THESIS

FLUX BALANCE ANALYSIS OF METABOLIC MODELS: A REVIEW OF RECENT ADVANCES AND APPLICATIONS

Submitted by

Forrest Earnest Estep

Department of Chemical and Biological Engineering

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2016

Master's Committee:

Advisor: Christie Peebles

Patrick Shipman
Chris Snow

ABSTRACT

FLUX BALANCE ANALYSIS OF METABOLIC MODELS: A REVIEW OF RECENT ADVANCES AND APPLICATIONS

Genome-level reconstructions of metabolic networks have provided new insight into the cellular functions of many organisms. These metabolic models are massive constructs, often including thousands of metabolic and transport reactions and metabolite species for even the most basic organisms. Construction of these models has typically involved an initial genomic analysis to identify known genes or genes with homologous structures for which the function may be inferred, followed by an intensive process of literature searching and experimental validation to refine the model. A number of automated algorithms have been developed to assist with this process. Once the model has been constructed, optimization techniques are applied to predict the distribution of fluxes through the reaction network. The systems then studied by FBA are generally static systems, assumed to be operating at a steady state, and thus constrained by the stoichiometries of the reactions rather than the kinetics. While these assumptions have shown to be valid under select laboratory conditions, evidence indicates that most organisms are not always at this steady state. A number of model improvements have been considered to bring predicted results more in line with experimental data, including the addition of regulatory controls, more detailed incorporation of thermodynamics, and the consideration of metabolite pool and flux data from metabolomics and labeled carbon studies, respectively. The improved predictive capabilities of these models readily find application in metabolic engineering in the custom strain design of organisms. Often this purpose is the production of some valuable

bioproduct. This review seeks to give overview the advances made on both the model construction and application ends, with particular emphasis on model improvements via more complex constraints and the incorporation of experimental data.

TABLE OF CONTENTS

INTRODUCTION

The metabolic systems of living organisms are quite complicated, being composed of a myriad of enzymes capable of catalyzing a number of different reactions. The increased use of bioinformatics to analyze complete genomes is allowing biologists to begin developing full system models of many organisms' metabolic networks. These models seek to include all chemical reactions occurring within the organism, their compartmentalized location within the cellular organelles, and the transport reactions both between intracellular compartments and between the organism and its environment. With over 1,200 metabolic and transport reactions occurring in simple unicellular organisms such as *E. coli* [1] , transient modeling of these systems is generally too computationally intensive to be feasible. Further, it is difficult to determine the *in vivo* reaction rates with much confidence, and intracellular metabolites are often present at such low concentrations that their accurate measurement is equally questionable. Flux Balance Analysis (FBA) is a modeling technique that helps to work around these issues.

CHAPTER 1 – WHAT IS FLUX BALANCE ANALYSIS?

The Basic Assumptions

A general overview of how a metabolic network is analyzed using FBA is illustrated in Figure 1 [2]. A set of ordinary differential equations describing cellular metabolism can be formed from the knowledge of the reactions involved. These equations take the following form:

$$\frac{dC_i}{dt} = \sum_j^n a_i v_j = \sum_j^n k_j C_i^{a_i} \prod_k^m C_k^{a_k}$$

Here, the time dependent concentration of species *i* is the sum of the n associated reactions' fluxes ($a_i v_i$), defined as the product of the reaction rate constant k_j and the *m* stoichiometrically weighted species also involved in the reaction. These equations involve a number of difficult to measure kinetic constants [3], and the complexity of these systems would require a computational solution.

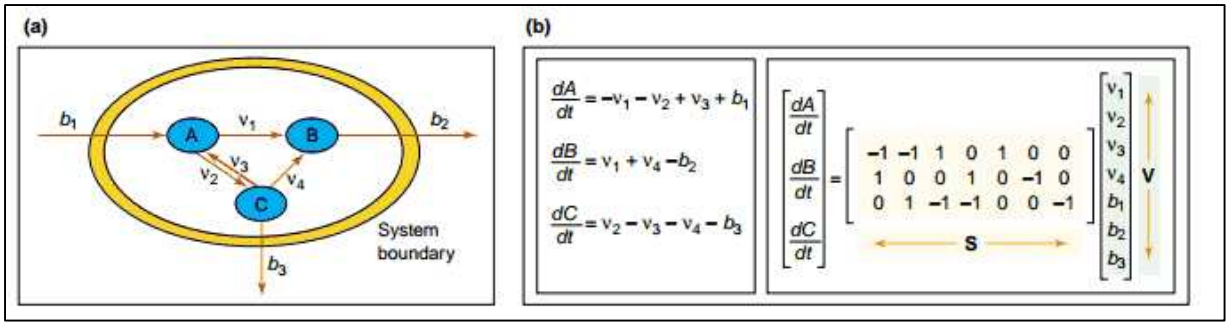

Figure 1: Constructing a metabolic model by (a) formulating a representation for the metabolic network and (b) converting the network into a set of differential equations

Flux balance analysis simplifies this problem by assuming that nature has over time optimized life towards some evolutionary goal. Further, it is assumed that at this optimum an organism behaves as if it were at steady state. This leads to the assumption that the internal metabolites, when in constant environmental conditions, will be at a pseudo steady state, being

produced and consumed at approximately the same rate [2]. The distribution of metabolites in this system is then only constrained by the stoichiometries of the reactions.

