From Genotypes to Phenotypes: Reconstruction & Mathematical Analysis of Metabolic Networks

Jennifer Reed, Ph.D.
Faculty Fellow at UCSD

Winter School in Genomics 2006
Simulating Phenotypes

**Model**

\[
S = \begin{pmatrix}
0 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & 1 & 0 & -1 \\
1 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\
-1 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\
0 & -2 & 1 & 0 & 0 & 0 & 0 & 0 & -2 & -1 & 0 \\
0 & 0 & 0 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 1 & -1 & 1 & 0 & 0 & 0
\end{pmatrix}
\]

\[
S \cdot v = 0
\]

\[
v_{min,i} \leq v_i \leq v_{max,i}
\]
Genome-scale Metabolic Model Reconstruction

- Genome Annotation
  - by homology, location

- Biochemical Data
  - protein characterized

- Physiological Data
  - indirect, pathway known

- Inferred Reactions
  - indirect, inferred from biomass requirements

- Quantitative Analysis
  - simulate cell behavior
  - drive experimental studies

Network Reconstruction

Metabolic Model

New Predictions
Emergent Properties
Model Development: an iterative process

- Biochemical data
- Revised ORF assignments

Computational, Biochemical Investigation

Genome-scale Metabolic Model Reconstruction

- ORGANISM
- Network Reconstruction
- Metabolic Model
- New Predictions
- Emergent Properties

Quantitative Analytical Methods

Inferred Reactions

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu
Part I: Metabolic Network Reconstruction
Escherichia coli Metabolism

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu
What is in a metabolic reconstruction?

**Genome:**
- Annotated genes
- Gene location
- Regulatory regions
- Wobble base pairs

**Biochemistry:**
- Stereochemistry
- pH and pKa (charge)
- Elemental balance
- Charge balance
- Multiple reactions/enzyme
- Multiple enzymes/reaction

**Physiology:**
- Flux data
- Knock-outs
- Balanced functions
- Overall phenotypic behavior
- Location of gene product compartmentalization
Defining Metabolic Reactions

Lactate Dehydrogenase

1st level: Metabolite Specificity
Primary metabolites
- LAC
- PYR

2nd level: Metabolite Formulas
Neutral Formulas
- \( \text{C}_2\text{H}_4\text{O}_3 \)
- \( \text{C}_3\text{H}_6\text{O}_3 \)

Charged Formulas
- \( \text{C}_2\text{H}_5\text{O}_3^- \)
- \( \text{C}_3\text{H}_5\text{O}_3^- \)

3rd level: Stoichiometry
- \( 1 \text{ LAC} + 1 \text{ NAD} \quad ? \quad 1 \text{ PYR} + 1 \text{ NADH} + 1 \text{ H} \)

4th level: Thermodynamic Considerations: Directionality
- \( 1 \text{ LAC} + 1 \text{ NAD} \leftrightarrow 1 \text{ PYR} + 1 \text{ NADH} + 1 \text{ H} \)

5th level: Localization
- prokaryotes:
  - [c]: cytoplasm
  - [e]: extracellular
  - [p]: periplasm
- eukaryotes:
  - [n]: nucleus
  - [g]: golgi apparatus
  - [v]: vacoule
  - [l]: lysosome
  - [r]: endoplasmic reticulum

STEPWISE INCORPORATION OF INFORMATION

University of California, San Diego
Department of Bioengineering

http://systemsbiology.ucsd.edu
Sources of Information

- Genome Sequence & Annotation
- Physiological Data
- Databases
  - Kyoto Encyclopedia of Genes and Genomes (KEGG)
- Localization
  - Signal sequences
    ...PLLPLISGSALP...

Available Literature

Identification of the *Escherichia coli* K-12 ybE Gene as pgp,
Encoding 6-Phosphogluconactonase
Lynn C. Thomason, Donald L. Grant, Alan E. Davis, Rino Khurana, and Judith L. Bremer

