A State Space Framework for the Network Analysis of Mammalian Metabolism

Adam C. Baughman, Lealon L. Martin, Susan T. Sharfstein

Rensselaer Polytechnic Institute, Troy, New York, USA 12180
Department of Chemical and Biological Engineering

For the 2007 Annual Meeting of the American Institute of Chemical Engineers
Salt Lake City, Utah, USA
Motivation

- Demand for biological compounds continues to expand.
  - The global market in therapeutic protein products is estimated to sustain 20% to 25% annual growth through 2010. [e.g. Pavlou 2004, Reichert 2004]
  - High therapeutic doses for many compounds creates ongoing concern regarding worldwide production capacity. [Browne 2007]
  - Demand for biological compounds in non-therapeutic applications (e.g. biocatalysis) is also predicted to increase. [Straathof 2002]

- Production cell lines are characterized by unique metabolic physiologies.
  - Customarily, cells do not manufacture any particular compound in great excess.
  - Specific productivity, primary metabolism, and cell growth are believed to be tightly linked. [e.g. Sharfstein 2005, Browne 2007]

- Modeling techniques have been proposed to guide development.
  - Several in silico techniques have been proposed, which attempt to quantify metabolic activity at the network-scale.
  - Such quantitative understanding should aid the development of optimal culture conditions and steer metabolic engineering efforts.
• Within the cell, mass must be conserved.
  • Changes in the overall cellular composition and volume are on the order of the overall cell growth rate → slow → pseudo-steady-state.

\[
\frac{\partial (C_i V)}{\partial t} = (U_i - Y_i) + \sum_{j=1}^{N} R_j^i = 0
\]

\[
\sum_{j=1}^{N} R_j^i = Y_i - U_i
\]

• By itself, this is insufficient to capture network metabolism.
  • An overall mass balance alone comprises an unstructured modeling approach.
  • Graph theory has previously been proposed to develop structured representations of metabolic networks [e.g. Arita 2000, Lacroix 2006].
• The intracellular trafficking of metabolites forms a directed graph.
  • Each node of the graph is a known metabolic conversion.
  • A single node may not both consume and produce the same metabolite.
  • The complete graph for any single metabolite will be disjoint.

\[
\begin{align*}
\mathcal{U}_1 &= w_1 + w_3 \\
\mathcal{U}_2 &= w_2 + w_5 \\
\mathcal{U}_3 &= w_5 + b_3 \\
\mathcal{Y}_4 &= x_3 \\
\mathcal{Y}_3 &= x_4 + b_3 \\
\mathcal{Y}_4 &= x_4 + b_3
\end{align*}
\]
What is State Space?

• Classically, State Space is a dynamic process control model.
  • The state of a process is quantitatively described as a set of fundamentally independent state variables.
  • This state is perturbed or influenced by one or more input/control variables.
  • The output of the process is a manifestation of one or more state and/or control variables.
  • State information is recovered from perturbation information.

\[
\begin{align*}
  s'(t) &= \mathcal{F}(s, u, t) \\
  y(t) &= \mathcal{G}(s, u, t) \\
  s(t) &= \Delta(s', t)
\end{align*}
\]

• If a linear, time-invariant, process model is appropriate...
  • The functions $\mathcal{F}$ and $\mathcal{G}$ may be represented as a linear combination of linear operators.
  • Each operator generates a particular mapping, or connectivity.
  • Any time dependence is removed.

\[
s' = \mathcal{F}(s, u) = A(u) + B(s)\\
y = \mathcal{G}(s, u) = C(u) + D(s)\\
s = \Delta(s')
\]
Our graph representation of metabolism may be *conceptually* integrated with a linearized *State Space* framework.

- The pseudo-steady-state approximation permits a *time-invariant* framework.
- Each linear operator corresponds precisely with a particular *mode* of network connectivity.
- Transformation/Recovery ($\Delta$) is captured through each *reaction/node* balance.

\[
\begin{align*}
U_i &= \sum_{j=1}^{N} w_j^i + b^i \\
Y_i &= \sum_{j=1}^{N} x_j^i + b^i \\
R_j^i &= \alpha_j^i R_j = \left( x_j^i + \sum_{k=1}^{N} \tau_{j,k}^i \right) - \left( w_j^i + \sum_{k=1}^{N} \tau_{k,j}^i \right)
\end{align*}
\]
Why Use State Space?

