Introduction to Steady State Metabolic Modeling

James Galagan
Genome Wide View of Metabolism

- Explore capabilities of global network
- Identify gaps in knowledge (missing genes)
- How do we go from a pretty picture to a model we can manipulate?
Metabolic Pathways

Metabolites
- glucose

Enzymes
- phosphofructokinase

Reactions & Stoichiometry
- 1 F6P => 1 FBP

Kinetics
- \[ v_{PFK} = \frac{v_{max}}{N_{PFK}} \frac{F6P}{K_{F6P}} \]

Regulation
- gene regulation
- metabolite regulation

From Mathews and van Holde: Biochemistry 2/e. © The Benjamin/Cummings Publishing Co., Inc.
Metabolic Modeling: The Dream

\[ \dot{X}_1 = -\beta_1 X_1^{h_1} X_2^{h_2} X_5^{h_3} \]

\[ \dot{X}_2 = \alpha_2 X_1^{h_1} X_2^{h_2} X_5^{h_3} - \beta_2 X_2^{h_2} ATP^{h_2,ATP} \]

\[ \dot{X}_3 = \beta_2 X_2^{h_2} ATP^{h_2,ATP} - \beta_3 X_3^{h_3} P_i^{h_3,PI} NAD^{h_3,NAD} \]

\[ \dot{X}_4 = 2\beta_3 X_3^{h_3} P_i^{h_3,PI} NAD^{h_3,NAD} + \alpha_4 X_5^{g_4} - \beta_4 X_4^{h_4} \]
Reaction Rates

\[ A + 2B \rightarrow 3C \]

**Formation rates**

\[ v_{fA} = \frac{d[A]}{dt} \quad v_{fB} = \frac{d[B]}{dt} \quad v_{fC} = \frac{d[C]}{dt} \]

**Reaction Rate = Reaction Velocity = Reaction Flux**

\[ v = \frac{d[A]}{dt} = \frac{1}{2} \frac{d[B]}{dt} = \frac{1}{3} \frac{d[C]}{dt} \]
Steady State Assumptions

• Dynamics are transient
• At appropriate time-scales and conditions, metabolism is in steady state

Two key implications
1. Fluxes are roughly constant
2. Internal metabolite concentrations are constant

\[
\frac{d[A]}{dt} = vin - v1 - v3 = 0
\]
Metabolic Flux

Input fluxes

Volume of pool of water = metabolite concentration

Output fluxes

Slide Credit: Jeremy Zucker
Reaction Stoichiometries Are Universal

The conversion of glucose to glucose 6-phosphate always follows this stoichiometry:

$$1\text{ATP} + 1\text{glucose} = 1\text{ADP} + 1\text{glucose 6-phosphate}$$

This is chemistry not biology.

Biology => the enzymes catalyzing the reaction

Enzymes influence rates and kinetics
- Activation energy
- Substrate affinity
- Rate constants

Not required for steady state modeling!
Metabolic Flux Analysis

Use universal reaction stoichiometries to predict metabolic network capabilities at steady state*


*Not precise, but more precision will come in later slides
Stoichiometry As Vectors

• We can denote the stoichiometry of a reaction by a vector of coefficients

• One coefficient per metabolite
  – Positive if metabolite is produced
  – Negative if metabolite is consumed

Example:

<table>
<thead>
<tr>
<th>Metabolites:</th>
<th>Reactions:</th>
<th>Stoichiometry Vectors:</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ A  B  C  D ]^T</td>
<td>2A + B -&gt; C</td>
<td>[ -2  -1  1  0 ]^T</td>
</tr>
<tr>
<td></td>
<td>C -&gt; D</td>
<td>[ 0  0  -1  1 ]^T</td>
</tr>
</tbody>
</table>
The Stoichiometric Matrix

Let $V$ be a vector of fluxes through each reaction

Then $S \cdot V$ is a vector describing the change in concentration of each metabolite per unit time

\[
\frac{dx}{dt} = S \cdot V
\]
A (Very) Simple System

We have introduced two new things

- **Reversible reactions** – are represented by two reactions that proceed in each direction (e.g. v4, v5)

- **Exchange reactions** – allow for fluxes from/into an infinite pool outside the system (e.g. vin and vout). *These are frequently the only fluxes experimentally measured.*
What Can We Use $S$ For?

From $S$ we can determine what combination of fluxes are possible in the system and what are not.

