataplerosis for citric acid cycle function. *J Biol Chem* 2002;277:30409-12.

**203.** Duerden JM, Gibbons GF. Secretion and storage of newly synthesized hepatic triacylglycerol fatty acids in vivo in different nutritional states and in diabetes. *Biochem J* 1988;255:929-35.

**204.** Hellerstein MK, Christiansen M, Kaempfer S, Kletke C, Wu K, Reid JS, Mulligan K, Hellerstein NS, Shackleton CH. Measurement of de novo hepatic lipogenesis in humans using stable isotopes. *J Clin Invest* 1991;87:1841-52.

**205.** Zielke HR, Zielke CL, Ozand PT. Glutamine: a major energy source for cultured mammalian cells. *Fed Proc* 1984;43:121-5.

**206.** Bettger WJ, Ham RG. The nutrient requirements of cultured mammalian cells. *Adv Nutr Res* 1982;4:249-86.

**207.** Kilberg MS, Handlogten ME, Christensen HN. Characteristics of an amino acid transport system in rat liver for glutamine, asparagine, histidine, and closely related analogs. *J Biol Chem* 1980;255:4011-9.

**208.** Mailliard ME, Kilberg MS. Sodium-dependent neutral amino acid transport by human liver plasma membrane vesicles. *J Biol Chem* 1990;265:14321-6.

**209.** Pawlik TM, Souba WW, Bode BP. Asparagine uptake in rat hepatocytes: resolution of a paradox and insights into substrate-dependent transporter regulation. *Amino Acids* 2001;20:335-52.

**210.** Lanks KW, Li PW. End products of glucose and glutamine metabolism by cultured cell lines. *J Cell Physiol* 1988;135:151-5.

**211.** Leigh D. Subacute necrotizing encephalomyelopathy in an infant. *J Neurol Neurosurg Psychiatry* 1951;14:216-21.

**212.** Dahl HH. Getting to the nucleus of mitochondrial disorders: identification of respiratory chain-enzyme genes causing Leigh syndrome. *Am J Hum Genet* 1998;63:1594-7.

**213.** Mannan AA, Sharma MC, Shrivastava P, Ralte AM, Gupta V, Behari M, Sarkar C. Leigh's syndrome. *Indian J Pediatr* 2004;71:1029-33.

**214.** Rahman S, Blok RB, Dahl HH, Danks DM, Kirby DM, Chow CW, Christodoulou J, Thorburn DR. Leigh syndrome: clinical features and biochemical and DNA abnormalities. *Ann Neurol* 1996;39:343-51.

**215.** Fernie AR, Trethewey RN, Krotzky AJ, Willmitzer L. Metabolite profiling: from diagnostics to systems biology. *Nat Rev Mol Cell Biol* 2004;5:763-9.

**216.** Reed JL. Model driven analysis of *Escherichia coli* metabolism Bioengineering. La Jolla: University of California, San Diego, 2005.

**217.** Vo TD, Palsson BO. Isotopomer analysis of myocardial substrate metabolism: A systems biology approach. *Biotechnol Bioeng* 2006;95:972-83.

**218.** Laderman KA, Penny JR, Mazzucchelli F, Bresolin N, Scarlato G, Attardi G. Aging-dependent functional alterations of mitochondrial DNA (mtDNA) from human fibroblasts transferred into mtDNA-less cells. *J Biol Chem* 1996;271:15891-7.

**219.** Sheikh K, Forster J, Nielsen LK. Modeling hybridoma cell metabolism using a generic genome-scale metabolic model of Mus musculus. *Biotechnology Progress* 2005;21:112-121.

**220.** Sorbi S, Blass JP. Abnormal activation of pyruvate dehydrogenase in Leigh disease fibroblasts. *Neurology* 1982;32:555-8.

**221.** Constantopoulos G, Greenwood MA, Sorrell SH. Mitochondrial abnormalities in fibroblast line GM3093 defective in oxidative metabolism. *Experientia* 1986;42:315-8.

**222.** Lie SO, McKusick VA, Neufeld EF. Simulation of genetic mucopolysaccharidoses in normal human fibroblasts by alteration of pH of the medium. *Proc Natl Acad Sci U S A* 1972;69:2361-3.

**223.** Chretien D, Benit P, Chol M, Lebon S, Rotig A, Munnich A, Rustin P. Assay of mitochondrial respiratory chain complex I in human lymphocytes and cultured skin fibroblasts. *Biochem Biophys Res Commun* 2003;301:222-4.

**224.** Blass JP, Schulman JD, Young DS, Hom E. An inherited defect affecting the tricarboxylic acid cycle in a patient with congenital lactic acidosis. *J Clin Invest* 1972;51:1845-51.

**225.** Hansen TL, Christensen E. Studies on pyruvate carboxylase from cultured human fibroblasts and amniotic fluid cells. *J Inherit Metab Dis* 1980;2:23-8.

**226.** Martin MA, Blazquez A, Gutierrez-Solana LG, Fernandez-Moreira D, Briones P, Andreu AL, Garesse R, Campos Y, Arenas J. Leigh syndrome associated with mitochondrial complex I deficiency due to a novel mutation in the NDUFS1 gene. *Arch Neurol* 2005;62:659-61.

**227.** Cameron JM, Levandovskiy V, MacKay N, Robinson BH. Respiratory chain analysis of skin fibroblasts in mitochondrial disease. *Mitochondrion* 2004;4:387-94.