- **State Space** offers certain conceptual advantages over a strictly graphical approach.
  - Explicitly (mathematically) decouples distribution and transformation operations.
  - Lends itself to a compact representation of network analysis.
  - Expresses network interactions elegantly as “input → state → output” relationships.

Savinell & Palsson, 1992a
More Characteristics of State Space

• The structured framework is mathematically advantageous.
  • “Node-wise” construction of metabolic networks ensures the model will consistently remain under-determined.

  \[ p \text{ metabolites consumed} \rightarrow j \rightarrow q \text{ metabolites generated} \]

  At least \( p + q \) flow variables + 1 conversion rate \( \Rightarrow \) At least \( p + q + 1 \) unknowns.
  \( p + q \) reaction (node) balances \( \Rightarrow p + q \) constraints.

• The solution space – the State Space – is a linear vector space.
  • The space is convex \( \Rightarrow \) optimizing a convex objective over this space ensures a global solution.
  • When the process input, output and conversion rates are bound, the entire space is bound.
  • The Euclidean norm (2-norm) is valid over this space.

- An IgG-secreting murine Hybridoma was cultured continuously (CSTR) and metabolite consumption or generation rates were measured at the steady-state.
- An MFA model was constructed, consisting of 45 reactions in 51 metabolites.
  - To reduce complexity, “directly calculable” fluxes were resolved and eliminated.
  - The smaller rank-deficient model consisted of 20 to 22 constraints in 22 metabolites.
  - The reduced model was fit to experimental observations using a least-squares objective.
- We constructed an identical State Space formulation.
  - We used the complete reaction and metabolite set – no fluxes were eliminated.
  - Metabolites were intuitively identified as “inputs” or “outputs” of the network.
  - Bounds on all network variables were very relaxed.
  - The resulting linear model (272 equations in 592 variables) was fit to observations using the same least-squares objective.

‡ Biotechnology & Bioengineering 50 (299-318)
Stating the Optimization Problem

**Metabolic Flux Analysis**

\[
\min_{\mathbf{v}} \| \mathbf{\delta} - \mathbf{d} \|_2 \\
\text{s.t. } \mathbf{Sv} - \mathbf{\delta} = 0 \\
|\mathbf{v}| < \mathbf{v}_{\text{MAX}}
\]

**State Space Analysis**

\[
\min_{\mathbf{x}} \| \mathbf{Cx} - \mathbf{d} \|_2 \\
\text{s.t. } \mathbf{U}_i - \sum_{j=1}^{N} \mathbf{w}_j^i - \mathbf{b}_i = 0 \quad \forall i \\
\mathbf{\gamma}_i - \sum_{j=1}^{N} \mathbf{x}_j^i - \mathbf{b}_i = 0 \quad \forall i \\
\alpha_i^i \mathbf{R}_j^i + \left( \mathbf{w}_j^i + \sum_{k=1}^{N} \mathbf{\tau}_k^i \right) - \left( \mathbf{x}_j^i + \sum_{k=1}^{N} \mathbf{\tau}_{j,k}^i \right) = 0 \quad \forall i, j \\
0 \leq \mathbf{x} \equiv \{ \mathbf{U}, \mathbf{\gamma}, \mathbf{R}, \mathbf{w}, \mathbf{x}, \mathbf{\tau} \} \leq 10^6
\]
## Modeling Comparison

<table>
<thead>
<tr>
<th>Enzyme/Pathway</th>
<th>Relaxed State Space Prediction</th>
<th>Bonarius et al MFA Prediction†</th>
</tr>
</thead>
<tbody>
<tr>
<td>[PGI] : g6p → f6p</td>
<td>0.47</td>
<td>0.51</td>
</tr>
<tr>
<td>[PFK,FBA] : f6p → gap</td>
<td>3.16</td>
<td>4.01</td>
</tr>
<tr>
<td>[G6PDH] : g6p → ru5p</td>
<td>5.28</td>
<td>6.25</td>
</tr>
<tr>
<td>[RPI] : ru5p → r5p</td>
<td>2.58</td>
<td>2.23</td>
</tr>
<tr>
<td>[GAPDH,PGI] : gap → g3p</td>
<td>7.49</td>
<td>9.77</td>
</tr>
<tr>
<td>[PYK] : pep → pyr</td>
<td>7.24</td>
<td>6.72</td>
</tr>
<tr>
<td>[PEPCK] : pep → oma</td>
<td>0.00</td>
<td>2.73</td>
</tr>
<tr>
<td>[ME] : oma → pyr</td>
<td>0.00</td>
<td>1.67</td>
</tr>
<tr>
<td>[PDH] : pyr → accoa</td>
<td>1.62</td>
<td>1.61</td>
</tr>
<tr>
<td>[CS] : accoa + oma → cit</td>
<td>2.38</td>
<td>1.83</td>
</tr>
<tr>
<td>[ASNA] : asn → asp</td>
<td>0.00</td>
<td>0.35</td>
</tr>
<tr>
<td>[GLUDH] : glu → akg</td>
<td>0.00</td>
<td>-0.05</td>
</tr>
</tbody>
</table>

