1. Review

In the last lecture, we discussed how to use linear algebra to model metabolic networks with flux-balance analysis (FBA), which depends only on the stoichiometry of the reactions, not the kinetics. The metabolic network is represented as an $n \times m$ matrix $M$ whose columns are the $m$ reactions occurring in the network and whose rows are the $n$ products and reactants of these reactions.

$$
\begin{bmatrix}
\frac{dA}{dt} \\
\frac{dB}{dt} \\
\frac{dC}{dt} \\
\frac{dD}{dt} \\
\frac{dE}{dt} \\
\frac{dF}{dt} \\
\frac{dG}{dt} \\
\frac{dH}{dt} \\
\frac{dI}{dt}
\end{bmatrix} =
\begin{bmatrix}
R1 & R2 & R3 & R4 & R5 & R6 & R7 & R8 & R9 & R10 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
v1 \\
v2 \\
v3 \\
v4 \\
v5 \\
v6 \\
v7 \\
v8 \\
v9 \\
v10
\end{bmatrix}
$$

The entry $M(i, j)$ represents the relative amount of metabolite $i$ consumed or produced by reaction $j$. A positive value indicates production; a negative value indicates consumption. The nullspace of this matrix consists of all the $m \times 1$ reaction flux vectors that are possible given that the metabolic system is in steady state; i.e. all the sets of fluxes that do not change the metabolite concentrations.

$$
\frac{dx}{dt} = S V \\
0 \quad V_j \geq 0, \quad \forall j
$$

Steady State fluxes are constrained to the nullspace of $S$

The nullspace ensures that the system is in steady state, but it there are also additional constraints in biological systems: fluxes cannot be infinite, and each reaction can only travel in the forward direction (in cases where the backward reaction is also biologically possible, it is included as a separate column in the matrix). After bounding the fluxes and constraining the directions of the reactions, the resulting space of possible flux vectors is called the constrained flux-balance cone. The edges of the cone are known as its extreme pathways. By choosing an objective
function—some linear combination of the fluxes—to maximize, we can calculate the optimal values for the fluxes using linear programming and the simplex algorithm. For example, the objective function may be a weighted sum of all the metabolite fluxes that represents the overall growth rate of the cell (growth objective). Alternatively, we can maximize use of one particular product by finding the \( m \times 1 \) flux vector that is in the flux-balance cone and has the largest negative value of that product.

Maximize \( F = \sum C_i V_j \)
This can be solved by linear programming.

If you want to knock out a gene \( G_4 \), there are several steps.

1. Calculate the FBA for the whole matrix
2. Knock one column, then calculate the FBA again
3. We’ll get the maximum growth rate

What if the knock out is lethal?

At the end of the last lecture, we also discussed knockout phenotype predictions, a classic application of metabolic modeling. Experimental biologists often gather information about the function of a protein by generating transgenic organisms in which the gene encoding that protein has been disrupted, or knocked out. We can simulate in silico the effect of knocking out a particular enzyme on metabolism by assuming that the reaction catalyzed by that enzyme does not occur at all in the knockout: zeroing out the \( j \)th column of \( M \) effectively removes the \( j \)th reaction from the network. What used to be an optimal solution may now lie outside of the constrained flux-balance cone and thus no longer be feasible in the absence of the \( j \)th reaction. We can determine the new constrained flux-balance cone and calculate the new optimum set of fluxes to maximize our
objective function, predicting the effect of the knockout on metabolism. There is a good example on the slides, but we can find that there are some errors existed.

What do errors tell us? Errors indicate gaps in model or knowledge. The authors discuss 7 errors in prediction (Please see the slides). Five out of seven cases due to possible regulatory mechanisms. Namely, there is a loop. Knock out one gene, and other genes will compensate it.

2. Quantitative Flux Prediction

In today’s lecture, we mainly focus on the idea of quantitative flux prediction. The question is: can models quantitatively flux and/or growth rate? We would like to predict externally measurable fluxes or growth rate as function of controlled uptake rates and/or environmental conditions. Using a quasi-steady state model, we can look to predict time-dependent changes in the cell or environment using FBA.