This reduces the set of transient ordinary differential equations to a set of linear algebraic equations, the solution of which is well studied. This simplified set can be represented in matrix form as $\mathbf{Sv}=0$. Here, $\mathbf{S}$ is a matrix of the stoichiometric reaction coefficients, and $\mathbf{v}$ is a vector of the associated reaction fluxes [2]. A "growth equation" is included in this set of reactions to represent the final consumption of metabolites necessary for cell replication. Additional equations are included to represent the uptake of nutrients and the excretion of certain molecules to the extracellular space. These equations together constrain the flux of metabolites through the network of reactions [4]. As there are many more reactions than there are metabolites, the system of equations is said to be underdetermined. To solve this system for a unique solution requires measuring enough of the specific metabolite fluxes to constrain the system [2]. It is difficult to accurately measure these low concentration metabolites, so the assumption that life is optimized towards some goal is applied again.

Flux Balance Analysis, Model Solution by Optimization

This correctly implies that the mathematical techniques of optimization can be used to determine a unique solution. Optimization seeks to maximize (or minimize) an *objective function* by manipulating certain *variables* within the bounds of some *constraints*. The desired information is often both the optimum value of the objective function and the values of the variables that define the optimum state.

The metabolic model is easily translated to this framework. Written explicitly in optimization terms, the general problem appears as follows:

$$\max_{\boldsymbol{v}} \boldsymbol{c}^T \boldsymbol{v}$$

$$s.t. \quad \boldsymbol{S} \cdot \boldsymbol{v} = \mathbf{0}$$

$$v_i \geq 0$$

$$v_i \leq UB$$

$$b_i \geq LB$$

$$b_i \leq UB$$

The variables solved for are the fluxes through each metabolic and transport reaction. Constraints take the form of both equality and inequality equations. The stoichiometric limitations (**S**) of the reaction network are represented as equality constraints. This includes all metabolic reactions and the growth reaction. Inequality constraints can be used to represent feasible upper and lower bounds on the system, or to limit specific reaction fluxes. Examples of these constraints include a maximum feasible substrate uptake rate or a minimum flux through an energy consumption equation representing maintenance energy required for cellular functions other than growth. The objective function is generally defined to maximize the growth rate. This assumption is derived from the idea that, over time, the fastest growing organisms eventually dominate a population, so the model should be directing metabolite flux in such a way to allow the fastest growth. Other objective functions such as maximizing production of energy or of some desired product have also been used.

At the base level, the metabolic model used for flux balance analysis is composed of all linear equations, so linear programming (LP) techniques can readily be applied to solve for a flux distribution. Additional details such as regulatory effects can introduce nonlinearities or on/off decisions, requiring nonlinear programing (NLP) or mixed integer programing (MILP/MINLP)

to solve. There are a number of useful programs available to aid in performing FBA, including the algebraic modeling software GAMS [5] and the COBRA Toolbox for Matlab [6] [7].

CHAPTER 2 – METABOLIC MODEL: CONSTRUSCTION AND REFINEMENT

Initial Model Development

Feist et al. have published a thorough review on the construction of metabolic network models. For all enzymes involved in metabolism the following information is required [8]: (a) what reaction(s) the enzyme catalyzes, (b) the reactants, products, and related stoichiometries for each associated reaction, (c) the reversibility of the reaction, and (d) the cellular compartment in which the reaction is localized. Transport reactions, both for cellular uptake and excretion as well as for transport between the cellular compartments, must also be identified. This includes passive transport through membrane diffusion and through pores in addition to active transport by membrane proteins [2].

Annotated genomic sequences are generally good initial resources for information regarding enzymes present in an organism. Though some enzymes identified in the genome sequence may be well studied, many will not. Homology searches comparing unknown genes to genes coding known enzymes in other organisms can help indicate which are most likely to be involved in metabolism [2]. Some useful databases of genomic and metabolic information are listed in Table 1 [3], and Tomar et al. give a short description of many of them in their review on system construction [9].

It is important to note the relationships from genes to enzymes to reactions when developing the model. Reaction rules provide the framework for connecting genes to reactions based on the enzymes they encode. These rules are defined by Boolean statements that relate the expression of individual genes to the presence of the reaction in the network. The basic reaction rule is a one to one true/false connection between a gene and the reaction the associated enzyme

Table 1. Useful resources for constructing and analyzing a metabolic network. [3]