To get there we need three concepts:

1. **Nullspace**
2. **Extreme Pathways**
3. **Constrained Flux Space**
The Steady State Assumption and S

- We have $\frac{dx}{dt} = S \bullet V$

- But also recall that at steady state, metabolite concentrations are constant: $dx/dt=0$

$$\frac{dx}{dt} = S \bullet V = 0$$

Steady State fluxes are constrained to the nullspace of S
The Nullspace of S

- Subspace of flux vectors that do not change metabolite concentrations
- Can describe nullspace with non-unique basis vectors, $b_i$
- All nullspace fluxes are linear combinations of this basis:

$$V = \sum_i \alpha_i b_i$$

- Can find a basis using standard methods (e.g. SVD)
Example Nullspace Basis

$$\begin{array}{cccccc}
\text{v1} & \text{v2} & \text{v3} & \text{v4} & \text{vin} & \text{vout} \\
A & -1 & 0 & -1 & 0 & 1 & 0 \\
B & 1 & -1 & 0 & 0 & 0 & 0 \\
C & 0 & 1 & 0 & 1 & 0 & -1 \\
D & 0 & 0 & 1 & -1 & 0 & 0 \\
\end{array}$$

$$\begin{array}{cccc}
b1 & b2 \\
1 & 1 \\
1 & 1 \\
0 & -1 \\
0 & -1 \\
1 & 0 \\
1 & 0 \\
\end{array}$$

All steady state fluxes are combinations of b1 & b2

b2 includes negative fluxes that are not thermodynamically possible

-> Need to constrain the nullspace
Extreme Pathways

- The most fundamental constraint is that all fluxes must be positive*
- In this case, we have the following linear homogeneous equation system:

  \[ 0 = S \cdot V, \quad v_i \geq 0, \quad i = 1..n \]

- Solution to this set of equations is an exercise in convex analysis
- Solution region can be described by a unique set of Extreme Pathways

*recall that reversible reactions are represented by two unidirectional fluxes
The Flux Cone

Extreme pathways circumscribe a *convex flux cone*

- *Every* steady state flux vector, \( v \), is a *non-negative combination* of these pathways:

\[
V = \sum_{i} \alpha_i p_i \quad \alpha_i \geq 0
\]

- Extreme pathways represent underlying pathway structure of system
Constraining the Solution Space

• No reaction has capacity for *infinite* flux
• Often one can estimate constraints on transfer fluxes
  – *Max glucose uptake measured at maximum growth rate*
  – *Max oxygen uptake based on diffusivity equation*
• Flux constraints result in constraints on extreme pathways
  – Need enough constraints to ‘cover’ extreme pathways
The Constrained Flux Cone

- Contains all achievable flux distributions given the constraints:
  - Stoichiometry
  - Reversibility
  - Max and Min Fluxes

- Only requires:
  - Annotation
  - Stoichiometry
  - Small number of flux constraints (small relative to number of reactions)
Selecting One Flux Distribution

• At any one point in time, organisms have a single flux distribution

• How do we narrow down the range of predicted flux distributions (ideally to one)?

What if we assume organisms are trying to maximize a “fitness” function that is a function of fluxes?
Linear Programming

If we assume the objective function is a linear function of fluxes, we can use linear programming to find a solution.

Linear Programming
Maximize:
\[ z = \sum_{i} c_i v_i = c^T v \]

Subject to:
\[ Ax < b \]
\[ x \geq 0 \]

Solution always lies at boundary of admissible space
Can be found using simplex algorithm

Example

Constraints:
\[ x_i \geq 0 \]
\[ r_a \leq \text{const.} \]

Optimizing \textit{E. coli} Growth

For one gram of \textit{E. coli} biomass, you need this ratio of metabolites

Assuming a matched balanced set of metabolite fluxes, you can formulate this objective function

\[ Z = 41.257v_{ATP} - 3.547v_{NADH} + 18.225v_{NADPH} + 0.205v_{G6P} + 0.0709v_{F6P} + 0.8977v_{R5P} + 0.361v_{E4P} + 0.129v_{T3P} + 1.496v_{3PG} + 0.5191v_{PEP} + 2.8328v_{PYR} + 3.7478v_{AcCoA} + 1.7867v_{OAA} + 1.0789v_{AKG} \]
FBA Summary

**Stoichiometric Matrix**
Gene annotation
Enzyme and reaction catalog

**Feasible Space**
\[ S^*v = 0 \]
Add constraints:
\[ v_i > 0 \]
\[ \alpha_i > v_i > \beta_i \]

**Optimal Flux**
Growth objective
\[ Z = c^*v \]
Solve with linear programming

Next some applications of FBA....
APPLICATIONS
in silico Deletion Analysis

Can we predict gene knockout phenotype based on their simulated effects on metabolism?