Phylogenetic Data

- Bacteria
- Archaea
- Eukarya

Physiological Data

Growth Measurements

<table>
<thead>
<tr>
<th>Time (Hours)</th>
<th>OD 600</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>4</td>
<td>0.2</td>
</tr>
<tr>
<td>9</td>
<td>0.3</td>
</tr>
<tr>
<td>14</td>
<td>0.4</td>
</tr>
<tr>
<td>19</td>
<td>0.49</td>
</tr>
<tr>
<td>24</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Univ.
Department of Bioengineering

http://systemsbiology.ucsd.edu
Gene: glk

Enzyme: Glucokinase

Reaction:
ATP + D-Glucose →
ADP + D-Glucose 6-phosphate

E.C.: 2.7.1.1

Human β-cell glk http://www.rcsb.org/pdb/
Biochemical Literature: Curation and Expansion of the Network

*H. pylori* Glycolysis according to KEGG:

\[
\text{Glucose} \rightarrow \text{G-6-P} \rightarrow \text{F-6-P} \rightarrow \text{FDP}
\]

*H. pylori* Glycolysis according to Hoffman *et al.* (1996):

\[
\text{Glucose} \rightarrow \text{G-6-P} \rightarrow \text{F-6-P} \rightarrow \text{FDP}
\]
Physiological Data and Inferred Reactions

Filling in the Gaps based on indirect evidence

KEGG Glycolytic Pathway for *Methanosarcina barkerii*:
No Orthologs for EC 2.7.1.11 or 1.2.1.12
### Filling in the Gaps – an Example

- Experiments determine which amino acids are taken up by *H. pylori* vs. which can be produced *in vivo*.
- Missing steps of amino acid biosynthesis are added if necessary on the basis of this physiological evidence.

#### Amino Acid Requirements

<table>
<thead>
<tr>
<th>AA</th>
<th>Reynolds</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Arg</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Asn</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Asp</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Cys</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Gln</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Glu</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Gly</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>His</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ile</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Leu</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lys</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Met</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Phe</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pro</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Ser</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Thr</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Trp</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Tyr</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Val</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*in vivo* | *in silico*
Inferred Reactions

• Some reactions are included based on indirect physiological evidence (by inference)
  – Assumption: the cell must be able to produce all biomass components to grow
  – Reactions are added if necessary
  – Generally transporters, etc.
  – Most tentative; should be examined more carefully
Reaction Confidence: Sources of Evidence

- **Biochemical**
  Enzyme has been tested biochemically.
- **Genetic**
  Gene overexpression and purification, gene deletions.
- **Sequence**
  There is significant sequence similarity to another gene with known function.
- **Physiological**
  There is physiological data to support inclusion in the model.
- **Modeling**
  Reaction is included to improve simulation results.

![Model Reaction Properties Table]

- Kinetic Assay
- Overexpression
- 86% Homology
- Grows on Ascorbate
Network Assembly and Representation

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Glycolytic reactions</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEX1</td>
<td>[c]GLC + ATP $\rightarrow$ G6P + ADP + H</td>
<td>glk</td>
</tr>
<tr>
<td>PGI</td>
<td>[c]G6P $\leftrightarrow$ F6P</td>
<td>pgi</td>
</tr>
<tr>
<td>PFK</td>
<td>[c]ATP + F6P $\rightarrow$ ADP + FDP + H</td>
<td>pfkA, pfkB</td>
</tr>
<tr>
<td>FBA</td>
<td>[c]FDP $\leftrightarrow$ DHAP + G3P</td>
<td>fbaA, fbaB</td>
</tr>
<tr>
<td>TPI</td>
<td>[c]DHAP $\leftrightarrow$ G3P</td>
<td>tpiA</td>
</tr>
<tr>
<td>GAPD</td>
<td>[c]G3P + NAD + Pi $\rightarrow$ 13DPG + H + NADH</td>
<td>gapA, gapC1, gapC2</td>
</tr>
<tr>
<td>PGK</td>
<td>[c]13DPG + ADP $\rightarrow$ 3PG + ATP</td>
<td>pgk</td>
</tr>
<tr>
<td>PGM</td>
<td>[c]3PG $\leftrightarrow$ 2PG</td>
<td>gpmA, gpmB</td>
</tr>
<tr>
<td>ENO</td>
<td>[c]2PG $\leftrightarrow$ H2O + PEP</td>
<td>eno</td>
</tr>
<tr>
<td>PYK</td>
<td>[c]ADP + H + PEP $\rightarrow$ ATP + PYR</td>
<td>pykA, pykF</td>
</tr>
</tbody>
</table>

Reed, et al. Nat. Reviews Genetics, 2006
**E. coli** Gene-Protein-Reaction (GPR) Associations