Resource	URL	Description
Metabolic reconstruction		
Kyoto Encyclopaedia of Genes and Genomes (KEGG)	http://www.genome.jp/kegg/	Pathway databases for several organisms
BioCyc	http://www.biocyc.org	Pathway databases for several organisms
PEDANT	http://pedant.gsf.de/	Genome annotations
Reactome	http://www.reactome.org/	Curated database of biological processes in humans
Biomodels.net	http://www.biomodels.net/	Kinetic models of pathways, many published models from literature
BRENDA	http://www.brenda-enzymes.info/	Biochemical and molecular information on enzymes
SABIO-RK Database	http://sabio.villa-bosch.de/	System for the analysis of biochemical pathways – reaction kinetics
Software tools		
Constraint-based reconstruction and analysis (COBRA) toolbox	http://www.bioeng.ucsd.edu/research/research.groups/gcrg/downloads/COBRAToolbox/	Interfaces with MATLAB for extensive analysis of networks using FBA; performs gene deletions – single and multiple (can interface with LINDO, GLPK, CPLEX)
MetaFluxNet	http://mbel.kaist.ac.kr/lab/mfn/	Metabolic flux analysis
CellNetAnalyzer	http://www.mpi-magdeburg.mpg.de/projects/cna/cna.html	Structural and functional analysis of cellular networks
SNA: Stoichiometric network analysis	http://www.bioinformatics.org/project/?group.id=546	Mathematica toolbox for stoichiometric network analysis
Yana	http://yana.bioapps.biozentrum.uni-wuerzburg.de/	Network reconstruction, visualization and analysis
PathwayAnalyser	http://sourceforge.net/projects/pathwayanalyser	FBA and MoMA of metabolic networks; gene deletion studies
Systems Biology Research Tool	http://www.bioc.uzh.ch/wagner/software/SBRT/	Multiple methods for analysing stoichiometric networks
SBML Software Guide	http://sbml.org/SBML_Software_Guide	Resource list for software tools, model databases
Solvers for FBA/MoMA		
LINDO	http://www.lindo.com/	Commercial solver for optimization problems
CPLEX	http://www.ilog.com/products/cplex/	Commercial optimization software package
GNU linear programming toolkit (GLPK)	http://www.gnu.org/software/glpk/	Solver for LP problems
Object oriented quadratic programming (OOQP)	http://pages.cs.wisc.edu/~swright/ooqp/	Solver for QP problems

performs. Therefore, when that gene is expressed in the organism, the reaction is present in the model, and when the gene is not active (due to transcription regulation or genetic modifications such as a knockout or mutation), then that reaction is absent from the model. When multiple enzymes perform the same reaction on the same substrate in the same cellular compartment, they are called isozymes. The presence of an isozyme is indicated in that reaction's rules by an "or" statement, so the reaction is included in the model network whenever any single gene is present, or multiple of the isozymes' coding genes are present. The remaining reaction rule deals with enzyme complexes. Enzyme complexes are a multimeric construction of individual proteins which together perform a specific function. The functional activity of a complex is inhibited (completely or significantly) when any of its subunits are absent. This feature of enzyme

complexes is captured by an “and” statement in the associated reaction rule. This means that that reaction requires all of the associated enzymes to be expressed in order for it to be included in the model. More complicated reaction rules can be seen when enzyme complexes are isozymes for other proteins.

The reactions identified in the initial search outline the process of substrate uptake from the surroundings, catabolism of that substrate into intermediate metabolites, and the anabolism of those metabolites into biologically necessary molecules. An organism cannot simply accumulate these end metabolites indefinitely, and once sufficient reserves of the necessary components are available to the organism, it will reproduce. This process is represented in the model by a biomass reaction equation which consumes a stoichiometric ratio of key end metabolites such as proteins, nucleic acids, lipids, and other components [4].

Developing a biomass equation begins with an analysis of the macromolecular biomass components just mentioned to determine the mass fractions of each macromolecule relative to the dry weight of the biomass. Next, the macromolecules are broken down into the metabolites necessary to synthesize them. For example, DNA is broken into the individual nucleic acids and proteins are broken down to their constituent amino acids. These metabolites are represented in the growth equation as being consumed from their general pools [8]. This base equation is then expanded to include the energy required to synthesize the biologically necessary macromolecules as a part of the cellular maintenance energy. Further requirements for various vitamins or cofactors necessary for growth can be included [10]. The importance of determining an accurate biomass equation cannot be stressed enough, as growth rate is one of the primary phenotypic indicators for quantifying the performance of the model.

Though genomic data provides a framework for the metabolic model, there will be many areas that require manual refinement. Homology searches may match the coded enzyme to its function, but many enzymes with the same enzyme commission (EC) number have different activities, substrate specificity, or compartmentalization within a specific organism. Other genes may have been misidentified, leading to network gaps, where certain pathways are missing key components, or synthetic pathways which are not actually present. A thorough literature search for enzyme information specific to the organism of interest is a first step towards correcting some of these errors [8].

Model Improvement by Experimental Validation

The most common indicators that a model contains an error are growth prediction inconsistencies. These are cases where, for a specific set of conditions, the model growth prediction is contradictory to experimental results. When the model predicts growth but there is no growth experimentally (G/NG), there are typically additional reaction(s) included in the model that are not actually present. Similarly, NG/G inconsistencies can indicate network gaps that must be closed with additional reactions to bring the model more closely in line with reality. Records of tested experimental growth results as well as those found in the literature should be kept judiciously, and all changes to the model verified against these experimental results. If a model modification would cause a NG/NG result to become a G/NG (or a G/G become a NG/G), the question must be asked if the modification is not truly reflective of the metabolic mechanism, or if the experimental results were due to subpar technique or some other experimental factor.