**228.** Huh TL, Casazza JP, Huh JW, Chi YT, Song BJ. Characterization of two cDNA clones for pyruvate dehydrogenase E1 beta subunit and its regulation in tricarboxylic acid cycle-deficient fibroblast. *J Biol Chem* 1990;265:13320-6.

**229.** Carrozzo R, Tessa A, Vazquez-Memije ME, Piemonte F, Patrono C, Malandrini A, Dionisi-Vici C, Vilarinho L, Villanova M, Schagger H, Federico A, Bertini E, Santorelli FM. The T9176G mtDNA mutation severely affects ATP production and results in Leigh syndrome. *Neurology* 2001;56:687-90.

**230.** Tulinius MH, Houshmand M, Larsson NG, Holme E, Oldfors A, Holmberg E, Wahlstrom J. De novo mutation in the mitochondrial ATP synthase subunit 6 gene (T8993G) with rapid segregation resulting in Leigh syndrome in the offspring. *Hum Genet* 1995;96:290-4.

**231.** Sgarbi G, Baracca A, Lenaz G, Valentino LM, Carelli V, Solaini G. Inefficient coupling between proton transport and ATP synthesis may be the pathogenic mechanism for NARP and Leigh syndrome resulting from the T8993G mutation in mtDNA. *Biochem J* 2006;395:493-500.

**232.** Stryer L. Biochemistry. New York: W.H. Freeman, 1995:xxxiv, 1064 p.

**233.** Chretien D, Rustin P, Bourgeron T, Rotig A, Saudubray JM, Munnich A. Reference charts for respiratory chain activities in human tissues. *Clin Chim Acta* 1994;228:53-70.

**234.** Morris AA, Leonard JV, Brown GK, Bidouki SK, Bindoff LA, Woodward CE, Harding AE, Lake BD, Harding BN, Farrell MA, Bell JE, Mirakhur M, Turnbull DM. Deficiency of respiratory chain complex I is a common cause of Leigh disease. *Ann Neurol* 1996;40:25-30.

**235.** Benit P, Slama A, Cartault F, Giurgea I, Chretien D, Lebon S, Marsac C, Munnich A, Rotig A, Rustin P. Mutant NDUFS3 subunit of mitochondrial complex I causes Leigh syndrome. *J Med Genet* 2004;41:14-7.

**236.** Rustin P, Chretien D, Bourgeron T, Wucher A, Saudubray JM, Rotig A, Munnich A. Assessment of the mitochondrial respiratory chain. *Lancet* 1991;338:60.

**237.** Rustin P, Chretien D, Bourgeron T, Gerard B, Rotig A, Saudubray JM, Munnich A. Biochemical and molecular investigations in respiratory chain deficiencies. *Clin Chim Acta* 1994;228:35-51.

**238.** Chretien D, Gallego J, Barrientos A, Casademont J, Cardellach F, Munnich A, Rotig A, Rustin P. Biochemical parameters for the diagnosis of mitochondrial respiratory chain deficiency in humans, and their lack of age-related changes. *Biochem J* 1998;329 ( Pt 2):249-54.

**239.** Mahoney DJ, Parise G, Tarnopolsky MA. Nutritional and exercise-based therapies in the treatment of mitochondrial disease. *Curr Opin Clin Nutr Metab Care* 2002;5:619-29.

**240.** Wijburg FA, Barth PG, Bindoff LA, Birch-Machin MA, van der Blij JF, Ruitenbeek W, Turnbull DM, Schutgens RB. Leigh syndrome associated with a deficiency of the pyruvate dehydrogenase complex: results of treatment with a ketogenic diet. *Neuropediatrics* 1992;23:147-52.

**241.** Di Rocco M, Lamba LD, Minniti G, Caruso U, Naito E. Outcome of thiamine treatment in a child with Leigh disease due to thiamine-responsive pyruvate dehydrogenase deficiency. *Eur J Paediatr Neurol* 2000;4:115-7.

**242.** Tarnopolsky M, Martin J. Creatine monohydrate increases strength in patients with neuromuscular disease. *Neurology* 1999;52:854-7.

**243.** Tarnopolsky MA, Roy BD, MacDonald JR. A randomized, controlled trial of creatine monohydrate in patients with mitochondrial cytopathies. *Muscle Nerve* 1997;20:1502-9.

**244.** Goldenberg PC, Steiner RD, Merkens LS, Dunaway T, Egan RA, Zimmerman EA, Nesbit G, Robinson B, Kennaway NG. Remarkable improvement in adult Leigh syndrome with partial cytochrome c oxidase deficiency. *Neurology* 2003;60:865-8.

**245.** Joyce AR, Palsson BO. The model organism as a system: integrating '*omics*' data sets. *Nat Rev Mol Cell Biol* 2006;7:198-210.

**246.** Rochfort S. Metabolomics reviewed: a new '*omics'* platform technology for systems biology and implications for natural products research. *J Nat Prod* 2005;68:1813-20.

**247.** Kiechle FL, Zhang X, Holland-Staley CA. The -*omics* era and its impact. *Arch Pathol Lab Med* 2004;128:1337-45.

**248.** Abu-Issa R, Kirby ML. Take heart in the age of '*omics*'. *Circ Res* 2004;95:335-6.