Values in $10^{-12}$mole/cell/day

†Results are given for simulation **including CO$_2$ and NADPH balances** (Rank = 20)
• A complete mammalian metabolic network of 84 reactions...
  • Reactions were selected on the basis of supposed metabolic significance.
  • Linear reaction paths, whose intermediates have limited alternative fates, were lumped.
  • Enzyme stoichiometry was referenced against a genome-scale mammalian model: Homo Sapiens Recon 1 [Palsson, bigg.ucsd.edu].
  • Biosynthetic reactions were constructed by known or measured stoichiometries (to the greatest extent possible).

• ...in 80 metabolic species.
  • Species are compartmentalized and may only exchange compartments if a suitable transporter is known or hypothesized to exist.
  • Plenary species are neglected – i.e. only the scarce species (by compartment) are accounted.
There may be certain “objectives” which govern cellular behavior.

- Minimize the secretion of toxic byproducts.
- Minimize the consumption of energetic substrates.
- Minimize the generation of reducing equivalents.

There may also be certain engineering objectives or constraints.

- Maintain or Maximize overall growth rate.
- Maximize specific productivity.

For initial explorations, we postulate a simple objective.

- A particular culture doubling time (e.g. 30 hours) suggests a particular growth rate and therefore a particular demand for biomass synthesis.
- We propose that the cell maintains this growth rate while minimizing its overall substrate/nutrient consumption.
Let us now consider results by Mancuso, Sharfstein et al (1998)

- Hybridomas were cultured continuously in a hollow-fiber bioreactor.
- At 240 hours post inoculation (well after steady-state growth was established), glutamine feed to the reactor was halted.
- Metabolite consumption/generation data was collected immediately prior to and immediately following this “glutamine shift”.

‡ Biotechnology & Bioengineering 57 (172-186)
Optimizing with State Space

- Using our postulated objective and prototype HSR-1 network, a State Space network simulation was attempted.
  - Metabolites were identified as network “inputs” or “outputs”.
  - The “output” flow variables were constrained to match their observed values.
  - To simulate the glutamine shift, a suitably small upper bound was imposed on the glutamine “input” flow variable.
  - No other network variables were constrained.
  - The complete model consisted of 412 equations in 802 variables.
## Optimization-Based Results