The first experimental conditions that are relatively straightforward to verify in the model using FBA are growth on different substrates. This is implemented in the model by manipulating

the bounds on the various exchange reactions representing substrate uptake. For example, to simulate growth on sucrose exclusively, the bounds on the sucrose uptake reaction would be set to some physically realistic level (often determined through experimentation) while the bounds for all other substrates would be set to zero:

$$Ex\_Suc_{LB} \geq -10;\ Ex\_Suc_{UB} \leq 0$$

$$Ex\_Glu_{LB} \geq 0;\ Ex\_Glu_{UB} \leq 0$$

$$Ex\_Fru_{LB} \geq 0;\ Ex\_Fru_{UB} \leq 0$$

The lower bounds are set to a negative number when representing cellular uptake. This is because exchange reactions are represented in most models as a conversion from intracellular to extracellular species:

$$Suc \leftrightarrow Suc_{ex}$$

so allowing the flux through that reaction to fluctuate negatively models the organism having the option to utilize that substrate.

This does not ensure that the simulation will indicate growth, as there must be a path for that substrate to be converted into each of the necessary biomass precursors. There must also be a path that consumes all additional metabolites produced in those necessary reactions, or the model will not predict flux through those paths. This is because these "dead end" metabolites would accumulate until reaching a level dangerous for the organism. GapFind and GapFill are algorithms available to identify these dead end metabolites and to determine the minimum number of model modifications necessary to reconnect that metabolite with the rest of the network, respectively [11]. The four modifications considered by the GapFill algorithm are reversing the directionality of reactions, adding reactions known to exist in other organisms found in various databases, adding additional extracellular uptake reactions, or by adding

transport reactions between intracellular compartments. Both algorithms use binary variables, GapFind to indentify if a metabolite has reactions leading to and away from it and GapFill to indicate whether or not a particular change should be made to the model. This changes the problem into a mixed-integer linear program (MILP). When these algorithms were applied to *E. coli* and *S. cerevisia* models it was found that changing reaction directionality was the primary mechanism used to fix model gaps [11]. SMILEY, a similar algorithm that only focuses on adding metabolic or transport reactions, was used by Reed et al. to identify two new enzymatic activities and four transport functions that were missing in their current *E. coli* model [12]. Zamorrodi et al. outline additional algorithms that attempt to reconcile growth prediction inconsistencies in their review [13].

CHAPTER 3 – MODEL EXPANSIONS

Detailed Thermodynamics

The thermodynamic consideration of metabolic models has typically been simplified down to the reversibility of the reactions. The inclusion of more complex thermodynamic constraints has been able to correct model predictions in many cases.

Beard et al. made an analogy to Kirchoff's Laws for electrical circuits to impose additional thermodynamic constraints. Here, they considered the chemical potential change of each reaction to be analogous to an electrical potential difference across a circuit element. Thus, the chemical potential changes along any closed reaction loop in the system must sum to zero for the loop to be feasible. This eliminates thermodynamically infeasible cycles from the flux search space. The chemical potential constraints are nonlinear, however, and thus do not ensure a unique solution [14].

Realistic reversibility constraints were later introduced by Hoppe et al. by relating the change in Gibb's free energy for the reaction to the specific metabolite concentrations. The process requires knowledge of the metabolite pool sizes associated with the reaction. From this knowledge, the actual reversibility of the reaction could be determined. Variable reversibility constraints were able to explain why certain predicted nonessential genes were in fact essential [15].

Regulatory Controls and Dynamic Simulations

Though regulation plays an important role in cellular operations, it has previously been typically neglected in FBA [2]. The absence of regulatory control has in the past limited the

predictive capabilities of metabolic models [16]. This trend is changing as we gain more insight into cellular metabolism.

Early regulation constraints introduced to metabolic models took the form of Boolean operators to represent flux constraints due to transcriptional regulation associated with specific environmental conditions [17] [18]. For a given initial condition, the Boolean rules determining the expression of each gene would be evaluated. A standard FBA is then performed under the constraints imposed by the regulatory model, and from those results a new set of environmental conditions is determined. The process then iterates through the regulatory model using the new environmental conditions until a state of agreement between the regulatory and metabolic systems is reached. This reduces the feasible space of reaction fluxes by eliminating a large number of the extreme pathways from consideration [19]. By cycling between the two models, a simulation of a time course of transcriptional events can be produced. This method has been used to predict the transcriptional regulation process of *E. coli* growth on glucose-lactose substrate [18], the repression of amino acid and other catabolite pathways [17], and to identify new regulatory mechanisms in *S. cerevisiae* using experimental growth data [20].

The diauxic shift in *E. coli* was also used to compare two methods of dynamic simulation when under the control of transcriptional regulation. The static approach divided the time period of interest into discreet blocks, performing a FBA for each block and integrating the results across the full range, similarly to the method explained above. A dynamic optimization approach based on nonlinear optimization of a set of ODEs was shown to be more flexible with respect to the introduction of experimental data, and but was much more computationally involved due to the nonlinearities [21].

Covert et al. later introduced ODEs to their regulatory mechanism in the form of a carbohydrate uptake control model. This model demonstrated the framework for how available kinetic information can be incorporated into a metabolic model. They used this integrated flux balance analysis (iFBA) model to better predict the diauxic shift of *E. coli* in nearly 25% of the 334 single knockout mutants examined [22].