Q: Why, given other computational methods exist? (e.g. protein/protein interaction map connectivity)

A: Other methods do not directly consider metabolic flux or specific metabolic conditions
in silico Deletion Analysis

“wild-type”

Gene knockouts modeled by removing a reaction
Mutations Restrict Feasible Space

- KO removes fluxes, and extreme pathways that depend on these fluxes

- Feasible space is constrained

- If original optimal flux is outside new space, new optimal flux is created

- Growth rate at new solution provides a measure of KO phenotype
The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities

J. S. Edwards* and B. O. Palsson†
Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093-0412
Communicated by Yuan-Cheng B. Fung, University of California, San Diego, La Jolla, CA, March 3, 2000 (received for review October 14, 1999)

PNAS| May 9, 2000 | vol. 97 | no. 10

Model of *E. coli* central metabolism
436 metabolites
720 reactions

Simulate mutants in glycolysis, pentose phosphate, TCA, electron transport

Edward & Palsson (2000) PNAS
E. coli KO simulation results

If $Z_{\text{mutant}}/Z = 0$, mutant is no growth (-), growth (+) otherwise

Compare to experiment (in vivo / in silico)
86% agree

Measured optimal growth of mutants ($Z_{\text{mutant}}$) versus non-mutant ($Z$)

Condition specific prediction

Edward & Palsson (2000) PNAS
What do the errors tell us?

- Errors indicate gaps in model or knowledge
- Authors discuss 7 errors in prediction
  - *fba* mutants inhibit stable RNA synthesis (not modeled by FBA)
  - *tpi* mutants produce toxic intermediate (not modeled by FBA)
  - 5 cases due to possible regulatory mechanisms (*aceEF, eno, pfk, ppc*)
Regulation means genes are expressed at different levels.

And we can measure this…
Coupling Expression with Metabolism

Can We Algorithmically Interpret Expression Data in a Metabolic Context?

Caroline Colijn, Aaron Brandes, Jeremy Zucker, Brian Weiner, Desmond Lun
Interpreting Array Data in Metabolic Context

Growth on Glucose

Clustering, GSEA

Model Flux Distribution From Expression?

Kegg, PathwayExplorer
What is the relationship between gene level and flux through a reaction?
Applying Flux Balance Analysis, we use expression data to model the maximum flux through each reaction.
Modeling Metabolism with Expression Data

Applying Flux Balance Analysis (FBA), we use expression data to model the maximum flux through each reaction.
Mycolic Acid Biosynthesis

- Major cell wall constituent
  - Antibiotic permeability
  - Intracellular growth

- Target of several first-line TB drugs

- Published FBA Model
  - 197 Metabolites
  - 219 Reactions
Boshoff TB Expression Compendium

75 drugs, drug combinations, and growth conditions

436 Experiments

Spotted array (Cy3 control, Cy5 experiment)

<table>
<thead>
<tr>
<th># samples</th>
<th>category</th>
</tr>
</thead>
<tbody>
<tr>
<td>103</td>
<td>respiration</td>
</tr>
<tr>
<td>73</td>
<td>cell wall synthesis</td>
</tr>
<tr>
<td>58</td>
<td>unclassified</td>
</tr>
<tr>
<td>50</td>
<td>DNA integrity</td>
</tr>
<tr>
<td>44</td>
<td>aromatic amides</td>
</tr>
<tr>
<td>36</td>
<td>iron scavengers</td>
</tr>
<tr>
<td>26</td>
<td>pH</td>
</tr>
<tr>
<td>20</td>
<td>protein synthesis</td>
</tr>
<tr>
<td>10</td>
<td>carbon source</td>
</tr>
<tr>
<td>10</td>
<td>transcription</td>
</tr>
<tr>
<td>6</td>
<td>nonreplicating persisting stage 1</td>
</tr>
</tbody>
</table>

Can we predict the impact of each drug/condition on mycolic acid synthesis (capacity)?
Experimental Approach

Control (Cy3)  

Drug (Cy5)  

FBA Mycolic Acid Flux Prediction

Log fold change mycolic acid capacity (arbitrary units)

Effect of Drug

Increased

Decreased

Optimize mycolic acid production
Significance and Specificity

**Significance**
- Compare pairs of controls
  - Would we see this effect due to noise in expression w/o drug?
- Resample controls and experiments
  - Is the effect due to noise between control and drug?