<table>
<thead>
<tr>
<th>Species</th>
<th>237.5 hr</th>
<th></th>
<th>241.25 hr</th>
<th></th>
<th>242.25 hr</th>
<th></th>
<th>243.25 hr</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>State Space</td>
<td>M &amp; S</td>
<td>State Space</td>
<td>M &amp; S</td>
<td>State Space</td>
<td>M &amp; S</td>
<td>State Space</td>
<td>M &amp; S</td>
</tr>
<tr>
<td>Glucose</td>
<td>86</td>
<td>135</td>
<td>107</td>
<td>205</td>
<td>97</td>
<td>220</td>
<td>89</td>
<td>240</td>
</tr>
<tr>
<td>Lactate</td>
<td>(155)</td>
<td>(155)</td>
<td>(185)</td>
<td>(185)</td>
<td>(175)</td>
<td>(175)</td>
<td>(160)</td>
<td>(160)</td>
</tr>
<tr>
<td>Oxygen</td>
<td>222</td>
<td>110</td>
<td>371</td>
<td>115</td>
<td>241</td>
<td>110</td>
<td>240</td>
<td>100</td>
</tr>
<tr>
<td>Antibody</td>
<td>(0.037)</td>
<td>(0.037)</td>
<td>(0.061)</td>
<td>(0.061)</td>
<td>(0.073)</td>
<td>(0.073)</td>
<td>(0.095)</td>
<td>(0.095)</td>
</tr>
<tr>
<td>Alanine</td>
<td>(9.2)</td>
<td>(9.2)</td>
<td>(12.4)</td>
<td>(12.4)</td>
<td>(12.7)</td>
<td>(12.7)</td>
<td>(11.2)</td>
<td>(11.2)</td>
</tr>
<tr>
<td>Aspartate</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>1.4</td>
<td>0.2</td>
<td>1.4</td>
<td>0.5</td>
</tr>
<tr>
<td>Glutamate</td>
<td>(1.9)</td>
<td>(1.9)</td>
<td>(2.5)</td>
<td>(2.5)</td>
<td>(2.1)</td>
<td>(2.1)</td>
<td>(134)</td>
<td>0.1</td>
</tr>
<tr>
<td>Glycine</td>
<td>0.5</td>
<td>0.4</td>
<td>(0.4)</td>
<td>(0.4)</td>
<td>(0.9)</td>
<td>(0.9)</td>
<td>(1.1)</td>
<td>(1.1)</td>
</tr>
<tr>
<td>Proline</td>
<td>(3.0)</td>
<td>(3.0)</td>
<td>(4.0)</td>
<td>(4.0)</td>
<td>(4.0)</td>
<td>(4.0)</td>
<td>(2.5)</td>
<td>(2.5)</td>
</tr>
<tr>
<td>Glutamine</td>
<td>155</td>
<td>37.7</td>
<td>10.0</td>
<td>5.5</td>
<td>10.0</td>
<td>7.2</td>
<td>10.0</td>
<td>7.1</td>
</tr>
<tr>
<td>Histidine</td>
<td>4.0</td>
<td>1.5</td>
<td>1.6</td>
<td>1.4</td>
<td>30.8</td>
<td>1.5</td>
<td>20.8</td>
<td>2.3</td>
</tr>
<tr>
<td>Leucine</td>
<td>15.2</td>
<td>5.3</td>
<td>19.9</td>
<td>7.2</td>
<td>18.8</td>
<td>8.1</td>
<td>18.8</td>
<td>8.4</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>3.8</td>
<td>3.9</td>
<td>56.9</td>
<td>5.8</td>
<td>60.3</td>
<td>6.9</td>
<td>64.5</td>
<td>7.6</td>
</tr>
<tr>
<td>Asparagine</td>
<td>(137)</td>
<td>0.4</td>
<td>3.1</td>
<td>0.5</td>
<td>3.1</td>
<td>0.6</td>
<td>3.0</td>
<td>0.6</td>
</tr>
<tr>
<td>Valine</td>
<td>5.3</td>
<td>3.6</td>
<td>5.5</td>
<td>5.7</td>
<td>5.6</td>
<td>6.7</td>
<td>5.7</td>
<td>7.6</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>3.7</td>
<td>2.1</td>
<td>3.8</td>
<td>2.6</td>
<td>3.8</td>
<td>2.9</td>
<td>3.9</td>
<td>3.3</td>
</tr>
<tr>
<td>Ammonium</td>
<td>(25.5)</td>
<td>(25.5)</td>
<td>(53.2)</td>
<td>(53.2)</td>
<td>(46.8)</td>
<td>(46.8)</td>
<td>(37.1)</td>
<td>(37.1)</td>
</tr>
</tbody>
</table>
Remarks

- *State Space* represents a mathematically advantageous framework.
  - The fundamental incorporation of network structure alleviates the need for additional constraints to obtain biologically consistent results.
  - Node-wise construction ensures sufficient degrees of freedom for analysis.
  - The framework uniquely permits the exploration of network connectivity.

- Like any modeling approach, *State Space* does have some limitations.
  - Enforcing a network structure with defined inputs and outputs requires very careful reaction selection.

- *State Space* complements and expands the set of established metabolic modeling techniques.

- *In Silico* results should guide *in vitro* experimentation.
  - We intend to apply this framework to quantify differences in the metabolic phenotypes of high and low producing CHO cell lines.
Acknowledgements

• Thanks to my research group for their support, inventive thinking, and many welcome distractions.
  • Ian Tolle
  • David Follansbee
  • Jessica Hronich
  • Tom Kiehl
  • Xinqun Huang

• Thanks also to the RPI Research Office for continued funding.
  • RPI #146086