**Enolase**
- 1 gene
- 1 reaction
- (342/904 Genes are one-to-one)

**Succinate Dehydrogenase**
- 4 subunits
- 2 reactions
- (135/904 Genes interact as Subunits)

**Xylose ABC Transporter**
- Protein complex
  - 3 proteins
- 1 reaction
- (153/904 Genes interact as PCs)

**Pyruvate Kinase**
- 2 isozymes
- 1 reaction
- (149/931 Reactions have multiple isozymes)

University of California, San Diego
Department of Bioengineering

Enolase

1 gene
1 reaction

(342/904 Genes are one-to-one)

**Succinate Dehydrogenase**

4 subunits
2 reactions

(135/904 Genes interact as Subunits)

**Xylose ABC Transporter**

Protein complex
- 3 proteins
1 reaction

(153/904 Genes interact as PCs)

**Pyruvate Kinase**

2 isozymes
1 reaction

(149/931 Reactions have multiple isozymes)
Network Evaluation

Reed, et al. Nat. Reviews Genetics, 2006
Where to Begin?

1. Choose your organism
   - Annotated Genome
   - Well-characterized (textbook and/or reviews), higher SKI values
   - Physiological studies
   - Relative importance—some reconstructions will generate more interest than others

2. Download annotated genome
   - TIGR or other website

3. Build network
   - Section by section
   - Test frequently
   - Check all genes to determine inclusion or exclusion
   - Annotate your reconstruction with confidence level, etc.

4. Iterative model-building
   - Test network capabilities against physiological data
Part II:
Constraint-Based Modeling for Calculating Phenotypes
From Component to Systemic Annotation: Fundamental to Systems Biology

Genome sequence

Component interactions

Component annotation

Genome sequence

Component interactions

Genome sequence

Component interactions

Regulatory interactions

Direct gene products

Effects on other compounds

-2
-1
+1

-2
-3
+1

-1
+1
Stoichiometric Coefficients

- Typically integral numbers
- Universal biochemical constants and are time invariant

**Chemical reaction:**  \((a)A + (c)C \xrightarrow{v_i} (e)E + (h)H\)

**Representation as a column in a matrix:**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-a</td>
<td>0</td>
<td>-c</td>
<td>0</td>
<td>+e</td>
<td>0</td>
<td>0</td>
<td>+h</td>
</tr>
</tbody>
</table>

|   |   |   | DNA-protein |   | protein-protein |   |   |
|---|---|----------------|---|----------------|---|---|
|   |   | -1             |   | +1             |   |   |
| A |   | -1             |   | +1             |   |   |
|   |   | +2             |   | +1             |   |   |
| B |   | +3             |   | +1             |   |   |
|   |   | +1             |   | +1             |   |   |

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu
Constraint-Based Analysis

How often have I said to you that when you have eliminated the impossible, whatever remains, however improbable, must be the truth?

–Sherlock Holmes, A Study in Scarlet

University of California, San Diego
Department of Bioengineering
Metabolic Constraints

1. Steady-State Mass Balance Constraints

For each metabolite \( X \): 
\[
\sum v_{\text{produce}_x} = \sum v_{\text{consume}_x}
\]

For all metabolites this can be represented as:

\[
S \cdot v = 0
\]

2. Enzyme Capacity Constraints: \( \alpha \leq v_j \leq \beta \)

3. Thermodynamic Constraints: \( v_j \geq 0 \)
Abstract | Microbial cells operate under governing constraints that limit their range of possible functions. With the availability of annotated genome sequences, it has become possible to reconstruct genome-scale biochemical reaction networks for microorganisms. The imposition of governing constraints on a reconstructed biochemical network leads to the definition of achievable cellular functions. In recent years, a substantial and growing toolbox of computational analysis methods has been developed to study the characteristics and capabilities of microorganisms using a constraint-based reconstruction and analysis (COBRA) approach. This approach provides a biochemically and genetically consistent framework for the generation of hypotheses and the testing of functions of microbial cells.
Constraint-Based Analysis Methods

Optimal Solutions
1. Flux Balance Analysis
2. Flux Variability

Flux Dependencies
1. Robustness
2. Phase Planes
3. Flux Coupling

All Allowable Solutions
1. Extreme Pathways
2. Elementary Modes
3. Sampling

Altering Genotypes
1. Genetic Mutations
2. Strain Design

Application of Additional Constraints
1. Regulation
2. Energy Balance


University of California, San Diego Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu
Optimization

- **Objective Function:** A function that is maximized or minimized to identify optimal solutions.