An automated regulatory-metabolic network generator called probabilistic regulation of metabolism (PROM) was developed to integrate high-throughput data available into a genome-scale, constraint based regulatory network. This process requires a fully reconstructed metabolic network, a structure of the regulatory network including transcription factors and their targets, and a significant amount of gene expression data under various environmental conditions and gene knockouts. Reconstructions of *E. coli* and *M. tuberculosis* using the PROM process predicted lethal knockout phenotypes with up 95% accuracy [23].

Alternative Objective Functions

Growth rate is often taken as the phenotype for which the organism is optimized [4]. This has shown to give predictions consistent with experimental data for both wild type [24] and knockout mutants [25] [26] under laboratory conditions. *E. coli* has also shown to evolve to match predicted growth rates when grown on a substrate like glycerol for which it had not previously adapted [27]. While this further confirms that life can optimize itself towards some goal, it also indicates that any given organism could be operating at a suboptimal level, or could be working towards some optimum other than growth rate.

While evolutionary pressure eventually moves an organism's metabolism towards an optimum, much of the useful work of metabolic engineering alters the network without waiting

generations for the organism to readapt. Minimization of metabolic adjustment (MOMA) reflects this by using quadratic programming to minimize the variation in flux distribution relative to the wild type, growth optimized flux distribution. MOMA has shown to give more accurate predictions for knockout strains relative to FBA [28].

A number of other alternative objective functions have been used. Maximization of ATP production has proven effective for modeling mitochondrial metabolism [29] [30]. Maximizing cellular energy efficiency can be accomplished by either minimizing the redox potential or by minimizing the production of ATP [31]. The objective function can also be set to maximize the production of a metabolite of interest.

The inclusion of metabolic flux data can also help to improve the accuracy of model predictions. As mentioned, optimization is used to identify the flux distribution that maximizes the given objective function. This is necessary because there are a large number of feasible distributions through the network that satisfy the stoichiometric constraints. Measurement of certain fluxes, such as the uptake of substrate from the media or the excretion of various products can be readily performed. For the difficult to measure intracellular fluxes, the radioactive carbon isotope 13C can be used to elucidate the intracellular flux distribution [32], as shown in Figure 2 [33].

To this end, the cells are grown on a labeled substrate until they reach a steady growth rate. Enough time is allowed to pass for the labeled carbon to be fully integrated through the amino acid synthesis pathways and distribute throughout the organisms' proteins. Mass spectroscopy or nuclear magnetic radiation techniques are then used to measure the distribution of 13C throughout the proteins. A model of the flux distribution can then be generated from this data. The distributions determined by the labeled carbon experiment can then be used to further

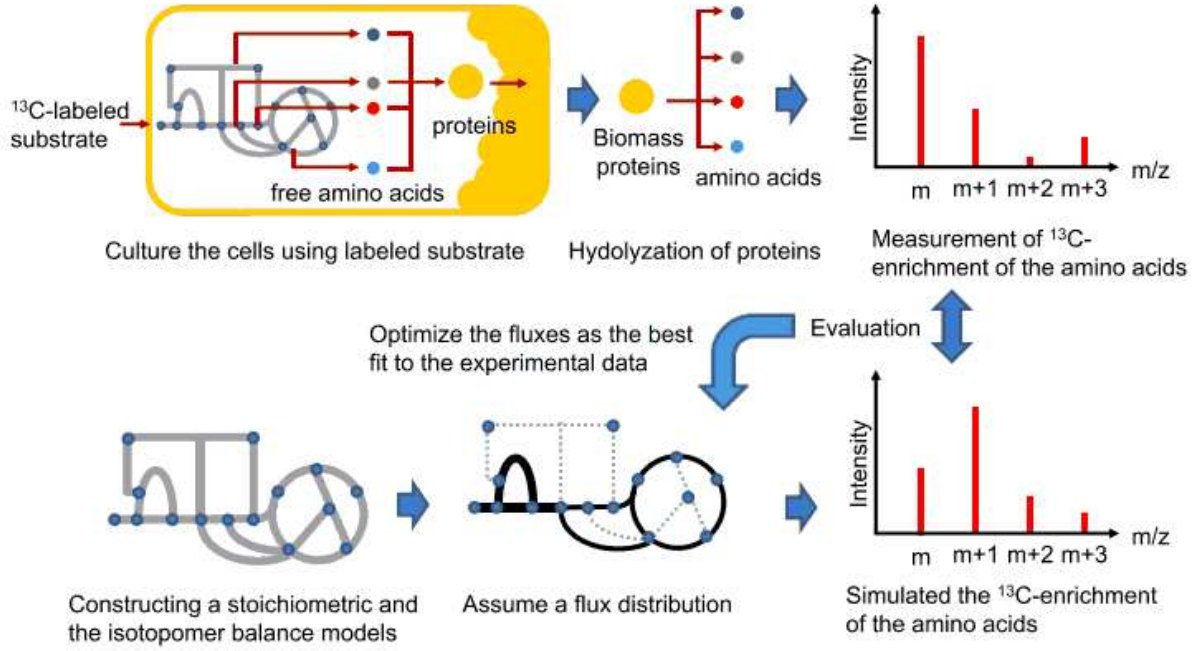

Figure 2. Overview of estimating a metabolite flux distribution using ^{13}C analysis [33]

refine the metabolic model. In particular, the flux distribution can be used as input for a MOMA objective function, reducing the deviation of predicted fluxes from wild type [28]. A similar algorithm, regulatory on/off minimization (ROOM), minimizes the total number of fluxes that are changed from the wild type [34].