**Specificity**
- Randomize gene labels
  - Is the effect specific to mycolic acid?
  - Or due to overall suppression (enhancement) of metabolism?
Predicted Inhibitors and Enhancers

<table>
<thead>
<tr>
<th>Predicted Inhibitors</th>
<th>Predicted Enhancers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isoniazid Specific Strong Known</td>
<td>Menadione/GSNO Non-specific Weak</td>
</tr>
<tr>
<td>Thiolactomycin Specific Strong Known</td>
<td>Chlorpromazine/GSNO Specific Weak* New</td>
</tr>
<tr>
<td>Ethambutol Specific Strong† Known</td>
<td>Rifampicin Non-specific Weak New</td>
</tr>
<tr>
<td>Ethionamide Specific Strong Known</td>
<td>Triclosan Non-specific Weak* Incorrect?</td>
</tr>
<tr>
<td>PA-824 Non-specific Weak Known</td>
<td>GSNO Specific Strong* New</td>
</tr>
<tr>
<td>Cerulenin Specific Weak‡ Known</td>
<td></td>
</tr>
<tr>
<td>PA-21 Specific Weak New</td>
<td></td>
</tr>
<tr>
<td>Streptomycin Non-specific Weak New</td>
<td></td>
</tr>
<tr>
<td>Valinomycin Non-specific Weak New</td>
<td></td>
</tr>
<tr>
<td>ZnSO4 Non-specific Strong New</td>
<td></td>
</tr>
</tbody>
</table>

*Experimental Testing and Validation in Progress

Tan-Yun Cheng, Branch Moody
Brigham and Women’s Hospital

*Prediction made only for certain replicates. †Prediction made only for certain doses.
More Than Expression Classification

Fatty Acid Synthesis Genes (29)

Drugs (437)

No labeled training set is required

- Predicted Enhancer
- Predicted Inhibitor
- Known Inhibitor

- Triclosan (10, 50, 100, 150 µg/ml)
- Rifampicin
- PA-824
- Valinomycin
- PA-21
- Ethambutol
- ZnSO₄
- Menadione-GSNO
- Rifampicin
- PA-824
- Streptomyacin
- PA-824
- Streptomyacin
- Thiolactomycin
- Isoniazid/Ethionamide

- Chloropromazine-GSNO
- Streptomycin
- PA-824
- Valinomycin
- PA-21
- Ethambutol
- ZnSO₄
- Menadione-GSNO
- Rifampicin
- PA-824
- Streptomyacin
- Thiolactomycin
- Isoniazid/Ethionamide
Predicting Nutrient Source

Predict nutrient source through predictions of metabolic state from expression data

• Organisms likely adjust metabolic state to available nutrients
• Expression data gives us a readout of metabolic state

Rank nutrients by how well they “match” a metabolic state
Geometric Interpretation

Applying expression constraints reshapes the flux cone

Growth on Glucose

Growth on Acetate
Nutrient 1 is a better fit for the expression data since the predicted metabolic state allows for a flux closer to the maximum optimal flux.
E. coli Nutrient Prediction

Apply expression data from growth on known carbon sources

Can we predict carbon source from expression and metabolism?
E. coli Nutrient Prediction

Color indicates distance from expression-constrained flux cone to optimal flux for that nutrient combination.
Ranking Multiple Nutrients
Other Approaches

  - If gene expression lower than some threshold, turn the gene off in the model.

  - Nested optimization problem
  - First, standard FBA
  - Second, maximize the number of enzymes whose predicted flux activity is consistent with their measured expression level
Resources

• Tools and Databases
  – Kegg
  – BioCyc
  – PathwayExplorer (pathwayexplorer.genome.tugraz.at)

• Metabolic Modeling
  – Palsson’s group at UCSD (http://gcrg.ucsd.edu/)
  – www.systems-biology.org
  – Biomodels database (www.ebi.ac.uk/biomodels/)
  – JWS Model Database (jjj.biochem.sun.ac.za/database/index.html)