- **Constraints:** Place limits on the allowable values the solutions can take on.

Maximize: $f(x)$

Such that $g(x) = a$

$h(x) \geq b$
How does LP work?
A 1D example

The solution space is the line of admissible in the positive orthant.

If we maximize ATP production the solution lies on the x-axis where all the flux would be through reaction \( x_1 \). Conversely, maximizing NADH production would give the point at the y-axis, where only reaction \( x_2 \) is active.

Note that the optimal solutions lie at the boundary of the admissible space.

An Illustrative Example

Consider two variables A and B, which are the amount of toy cars and trucks you can produce.

Do to resource limitations you can make no more than 60 cars a day and no more than 50 trucks a day.

\[ 0 < A < 60 \]
\[ 0 < B < 50 \]

You are also limited by shipping such that the number of cars plus twice the number of trucks must be less than 120.

\[ A + 2B < 150 \]

You can sell the toys at $20/car and $30/truck your earnings (Z) are given by:

\[ Z = 20A + 30B \]
Graphical Representation of Feasible Solution Space

3 Constraints:
0 < A < 60
0 < B < 50
A + 2B < 150
A + 2B = 150

Feasible Solution Space

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu
Graphical Representation of the Objective Function: \( Z = 20A + 30B \)

Optimal value within feasible set

Feasible Solution Space

- \( Z = 2550 \)
- \( Z = 2500 \)
- \( Z = 1500 \)
Types of solutions: the impact of the objective function

- **Single solution**
- **Degenerate solution**
- **No solution**

Optimal solution in a corner

Optimal solution along an edge

Optimal solution not found--region unbounded

Lines of constant Z
Determining optimal or best states

Flux balance analysis
Enumeration of alternate optima
Identification of objective functions
Flux variability analysis
Determining Optimal States

- Increase in Stated Objective Function
- Flux Variability for $V_1$
- Alternate or Equivalent Optima
- Experimental measurement
- Potential Objective Functions

$V_2$
$V_1$
1. Linear Programming: Finding an Optimal Solution

**Biological Significance:** Identifies a theoretical maximum for a defined cellular objective

**Example:** Optimizing for growth in *Escherichia coli* led to correct prediction of the endpoint of an adaptive evolution [Ref]

**Mathematics**

Maximize $c^T \cdot v$

$S \cdot v = 0$

$v_{i,\text{min}} \leq v_i \leq v_{i,\text{max}}$

*References*

Domach '90, Edwards NBT, Ibarra Nature
7. Multiple Alternative Optima
Multiple Integer Linear Programming (MILP)

**Defined Cellular Objective**

**Mathematics**

**Biological Significance**: Multiple alternative optima suggests that cells have many ways of achieving an objective. This relates to silent phenotypes, robustness, and redundancy.

**Example**:

Used to identify and analyze many alternative optimal solutions through a reconstructed genome-scale network of *E. coli* metabolism.

**Mathematics**

(a) Solve the master problem:

\[
\begin{align*}
\min Z^k &= \alpha^T z \\
\text{s.t.} \quad Bz &= q \\
\sum_{i \in N^k} w_i &\leq |NZ^k| - 1, \quad k = 1, 2, ..., K - 1 \\
0 &\leq z_i \leq U_w, \quad i \in I \\
y_i + w_i &\leq 1, \quad i \in NZ^{k-1} \\
z &\geq 0
\end{align*}
\]

(b) Define set $NZ^K$ and continue until $Z^K > (Z^1)^*$.

**References**


Dasika MS, Gupta A, Maranas CD. Pac Symp Biocomput. 2004;:474-85
10. Identifying Candidate Cellular Objectives

Calculating the cone of possible objective functions

**Biological Significance**: Given an experimentally measured cell state, calculates range of possible objectives for which the cell could be optimizing

**Example**: Calculating potential objectives for *Escherichia coli* led to showed that optimal growth was a candidate objective function [Burgard and Maranas]

**Mathematics**

Minimize: \( c_j \sum_{j \in E} (v_j - v_{j,exp})^2 \)  

Subject to:

Maximize: \( \sum_{j \in P} c_j v_j \)

Subject to: \( \sum_{j=1}^{M} S_j v_j = 0, \quad \forall i \in N \)