2.1 Quantitative Modeling of Growth vs. Uptake Fluxes (Edwards, Ibarra, & Palsson, 2001)

Edwards et al. did an in silico experiment to model the relationship between the uptake rate of two different carbon sources, oxygen uptake rate, and maximal cellular growth rate. From the experiment, they hypothesized that E. coli has an optimized metabolic network which maximizes growth under the given environmental conditions. They then replicated the experiment in batch reactors to test how the computational methods compared to traditional biological methods.

To study the relationship between the factors under consideration, Edwards et al. controlled two variables and predicted a third, then observed actual measurements of growth in batch reactors for the free variables. For example, growth rate was predicted as a function of oxygen and acetate uptake rates.

Growth rate here is the value of projected function.
The predictions were then compared to measurements of all three variables taken from batch reactors. The results showed that E. coli do seem to have a metabolic network that is designed to maximize growth for a given condition. The paper suggests an interesting style of experimentation which combines computational and biological methods to study metabolism. Another surprising thing is, although they did not consider the effect of kinetic in this model, the prediction results are still matched to the experimental results.

2.2 Quasi Steady State Modeling (Palsson et al., 1994)

The previous example used FBA to make quantitative growth predictions under specific environmental conditions (point predictions), but is it also possible to predict time-dependent changes in response to varying environmental or cellular conditions? Varma and Palsson (1994) demonstrate the use of FBA to predict time-dependent changes in E. coli metabolism. Their quasi steady state model is based on the idea that time-dependent changes in fluxes can be modeled as transitions through individual time points that are themselves in steady state. There are two assumptions in the Quasi steady state modeling.

1. Metabolism adjusts to changes in environment/cell more rapidly than the changes themselves
2. Cell and environment concentrations may be changing, but metabolism operates as if concentration is static at each point in time (i.e. steady state)

To model a time profile of the system under the quasi steady state assumption, we use FBA to calculate the fluxes within each time interval. First, we divide time into slices of $\Delta t$. Because fluxes represent the derivatives of the metabolite concentrations, we can assume that the derivatives are constant over each $\Delta t$ and integrate to find the starting metabolite concentrations at the next time interval.

If we choose $\Delta t$ right, for each time, $t$, we can use FBA to predict substrate uptake ($Su$) and growth rate ($g$) by cells during interval $\Delta t^*$. Following, we can integrate it to obtain biomass ($B$) and substrate concentration at next time point $t + \Delta t$.

Namely, the Quasi steady state modeling is an iterative process in which we calculate the optimal fluxes and growth rate for the system at one time point, and then use those optimal fluxes to derive the initial environmental conditions for the next time point.

We are looking at the steady state. However, if $\Delta t$ is small enough, then we can predict the dynamic state. We should look at the shape of curve. From this prediction, we can not only get the qualitative phenotype, but also quantitative phenotype.
3. Regulation in metabolic models

We already know that the metabolic flux controlled by many levels of regulation, like transcriptional regulation, translational regulation, etc. Errors in FBA prediction can be explained by these gene regulations. However, how can we incorporate these regulations into metabolic models?

- We can treat regulations as a Boolean logic

\[
\text{trans} = \text{IF (G) AND NOT (B)}
\]

\[
\text{rxn} = \text{IF (A) AND (E)}
\]

(also must specify protein synthesis and degradation delays)

3.1 Regulation of gene expression in Flux Balance models of metabolism (Palsson et al., 2001)

The researchers used a set of known transcriptional regulatory events to constrain their analysis of a metabolic regulatory network. They simulate several environmental conditions in order to construct networks representing known systemic effects in the metabolic pathway. A simplified example of a regulatory circuit is shown in the slides. The example shows how a reaction can be represented as a Boolean variable, where if \( \text{rxn}=1 \), the reaction occurs, but if \( \text{rxn}=0 \), the reaction does not occur. Treating reactions as binary variables does not allow for quantitative analysis, but qualitative shifts in the metabolic flux can be predicted by incorporating regulatory information into the system in this way.

A particular example from the study presented by Covert et al. is shown in the slides. The researchers simulated a diauxic shift, a shift from one carbon source to another. The process includes two genes, \( \text{RpC1} \), which senses Carbon 1, and \( \text{Tc2} \) which transports Carbon 2. If Carbon 1 is sensed, \( \text{tTc2} \), transcribed \( \text{Tc2} \), will be absent, as the network will be using Carbon 1 as a carbon source.