CHAPTER 4 – FLUX BALANCE ANALYSIS APPLICATIONS

Identifying "Essential" and "Synthetic Lethal" Genes

A common use for FBA is the determination of "essential" genes and "synthetic lethal" gene pairs. Essential genes are those that are required for cell growth. To determine essentiality, each gene is "knocked out" in turn and a FBA is performed on the new reaction network. If the predicted growth rate for the knockout model is low, often taken between 0.1 and 1e-9 of the wild type growth rate, that gene which was knocked out is designated an essential gene. Synthetic lethal gene pairs are sets of two genes which, when knocked out individually, do not stop cellular growth, but do stop growth when both are knocked out. Synthetic lethals are determined in a similar manner to essentials. Typically a single gene deletion is first performed to determine the full set of essential genes. Then, starting from the set of non-essential genes, each pair is sequentially removed and an FBA is performed, using the same criteria for lethality as for essentiality. Theoretically, synthetic triples and other higher order gene sets could be determined the same way, but for full metabolic networks this process quickly becomes computationally prohibitive [13].

Maximizing Production of Biomolecules

One of the primary goals of metabolic engineering is to manipulate an organism to begin producing or to over produce a product of interest. FBA lends itself to this task readily, giving rapid insight into the effect of a wide range of metabolic engineering tools, including reaction knockouts and alternative path analysis, the addition of non-native reaction pathways, and up or down regulation of certain genes [35].

Gene knockouts are an effective strategy to maximize the production of a molecule of interest. They are often implemented to eliminate competing pathways that use the same precursor metabolites as the target compound. As chemical accumulation cannot occur in a system growing at steady state, increasing those metabolite precursor pools leads to increasing the driving force through the desired reaction. It should be evident that when designing knockout strategies, prior knowledge of the essential and synthetic lethal genes greatly reduces the number of unfruitful attempts which result in no cell growth.

The OptKnock algorithm is used to design knockout strategies for the overproduction of a metabolite of interest [36]. It operates using a bi-level optimization program which seeks to maximize both growth rate and metabolite production simultaneously. When identifying genes for deletion, the program seeks ways of fixing the production of the target metabolite to the production of biomass precursors. In other words, when the biomass precursor is produced, the metabolite must also be produced due to stoichiometric constraints. The OptForce algorithm, also developed by Maranas and coworkers, uses a similar bi-level optimization framework to identify schemes for product maximization that involve both the up- and down-regulation of genes in addition to knockouts [37]. Table 2 presents many similar optimization algorithms available for performing metabolic manipulations [13]. For a thorough review of these algorithms, see Zomorrodi, 2012.

One recent application of manipulating the existing metabolic network was the reversal of the β-oxidation cycle in *E. coli* by Gonzalez and coworkers [38]. This cycle is natively the process by which long chain fatty acids are broken down into small chain metabolites which the cell can use for the construction of other biomolecules. Here, the authors found manipulations that allowed the cycle to function in reverse, using two carbon sugars to construct long chain

Table 2. Algorithms for the targeted redesign of reconstructed metabolic networks [13]

Name	Type of optimization problem	Type of intervention	Accessibility	Reference
OptKnock	Bilevel, MILP	Knockouts	GAMS	Burgard et al. (2003)
RobustKnock	Multi-level, MILP	Knockouts	MATLAB	Tepper and Shlomi (2009)
OptGene	Evolutionary	Knockouts	Online (as part of OptFlux)	Patil et al. (2005)
Objective tilting	Bilevel, MILP	Knockouts	MATLAB (via COBRA toolbox)	Feist et al. (2010)
OptStrain	Bilevel, MILP	Addition of non-native reactions/pathways	N/A	Pharkya et al. (2004)
SimOptStrain	Bilevel, MILP	Knockouts and addition of non-native reactions/pathways	N/A	Kim et al. (2011)
BiMOMA	Bilevel, MINLP	Knockouts	N/A	Kim et al. (2011)
OptReg	Bilevel, MILP	Knockouts, upregulations and downregulations	N/A	Pharkya and Maranas (2006)
GDLS	Heuristic	Knockouts, upregulations and downregulations	N/A	Lun et al. (2009)
FSEOF	LP	Upregulations and downregulations	N/A	Choi et al. (2010)
OptORF	Bilevel, MILP	Knockouts and overexpressions (of both metabolic and regulatory genes)	N/A	Kim and Reed (2010)
OptForce	Bilevel, MILP	Knockouts, upregulations and downregulations	GAMS	Ranganathan et al. (2010)
EMILiO	Bilevel, MILP	Knockouts, upregulations and downregulations	N/A	Yang et al. (2011)

hydrocarbons. This was done using a combination of targeted gene knockouts of competing pathways and some genes involved in metabolite regulation, over and under expression of certain enzymes to maximize the short chain carbon precursor pool (to maximize the driving force through the reversed cycle), and the introduction of select non-native termination enzymes to pull finished products from the cycle. These termination enzymes added a modular dimension to the process, giving the ability to generate different functional products of different chain lengths dependent upon the specific termination enzyme expressed. This procedure showed both the novel ways in which the native components of an organism can be manipulated towards some goal, and the introduction of non-native enzymes to a host for additional functionality.