\( v_{GLC} = \text{uptake}, \quad \forall j \in \text{glucose uptake} \)

\( v_j \geq 0, \quad \forall j \in M \)

\( \sum_{j \in P} c_j = 1 \)

\( c_j \geq 0, \quad \forall j \in P \)

**References**

11. Flux Variability

**Mathematics:**
\[
\text{max } v_i \text{ or min } v_i
\]
such that \( S \cdot v = 0 \)
\[
0 \leq v \leq v_{max}
\]
\[
c^T \cdot v = Z \text{ (optional)}
\]

**Biological Significance:** Can determine for some fixed fluxes (e.g. growth rate) what the range is for all other fluxes in the network.

**Example:** For the *E. coli* metabolic network, 3% of the metabolic fluxes can vary and still allow for optimal biomass production on glucose [Mahadevan MetabEng].

**Key References**
Quantifying Flux Dependencies

Robustness analysis
Phenotypic phase plane analysis
Flux coupling
Parameter Variation

Robustness Analysis:
Projection of PhPP for Maximum Growth rate vs. O$_2$ uptake

Phenotypic Phase Plane (PhPP)

Robustness Analysis:
Projection of PhPP for Maximum Growth rate vs. Succinate uptake

Biomass Production vs. O$_2$ uptake

Biomass Production vs. Succinate uptake

Line of Optimality (LO)
2. Robustness to Gene Deletions and Enzyme Defects

**Biological Significance:**
The impairment of an enzyme can have a system wide effect and affect the optimal growth rate achievable by an organism.

**Example:**
Fluxes in E. coli have been analyzed to study how a continuous impairment of the enzyme will affect the predicted optimal growth rate.

**Mathematics**

Maximize/Minimize $Z = c_j v_j$

subject to $S y_j = 0$, $\alpha_j \leq v_j \leq \beta_j$

**References**

**Example**

In this example we vary the maximum allowable uptake rate of oxygen. The whole range of oxygenation is shown, from fully aerobic conditions to fully anaerobic conditions.

The growth rate is graphed in the upper panel and the by-product secretion rates in the lower.
Shadow prices can be used to interpret the changes in the optimal flux distribution.
6. Phenotypic Phase Planes

**Mathematics:** Shadow prices from the dual solution are calculated for different uptake rates. Shadow prices are constant within a region, changes in shadow prices delineate the different regions.

**Biological Significance:** Can determine what the optimal nutrient uptake rates to allow for maximal biomass production (Line of Optimality) and what uptake rates are not feasible.

**Example:** Comparison of experimentally measured uptake rates shows that *E. coli* uses its metabolic network to maximize biomass for some carbon sources (operates along the line of optimality) [Edwards NBT, Ibarra Nature]

---

**Key References**


Phenotypic Phase Plane Characteristics

- Uptake B
- Uptake A

- Single Substrate Limitation
- Dual Substrate Limitation
- "Futile" Region
- LO

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu
Hypothesis:
Metabolic regulation will drive the operation of the metabolic network toward the line of optimality.

Lines of constant growth rate
‘isoclines’

Acetate-Oxygen Phenotype Phase Plane

Hypothesis: Metabolic regulation will drive the operation of the metabolic network toward the line of optimality

Lines of constant growth rate
‘isoclines’

Acetate-Oxygen Phenotype Phase Plane

Hypothesis: Metabolic regulation will drive the operation of the metabolic network toward the line of optimality.

Lines of constant growth rate
‘isoclines’
Acetate Phase Plane: Experimental Data

\[ y = 0.9975x + 1.8663 \]

\[ R^2 = 0.8423 \]
Cellular Evolution: Growth rates on Glycerol


University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu
14. Flux Coupling

**Biochemical Significance:** Can determine how fluxes are related to other fluxes, these can be related to co-regulated gene sets.