We can represent this information as Booleans (see slides for details). But using Boolean logics only is not quantitative, can we predict qualitative flux shift? The slides have an overview of the procedure. We can calculate flux of each step, and then use Boolean to turn on or turn off genes.

4. Coupling Expression with Metabolism

But actually, we don’t need to artificially model gene levels, we can actually measure them. As discussed previously in lecture, we can measure mRNA expression throughout an organism by microarray experiments. As this information provides a set of expression levels of different genes under a certain condition, it would be extremely useful to incorporate it into the FBA. Usually,
data from microarray experiments is clustered, and unknown genes are hypothesised to have function similar to the function of those known genes with which they cluster. This analysis can be faulty, however, as genes with similar actions may not always cluster together. Incorporating microarray expression data into FBA could allow an alternate method of interpretation of the data.

Here arises a question, what is the relationship between gene level and flux through a reaction?

Colijn et al. address the question of algorithmic integration of expression data and metabolic networks. They apply FBA to model the maximum flux through each reaction in a metabolic network. For example, if microarray data is available from an organism growing on glucose and from an organism growing on acetate, significant regulatory differences will likely be observed between the two datasets. An example is pictured in the slides (Modeling Metabolism with Expression Data). $V_{\text{max}}$ tells us what the maximum we can reach. Microarray detects the level of transcripts, and it gives an upper boundary of $V_{\text{max}}$.

A visual of how the two sets of values compare is found in the slides (Modeling Metabolism with Expression Data).

In addition to predicting metabolic pathways under different environmental conditions, FBA and microarray experiments can be combined to predict the state of a metabolic system under varying drug treatments. For example, several TB drugs target mycolic acid biosynthesis. In a 2004 paper by Boshoff et al., researchers tested 75 drugs, drug combinations, and growth conditions to see what effect different treatments had on mycolic acid synthesis. In 2005, Raman et al. published an FBA model of mycolic acid biosynthesis, consisting of 197 metabolites and 219 reactions. The FBA model was used as illustrated in the slides (Experimental Approach). The basic flow of the prediction was to take a control expression value and a treatment expression value for a particular set of genes, then feed this information into the FBA and measure the final effect on the treatment on the production of mycolic acid. The results were fairly encouraging. Several known mycolic acid inhibitors were identified by the FBA. Interesting results were also found among drugs not specifically known to inhibit mycolic acid synthesis. 4 novel inhibitors and 2 novel enhancers of mycolic acid synthesis were predicted. One particular drug, Triclosan, appears to be an enhancer according to the FBA model, whereas it is currently “known” as an inhibitor. Further study of this particular drug would be interesting.

The slides showed why clustering may be ineffective in identifying function of various treatments. From the figure, we can see that known inhibitors, predicted inhibitors, and predicted enhancers of mycolic acid synthesis are not clustered
together. In addition, no labeled training set is required for FBA-based algorithmic classification, whereas it is necessary for supervised clustering algorithms. (I’m not sure about the reason…)

5. Predicting Nutrient Source
We can also predict nutrient source through predictions of metabolic state from expression data, because organisms likely adjust metabolic state to available nutrients. The expression data can give us readout of metabolic data. We can rank nutrients by how well they “match” a metabolic state.

Here is an example. Lun & Brandes use FBA to predict which nutrient source E. coli is growing on based on expression data. They make predictions by calculating the maximal flux-balance cone, and finding optimal fluxes for a set of potential nutrients. They then calculate the expression-constrained flux-balance cone and study the relative positions of the nutrients with respect to the new cone. Those nutrients which had optimal fluxes that remain within the expression-constrained flux-balance cone are likely sources; if no nutrients are left inside of the cone after addition of expression constraints. Lun & Brandes find the nutrient that is closest to the cone and label it as the most likely nutrient source for the given metabolic system.

Another example is TB. We know TB uses fatid acid, but don’t know which kind of fatid acid is used by TB. So we can use TB model to predict which fatid acid will be used by it.

There are also some other approaches to model the metabolic behaviors in cells.
Palsson et al. constructed context-specific metabolic networks. The gene will be turned off in this model if the expression value of this gene is lower than a specific threshold. Cabili et al. provided another way. First, they standardized FBA, and then maximized the number of enzymes whose predicted flux activity is consistent with their measured expression level.