

MIT OpenCourseWare  
<http://ocw.mit.edu>

6.047 / 6.878 Computational Biology: Genomes, Networks, Evolution  
Fall 2008

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

# Lecture 21: Introduction to Steady State Metabolic Modeling

November 18, 2008

## 1 Introduction

Metabolic modeling allows us to use mathematical models to represent complex biological systems. This lecture discusses the role that modeling biological systems at the steady state plays in understanding the metabolic capabilities of interesting organisms and how well steady state models are able to replicate *in vitro* experiments.

### 1.1 What is Metabolism?

According to Matthews and van Holde, metabolism is “the totality of all chemical reactions that occur in living matter”. This includes catabolic reactions, which are reactions that lead the break down of molecules into smaller components and anabolic reactions, which are responsible for the creation of more complex molecules (e.g. proteins, lipids, carbohydrates, and nucleic acids) from smaller components. These reactions are responsible for the release of energy from chemical bonds and the storage of energy respectively. Metabolic reactions are also responsible for the transduction and transmission of information (for example, via the generation of cGMP as a secondary messenger or mRNA as a substrate for protein translation).

### 1.2 Why Model Metabolism?

An important application of metabolic modeling is in the prediction of drug effects. An important subject of modeling is the organism *Mycobacterium tuberculosis* [1]. The disruption of the mycolic acid synthesis pathways of this organism would help control TB infection. Computational modeling gives us a platform for identifying the best drug targets in this system. Gene knock-out studies in *Escherichia coli* have allowed scientists to determine which genes and gene combinations affect the growth of that important model organism [2]. Both agreements and disagreements between models and experimental data can help us assess our knowledge of biological systems and help us improve our

predictions about metabolic capabilities. In the next lecture, we will learn the importance of incorporating expression data into metabolic models.

## 2 Model building

### 2.1 Chemical reactions

In metabolic models, we are concerned with modeling chemical reactions that are catalyzed by enzymes. Enzymes work by facilitating a transition state of the enzyme-substrate complex that lowers the activation energy of a chemical reaction. The diagram on slide 5 of page 1 of the lecture slides demonstrates this phenomenon. A typical rate equation (which describes the conversion of the substrates of the enzyme reaction into its products  $S=P$ ) can be described by a Michaelis-Menten rate law:  $\frac{V}{V_{max}} = \frac{[S]}{K_m + [S]}$ , where  $V$  is the rate of the equation as a function of substrate concentration.  $K_m$  and  $V_{max}$  are the two parameters necessary to characterize the equation.

The inclusion of multiple substrates, products, and regulatory relationships quickly increases the number of parameters necessary to characterize enzyme function. The figures on slides 1, 2, and 3 of page 2 of the lecture notes demonstrate the complexity of biochemical pathways. Kinetic modeling quickly becomes infeasible because the necessary parameters are difficult to measure and also vary across organisms [3]. Thus, we are interested in a modeling method that would allow us to avoid the use of large numbers of poorly-determined parameters.

### 2.2 Steady-state assumption

The steady state assumption allows us to represent reactions entirely in terms of their chemistry (the stoichiometric relationships between the components of the enzymatic reaction) by assuming that there is not accumulation of any metabolite in the system. This does not mean that there is not flux through any of the reactions, simply that accumulation does not occur. An analogy is to a series of waterfalls that contribute water to pools. As the water falls from one pool to another, the water levels do not change even though water continues to flow (see page 2 slide 5). This framework prevents us from seeing transient kinetics that can result from perturbations of the system, but if we are interested in long-term metabolic capabilities (functions on a scale longer than milliseconds or seconds) then steady state dynamics may give us all the information that we need.

An important aspect of the steady-state assumption is that reaction stoichiometries are conserved across species, whereas the biology of enzyme catalysis (and the parameters that characterize it) are not conserved across species. This makes the ability to generalize across species and reuse conserved pathways in models much more feasible.

## 2.3 Reconstructing Metabolic Pathways

There are several databases that can provide the information necessary to reconstruct metabolic pathways *in silico*. These databases allow reaction stoichiometry to be accessed using Enzyme Commission numbers, which are the same in each organism that uses that particular enzyme. Among the databases of interest are ExPASy [4], MetaCyc [5], and KEGG [6]. These databases often contain pathways organized by function that can be downloaded in a format such as SBML, often making pathway reconstruction very easy for well-characterized pathways. ExPASy [4], MetaCyc [5], and KEGG [6]. These databases often contain pathways organized by function that can be downloaded in a format such as SBML, often making pathway reconstruction very easy for well-characterized pathways.

## 3 Metabolic Flux Analysis

Metabolic flux analysis (MFA) is a way of computing the distribution of reaction fluxes that is possible in a given metabolic network at steady state. We can place constraints on certain fluxes in order to limit the space described by the distribution of possible fluxes.

### 3.1 Mathematical representation

We can represent the flux of each substrate  $x_i$ , we can represent the rate of change of that substrate as  $\frac{dx_i}{dt} = S \bullet v$ , where  $S$  is a vector that describes the stoichiometric coefficients of that metabolite in each reaction in which it is a substrate or product. With this steady state assumption, we have  $0 = S \bullet v$ . Because we are interested modeling pathways, we represent stoichiometric as an  $m \times n$  matrix  $\mathbf{S}$  and fluxes as an  $n$ -dimensional matrix  $\mathbf{v}$ . Page 3 slides 3 and 4 show figures of these mathematical representations.

### 3.2 Null space of $\mathbf{S}$

The feasible flux space through the reactions of the model are defined by the null space of  $\mathbf{S}$  (page 4 slide 5). In addition, there are a series of non-unique basis vectors  $b_i$  that describe the null space (page 4 slide 6). All of the null space fluxes are this linear combinations of the basis and  $v = \sum_i a_i b_i$ . These can be found by standard methods such as singular value decomposition (SVD) [7]. These basis vectors represent extreme pathways at the edge of the flux cone, which is a geometric representation of the flux space and give us important information about the metabolic capabilities of the organisms in which we are interested (page 5 slide 2).