**Example:** A higher percentage of reactions belong to coupled reaction sets in *H. pylori* than *E. coli* and *S. cerevisiae* indicating that these larger metabolic networks are more flexible (Burgard, Genome Research)

**Mathematics:**

\[
\text{maximize } R_{\text{max}} = \hat{v}_1 \quad \text{(or minimize)}
\]

subject to

\[
\sum_{j=1}^{M} S_{ij} \hat{v}_j = 0
\]

\[
\hat{v}_2 = 1
\]

\[
\hat{v}_j \text{uptake} \leq \hat{v}_j \text{uptake_max} \cdot t
\]

\[
\hat{v}_j \geq 0
\]

\[
t \geq 0
\]

where \(\hat{v}\) are the fluxes normalized by \(v_2\)

**Key References**

Characterizing the Whole Solution Space (Unbiased Methods)

Elementary mode analysis
Extreme pathway analysis
$\alpha$-spectrum
Random sampling
Characterizing the Whole Solution Space

Any steady state flux distribution, \( \mathbf{v} \), can be decomposed into the extreme pathways:

\[
\mathbf{v} = \alpha_1 \mathbf{P}_1 + \alpha_2 \mathbf{P}_2
\]

For the point shown, the \( \alpha \)-spectrum would be:

- Uniform random samples in the solution space
- Any steady state flux distribution, \( \mathbf{v} \), can be decomposed into the extreme pathways

\[
\begin{bmatrix}
\alpha_1 \\
\alpha_2
\end{bmatrix}
\]
4. Network-Based Biochemical Pathways

Extreme Pathways and Elementary Modes

**Biological Significance:**

Extreme pathways and elementary modes provide a set of biochemically valid pathways into which all possible steady states can be decomposed.

**Example:**

Elementary modes and extreme pathways have been used to characterize the metabolic potential of many systems, including the red blood cell, core E. coli, H. pylori and H. influenzae amino acid synthesis, core yeast metabolism and more.

**Mathematics**

\[ S \cdot v = 0 \]

\[ V_{i,\text{min}} \leq v_i \leq V_{i,\text{max}} \]

\[ C = \{ v: v = \Sigma \alpha_i p_i \land i \} \]

---

**References**


12. $\alpha$- Spectrum

**Biological Significance:** Given the set of extreme pathways and a flux distribution within the solution space, can determine what convex combination of pathways can lead you to that solution.

**Example:** For the human red blood cell only 14 of the 39 extreme pathways are needed to reconstruct the nominal steady state solution, and for two of these pathways the $\alpha$ weightings are fixed [Wiback JTB].

**Mathematics:**

\[
\text{max } \alpha_i \text{ or min } \alpha_i \\
\text{such that } P \cdot \alpha = v \\
0 \leq \alpha_i \leq 1
\]

**Key References**

Altering phenotypic potential

Gene Knockout predictions: FBA, MoMA, ROOM
Strain Design: OptKnock
Regulatory Considerations
Altered Solution Spaces

Metabolic Model

Wildtype Solution Space
Knockout Solution Space

Growth Rate

Glucose Uptake Rate

FBA
MOMA
5. Predicting Knockout Phenotypes
Minimization of Metabolic Adjustment (MOMA)

Biological Significance: Hypothesizes that knockout strains are not evolved to optimality and thus will use their metabolic networks ‘as similar as possible’ to the wildtype

Example: Comparison of experimentally measured flux distributions in mutant strains of E. coli were better predicted using MOMA than with FBA (LP) [Segre PNAS]

Mathematics
Minimize $L \cdot x + \frac{1}{2} x^T Q x$

References
Segre D, Vitkup D, Church GM. Proc Natl Acad Sci U S A. 2002 Nov 12;99(23):15112-7
Calculating Intracellular Fluxes and Growth Rates
(FBA and MOMA)

\[ \Delta_{\text{pyk}} \text{ FBA} \quad P = 6.0 \times 10^{-1} \]

\[ \Delta_{\text{pyk}} \text{ MOMA} \quad P = 7.4 \times 10^{-3} \]

Segre, D., Vitkup, D., and Church, G.M. 

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu
Deletion Strain Evolution

- 39 of 50 cases correctly predicted computationally
- Parallel cultures exhibit similar endpoint phenotypes
- Average GR increase of 87% observed

Fong & Palsson, Nat. Gen 2004
Computational Design of Mutant Strains

OptKnock: Find gene deletions needed such that maximizing biomass is coupled with maximizing metabolic engineering objective

Strain Designs for:
- Lactate Production
- Succinate Production
- 1,3 Propanediol Production
- Chorismate Production
- Alanine Production
- Serine Production
- Aspartate Production
- Glutamate Production

REFERENCES

OptKnock Design for Lactate pta-adhE evolution (37°C)


University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu
Regulation: Shaping the Space