CHAPTER 5 – MODEL LIMITATIONS AND FUTURE EXPANSIONS

There are a number of areas related to the analysis of organisms by FBA that continue to be refined. The availability of high-throughput data analysis to analyze genomic information is allowing metabolic networks to be constructed for a number of organisms. Due to incomplete data or inaccurate annotations, these reconstructions continue to have many errors. Careful experimental studies and judicious record keeping are essential to the development of accurate, applicable, and distributable metabolic models.

Currently, most analysis of these models uses a static picture of the network to gain insight into how the organism functions. Some expansions on traditional FBA have been made which begin to predict the dynamic functions of some organisms [19] [20] [22], but these are largely limited in scope. Many organisms, such as photosynthetic autotrophs, do not operate at a constant steady state. Rather, they alternate through a range of operational modes as they transition through their light and dark reaction cycles. A better grasp on these dynamics will greatly boost the predictive capabilities of the related models.

Related to the cyclical nature of some organisms is the need to identify more representative objective functions. The maximization of biomass has been shown to apply to some organisms, but only for constant laboratory settings after enough generations have been exposed to that environmental condition to evolutionarily adapt to it. Work has been ongoing to attempt to identify what classes of objective functions best apply given specific circumstances, often incorporating experimental data in the search for more applicable objective functions [39] [40] [41].

CONCLUSION

The application of Flux Balance Analysis to stoichiometric metabolic models has shown to have useful predictive capabilities for a number of areas. As additional information such as regulatory controls and more realistic thermodynamic considerations are included, these predictions are becoming more more accurate. The incorporation of experimental data is essential to identifying areas for improvement in the model. The rapid increases of computational power seen in recent years also lends itself to more complex methods of analysis, and many algorithms have been developed to assist in model development, refinement, and analysis. The versatility of FBA and related techniques make these constraint based analysis an appealing starting point when designing metabolic engineering strategies for organisms used in bioprocesses.

REFERENCES

[1] J. D. Orth, T. M. Conrad, J. A. Lerman, H. Nam, A. M. Feist and B. O. Palsson, "A comprehensive genome-scale reconstruction of Escherichia coli metabolism," *Molecular systems biology,* vol. 7, no. 1, 2011.

[2] K. J. Kauffman and J. S. Edwards, "Advances in flux balance analysis," *Current Opinion in Biotechnology,* vol. 14, no. 5, pp. 491-496, 2003.

[3] K. Raman and N. Chandra, "Flux balance analysis of biological systems: applications and challenges," *Breifings in bioinformatics,* vol. 10, no. 4, pp. 435-449, 2009.

[4] J. D. Orth, I. Thiele and B. O. Palsson, "What is flux balance analysis?," *Nature Biotechnology,* vol. 28, no. 3, pp. 245-248, 2010.

[5] A. Brook, D. A. Kendrick, A. Meeraus and R. E. Rosenthal, "GAMS: A user's guide," *Course Technology,* 1988.

[6] S. A. Becker, A. M. Feist, M. L. Mo, G. Hannum, B. O. Palsson and M. J. Herrgard, "Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox," *Nature protocols,* vol. 2, no. 3, pp. 727-738, 2007.

[7] J. Schellenberger, R. Que, R. M. Fleming, I. Thiele, J. D. Orth, A. M. Feist and B. O. Palsson, "Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0," *Nature protocols,* vol. 6, no. 9, pp. 1290-1307, 2011.

[8] A. M. Feist, M. J. Herrgard, I. Thiele, J. L. Reed and B. O. Palsson, "Reconstruction of biochemical networks in microorganisms," *Nature Reviews Microbiology,* vol. 7, no. 2, pp. 129-143, 2009.

[9] N. Tomar and R. K. De, "Comparing methods for metabolic network analysis and an application to metabolic engineering," *Gene,* vol. 521, no. 1, pp. 1-14, 2013.

[10] A. M. Feist and B. O. Palsson, "The biomass objective function," *Current opinion in microbiology,* vol. 13, no. 3, pp. 344-349, 2010.

[11] V. S. Kumar, M. S. Dasika and C. D. Maranas, "Optimization based automated curation of metabolic reconstructions," *BMC bioinformatics,* vol. 8, no. 1, p. 212, 2007.

[12] J. L. Reed, T. R. Patel, K. H. Chen, A. R. Joyce, M. K. Applebee, C. D. Herring and B. O. Palsson, "Systems approach to refining genome annotation," *Proceedings of the National Academy of Sciences,* vol. 103, no. 46, pp. 17480-17484, 2006.

[13] A. R. Zomorrodi, P. F. Suthers, S. Ranganathan and C. D. Maranas, "Mathematical optimization applications in metabolic networks," *Metabolic engineering,* vol. 14, no. 6, p. 672, 2012.

[14] D. A. Beard, S. D. Liang and H. Qian, "Energy balance for analysis of complex metabolic networks," *Biophysical journal,* vol. 83, no. 1, pp. 79-86, 2002.

[15] A. Hoppe, S. Hoffmann and H. G. Holzhutter, "Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks," *BMC systems biology,* vol. 1, no. 1, p. 23, 2007.

[16] A. K. Gombart and J. Nielsen, "Mathematical modelling of metabolism," *Current Opinion in Biotechnology,* vol. 11, no. 2, pp. 180-186, 2000.