### 3.3 Constraining the Flux Space

One obvious constraint that we can place on our fluxes is that they must be positive. Within this framework we can represent reverse reactions as separate reactions within the stoichiometric matrix and then set a lower bound on all the reactions of 0. Likewise, if we happen to know the maximum fluxes of any of our reactions (these values correspond to our  $V_{max}$  parameters), then we can also place these constraints on the flux space. We can also add input and output fluxes that represent transport into and out of our cells, which are often much easier to measure than internal fluxes and can thus serve to help us to generate a more biologically-relevant flux space. An example of an algorithm for solving this problem is the simplex algorithm [8]. Page 5 slides 4-6 demonstrate how constraints on the fluxes change the geometry of the flux cone. In reality, we are dealing with problems in higher-dimensional spaces.

### 3.4 Linear Programming

The constraints described above give us the linear programming problem described on slide 31. We justify the formulation of the flux determination problem as a linear programming problem by recognizing that for a given point in time and a given environment an organism will demonstrate one particular flux distribution. In order to try and narrow down our feasible flux, we assume that there exists a fitness function which is a linear combination of any number of the fluxes in the system. Linear programming (or linear optimization) involves maximizing or minimizing a linear function over a convex polyhedron specified by linear and non-negativity constraints.

We solve this problem by identifying the flux distribution that maximizes an objective function  $z = \sum_i c_i v_i = c^T v$ . Our solutions lie at the boundaries of the permissible flux space and can be on points, edges, or both. This concept is demonstrated on page 6 slide 1. In that slide,  $A$  is the stoichiometric matrix,  $x$  is the vector of fluxes, and  $b$  is a vector of maximal permissible fluxes.

In microbes such as *E. coli*, this objective function is often a combination of fluxes that contributes to biomass. However, this function need not be completely biologically meaningful. For example, we might simulate the maximization of mycolates in *M. tuberculosis*, even though this isn't happening biologically. It would give us meaningful predictions about what perturbations could be performed *in vitro* that would perturb mycolate synthesis even in the absence of the maximization of the production of those metabolites.

Flux balance analysis was pioneered by Palsson's group at UCSD and has since been applied to *E. coli*, *M. tuberculosis*, the human red blood cell [9].

## 4 Applications of metabolic modeling

One application of metabolic modeling is called deletion analysis. In this technique, we develop a metabolic model for an organism of interest and calculate the feasible flux space with constraints while systematically removing reactions

to generate *in silico* mutants (page 6 slide 6). Using this method we can make computational predictions about the metabolic capabilities of microbial knockout strains [10]. Geometrically, each knockout changes the shape of the flux cone and constrains the feasible flux space (page 7 slide 1). If we are using growth as our objective function, we can then use the value of the objective function at the optimal solution as a measure of phenotype. Errors (discordance with experimental observations) can then help us to assess the validity of our model and observations. This technique has been used successfully in *E. coli* and yeast [11].

In [10], Edwards and Palsson modeled the central metabolism of *E. coli* using 720 reactions and 436 metabolites. This large model is an good example of a system that is modeled well by a steady-state model and FBA but would be exceedingly difficult to describe well using ordinary differential equations. Edwards and Palsson simulated knockouts with perturbations in processes including glycolysis, electron transport, the pentose phosphate pathway, and TCA. Simulation results were compared with experimental results and good predictions of lethal phenotypes, reduced growth phenotypes, and well-functioning phenotypes were achieved.

The results of [11] may be less impressive. Although good results with experimental data are achieved, most of the phenotypes are predicted to survive perturbations. Nevertheless, this model and the model of Edwards and Palsson both provide a computational foundation for the improvement of our biological knowledge.

## References

- [1] Jamshidi, N. and Palsson, B. . *BMC Systems Biology* **1**, 26 (2007). PMC1925256.
- [2] Edwards, J. S., Ibarra, R. U., and Palsson, B. O. *Nat Biotech* **19**(2), 125–130 February (2001).
- [3] Holmberg, A. *Mathematical Biosciences* **62**(1), 23–43 November (1982).
- [4] Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. *Nucl. Acids Res.* **31**(13), 3784–3788 July (2003).
- [5] Caspi, R., Foerster, H., Fulcher, C. A., Kaipa, P., Krummenacker, M., Latendresse, M., Paley, S., Rhee, S. Y., Shearer, A. G., Tissier, C., Walk, T. C., Zhang, P., and Karp, P. D. *Nucl. Acids Res.* **36**(suppl\_1), D623–631 (2008).
- [6] Kanehisa, M., Goto, S., Kawashima, S., and Nakaya, A. *Nucl. Acids Res.* **30**(1), 42–46 (2002).
- [7] Price, N. D., Reed, J. L., Papin, J. A., Famili, I., and Palsson, B. O. *Biophys. J.* **84**(2), 794–804 February (2003).

- [8] [http://en.wikipedia.org/wiki/Simplex\\_algorithm](http://en.wikipedia.org/wiki/Simplex_algorithm).
- [9] Edwards, J. S., Covert, M., and Palsson, B. *Environmental Microbiology* **4**(3), 133–140 (2002).
- [10] Edwards, J. S. and Palsson, B. O. *Proceedings of the National Academy of Sciences of the United States of America* **97**(10), 5528–5533 May (2000). PMC25862.
- [11] Förster, J., Famili, I., Palsson, B. O., and Nielsen, J. *Omics: A Journal of Integrative Biology* **7**(2), 193–202 (2003). PMID: 14506848.