1. Temporary, adjustable constraints
2. Flux(es) constrained to zero
3. Extreme vector(s) removed
4. Dimension or volume of space is reduced

No ‘green’: solution achievable but solution space smaller

No ‘red’: solution not achievable

Covert, Schilling and Palsson, JTB 2001
Regulation w/FBA: Overall Method

Reconstruction:
Databases
Literature

Constraints:
Mass Balance
Sv = 0

Capacity
\( \alpha_i < v_i < \beta_i \)

Dynamics:
Quasi Steady-State Assumption

Integration

Time course of growth (phenotype)

General Solution Space

Solution Shifts

Extreme Pathways

Solution Shifts

New Growth Behavior

External Signal
Regulatory Proteins
Transcriptional Regulation
Altered Network Capabilities
Restricted Solution Space

Systems Biology Research Group
http://systemsbiology.ucsd.edu
Regulatory Network: Boolean Representation

Less connected than metabolic networks (all connections shown!)

**Regulatory Network Components**

Stimuli: **102**

Transcription Factors: **104**

Metabolic Genes: **479** (out of 906)
Iterative Model Building as for ‘Large-Scale’ Hypothesis Verification and Data Integration
Model Driven Discovery Via High Throughput Testing

IN SILICO MODEL

EXPERIMENTAL MEASUREMENTS

Comparisons:
Model vs. Experiment
eg. Minimal Media
WT and ∆ Growth Capabilities
Gene Expression

Model Improvement & Refinement

Experimental Design

Interpretation

INCONSISTENCIES
Can be used to design experiments and update the model

CONSISTENCIES
Can be used to help interpret experimental observations

Growth Data Comparisons: Two Failure Modes
i. Predicted Growth but NO Experimental Growth
   - Missing Regulation or Falsely Included Reactions
ii. Experimental Growth but NO Predicted Growth
   - Missing metabolic transport or enzymatic reactions
Example: Thymidine can be used as a sole carbon source, but metabolic transformations enabling use of this compound are uncharacterized in this strain.

Example: Hypothesis generation: transcriptional regulation in *E. coli*

Step one: Reconstruct computational model based on available data

Step two: Compare new observations to computational predictions

Step three: Expand model via hypothesis generation
Regulatory Network Interrogation

- Gene Chip
  - Wt + O2
  - Wt - O2
  - ΔArcA ΔFNR
  - ΔArcA ΔFNR

- Model
  - a
  - b
  - c
  - d
  - e
  - f
  - g
  - h

- Non-model
  - ArcA
  - FNR

- Discrepancies
  - ΔArcA
  - ΔFNR

- Improved Model
  - ArcA: if not O2
  - FNR: if not O2
  - a: If not FNR
  - b: If ArcA
  - c: If PdhR or FNR
  - d: If not ArcA

- Regulatory Network Perturbation Analysis (Ideker et al., Science 2001)
  - Generate knockout strains (Datsenko and Wanner, 2000)
  - Determine transcriptome (Gene chips, qPCR)
  - Iteratively improve model

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
http://systemsbiology.ucsd.edu
Model-Centric Hypothesis Generation

- Genome-scale regulatory/metabolic model of *E. coli*
  - 1,008 genes
- Systematic network perturbation analysis
  - ArcA, Fnr, ArcA/Fnr, AppY, OxyR, SoxS
- Generate new rules for model
- Hypotheses generation
Model-driven hypothesis generation

**Step one:** Reconstruct computational model based on available data

- **iMC1010v1**
  - Phenotypic Predictions
    - 79% (10828/13750) accuracy
  - Expression Predictions
    - 49% (23/47) accuracy
    - 15% (23/151) coverage

**Step two:** Compare new observations to computational predictions

- **Gene Expression Study**
  - Added new rules for 78 genes
  - Removed old rules for 27 genes
  - Changed old rules for 10 genes
  - Total of 115 changes in regulatory rules

**Step three:** Expand model via hypothesis generation

- **iMC1010v2**
  - Phenotypic Predictions
    - 79% (10833/13750) accuracy
  - Expression Predictions
    - 98% (100/102) accuracy
    - 66% (100/151) coverage

**110 new regulatory hypotheses overall**

University of California, San Diego
Department of Bioengineering

Covert et al, Nature 2004

Systems Biology Research Group
http://systemsbiology.ucsd.edu