[17] M. W. Covert, C. H. Schilling and B. O. Palsson, "Regulation of gene expression in flux balance models of metabolism," *Journal of theoretical biology,* vol. 213, no. 1, pp. 73-88, 2001.

[18] M. W. Covert, W. Markus and B. O. Palsson, "Transcriptional regulation in constraints-based metabolic models of Escherichia coli," *Journal of Biological Chemistry,* vol. 277, no. 31, pp. 28058-28064, 2002.

[19] M. W. Covert and B. O. Palsson, "Constraints-based models: regulation of gene expression reduces the steady-state solution space," *Journal of theoretical biology,* vol. 221, no. 3, pp. 309-325, 2003.

[20] M. J. Herrgard, B. S. Lee, V. Portnoy and B. O. Palsson, "Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae," *Genome research,* vol. 16, no. 5, pp. 627-635, 2006.

[21] R. Mahadevan, J. S. Edwards and F. J. Doyle III, "Dynamic Flux Balance Analysis of Diauxic Growth in Escherichia coli," *Biophysical Journal,* vol. 83, no. 3, pp. 1331-1340, 2002.

[22] M. W. Covert, N. Xiao, T. J. Chen and J. R. Karr, "Integrating metabolic, transcriptional regulatory and signal tranduction models in Escherichia coli," *Bioinformatics,* vol. 24, no. 18, pp. 2044-2050, 2008.

[23] S. Chandrasekaran and N. D. Price, "Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis," *Proceedings of the National Academy of Sciences,* vol. 107, no. 41, pp. 17845-17850, 2010.

[24] J. S. Edwards, R. U. Ibarra and B. O. Palsson, "In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data," *Nature biotechnology,* vol. 19, no. 2, pp. 125-130, 2001.

[25] D. Deutscher, I. Meilijson, M. Kupiec and E. Ruppin, "Multiple knockout analysis of genetic robustness in the yeast metabolic network," *Nature genetics,* vol. 38, no. 9, pp. 993-998, 2006.

[26] N. C. Duarte, M. J. Herrgard and B. O. Palsson, "Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model," *Genome research,* vol. 14, no. 7, pp. 1298-1309, 2004.

[27] R. U. Ibarra, J. S. Edwards and B. O. Palsson, "Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth," *Nature,* vol. 420, no. 6912, pp. 186-189, 2002.

[28] D. Segre, D. Vitkup and G. M. Church, "Analysis of optimality in natural and perturbed metabolic networks," *Proceedings of the National Academy of Sciences,* vol. 99, no. 23, pp. 15112-15117, 2002.

[29] R. Ramakrishna, J. S. Edwards, A. McCulloch and B. O. Palsson, "Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology,* vol. 280, no. 3, pp. R695-R704, 2001.

[30] T. D. Vo, H. J. Greenberg and B. O. Palsson, "Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and

biochemical data," *Journal of Biological Chemistry,* vol. 279, no. 38, pp. 39532-39540, 2004.

[31] J. M. Savinell and B. O. Palsson, "Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism," *Journal of theoretical biology,* vol. 154, no. 4, pp. 421-454, 1992.

[32] W. Wiechert, "13C Metabolic Flux Analysis," *Metabolic engineering,* vol. 3, no. 3, pp. 195-206, 2001.

[33] Y. Toya and S. Hiroshi, "Flux analysis and metabolomics for systematic metabolic engineering of microorganisms," *Biotechnology advances,* vol. 31, no. 6, pp. 818-826, 2013.

[34] T. Shlomi, O. Berkman and E. Ruppin, "Regulatory on/off minimization of metabolic flux changes after genetic perturbations," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 102, no. 21, pp. 7695-7700, 2005.

[35] R. Mahadevan, A. P. Burgard, I. Famili, S. Van Dien and C. H. Schilling, "Applications of metabolic modeling to drive bioprocess development for the production of value-added chemicals," *Biotechnology and Bioprocess Engineering,* vol. 10, no. 5, pp. 408-417, 2005.

[36] A. P. Bugard, P. Pharkya and C. D. Maranas, "Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization," *Biotechnology and Bioengineering,* vol. 84, no. 6, pp. 647-657, 2003.

[37] S. Ranganathan, P. F. Suthers and C. D. Maranas, "OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions," *PLoS computational biology,* vol. 6, no. 4, p. e1000744, 2010.

[38] C. Dellomonaco, J. M. Clomburg, E. N. Miller and R. Gonzalez, "Engineered reversal of the [bgr]-oxidation cycle for the synthesis of fuels and chemicals," *Nature,* vol. 476, no. 7360, pp. 355-359, 2011.

[39] A. P. Burgard and C. D. Maranas, "Optimization-based framework for inferring and testing hypothesized metabolic objective functions," *Biotechnology and bioengineering,* vol. 82, no. 6, pp. 670-677, 2003.

[40] R. Schuetz, L. Kuepfer and U. Sauer, "Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli," *Molecular systems biology,* vol. 3, no. 1, 2007.

[41] A. L. Knorr, R. Jain and R. Srivastava, "Bayesian-based selection of metabolic objective functions," *Bioinformatics,* vol. 23, no. 3, pp. 351-357, 2007.