Text S1 (Analog regulation of metabolic demand)
Nikolaus Sonnenschein1,*, Marcel Geertz2, Georgi Muskhelishvili1, Marc-Thorsten Hüt1
1 School of Engineering and Science, Jacobs University Bremen, Bremen, Germany
2 Department of Molecular Biology, University of Geneva, Geneva, Switzerland
* E-mail: n.sonnenschein@jacobs-university.de

Metabolic network representations

The strength of graph theory is that it can represent a complex system in a unified formal language of
nodes and links. Examples of graph-theoretical analyses of metabolic systems are [1] and [2]. Essentially,
the pattern of zero and non-zero entries of the stoichiometric matrix defines a graph representation of the
metabolic system. The level of information conveyed by looking at metabolism from a graph-theoretical
perspective is still subject to constant scrutiny [3].

A suitable approach for analyzing the correspondence between expression changes and metabolism and
thus quantifying metabolic coherence is the application of the tools developed for the effective TRNs [4]
to a gene-centric representation of metabolism, where the nodes are metabolic genes and a link is drawn
between two genes, if the associated reactions share a common metabolite [5]. This representation is the
gene-centric variant of one of the standard projections of a bi-partite graph representation of metabolism,
where both, metabolites and reactions serve as node sets (see, e.g., 6, for a discussion of these and other
representations).

Additional forms of representation include metabolite-, reaction-, enzyme- or gene-centric views. The
metabolite-centric view represents the interconversion possibilities of the different substrates, whereas
other views concentrate more on the processes (reactions), protein (enzymes) and genomic (genes) levels
respectively. We chose the gene-centric view for our analysis as it allowed us a direct comparison of
expression patterns with metabolic pathways.

Reaction to gene mappings

Imagine the following scenario: a reaction is catalyzed by a single enzyme (no isozymes involved), encoded
by a unique gene (no enzyme complexes involved). With this simple scheme in mind one might conclude
that reaction-, enzyme- and gene-centric representations are redundant. However, most of the time
reactions and their associated enzymes and genes are not interchangeable. Figure S2 visualizes the
amount of multiplicities between the reaction- and gene-level. The columns of the colored grid represent
the absolute number of genes per reaction pair. The rows represent the number of unique genes. So
the number of consecutive reaction pairs sharing a single gene can be found in column 2, row 1, as both
reactions are associated with a single gene and it happens that it is the same for both. In contrast,
the previously described simple scheme can be found in column 2, row 2. In order to assess the impact
of these ambiguities on our results we constructed other graphs in addition to our gene-networks by
applying the following rules: (1) the removal of reaction pairs lying beside the diagonal of the grid (see
Figure S2) excludes situations where a single or multiple genes are involved with both reactions; (2) taking
into account only reaction pairs fulfilling the condition of the second column and row, thus effectively
excluding enzyme complexes and the situations described in (1).

Currency metabolites

Another problem emerges through highly connected compounds, which have been termed current or
currency metabolites in the past [7, 8]. They have caused reports of questionable average path lengths
as they represent unrealistic shortcuts obscuring the essential pathway structures that have been assembled by biochemists over the last century.

For example Figure S3 demonstrates the huge impact of currency metabolites on our MC results by showing the z-score pattern for the untreated iAF1260 network. All scores are basically below or a little above 1 and such no significant coherence could be measured.

However, the KEGG and EcoCyc data sets provide currency metabolite free representations through their human readable pathway maps (in contrast to their complete reaction databases). For iAF1260 [10] we employed on the one hand a threshold heuristic to remove a certain percentage (i.e. 4 % for the results shown in the main article) of the most highly connected metabolites as described by Kharchenko et al. [5], and on the other hand a manual curation of the network where currency metabolites were removed on a reaction to reaction basis, i.e. the approach described by Ma and Zeng [8]. Figure 3C (in the main text) shows the MC result for the manually curated network for comparison with the untreated one (see Figure S3). Figure S4 shows the dependency of the MC on the percentage threshold.

**Constraint-based modeling**

For a metabolic system consisting of \( N \) reactions and \( M \) compounds the linear programming (LP) formulation of FBA can generally be stated as follows:

Maximize \( Z = c \cdot v \)

subject to

\[
\begin{align*}
\sum_{j=1}^{N} S_{ij} v_j, & \quad i = 1, \ldots, M \\
v_j^{\text{min}} & < v_j < v_j^{\text{max}}, & \quad j = 1, \ldots, N \\
v_j^{(m,s)} & < v_j^{(t)} < v_j^{(m,u)}, & \quad j = 1, \ldots, N^{(t)},
\end{align*}
\]

where \( v \) is a vector of reaction fluxes constrained by the stated boundary conditions, \( S \) is a matrix storing the stoichiometric information of the system (i.e. \( S_{ij} \) is the stoichiometric coefficient of metabolite \( i \) in reaction \( j \) ) and \( Z \) denotes the objective to be maximized represented by a linear combination of fluxes \( v_j \) and objective coefficients \( c_j \). Here, \( v^{(t)} \) denotes a transport reaction, i.e. a reaction either secreting metabolites from the system or taking them up. The quantity \( N^{(t)} \) is the number of transport reactions. As an approximation to a rich medium condition we allowed for every available transport reaction \( v^{(t)} \) unlimited secretion \( v_j^{(m,s)} = -\infty \) and \( v^{(m,u)} = 20 \) [in units of mmol/g · dw · h] as an arbitrary upper bound to influx. With the exception of \( v^{(t)} \), all reversible reactions were treated as two distinct irreversible reactions. Maximization of biomass production [10] and simultaneous minimization of all other fluxes was used as \( Z \) in order to avoid accumulation of flux in cycles. As all constraints are linear and the solution space is convex, a global maximum can always be found using linear programming (assuming the problem is well defined and not unbounded), though multiple global optima cannot be excluded [11].

**References**


**Figure S1.** The number of active reactions and genes in the effective networks increases when moving from rich to minimal media conditions. Only cytosolic reactions and genes were counted (i.e. transport and periplasmic reactions were excluded).
**Figure S2.** The multiplicities of reaction-gene relations depicted as a colored grid.

**Figure S3.** Effect of currency metabolites as seen for the untreated iAF1260 network.
Figure S4. (A) MC values plotted against the percentage of removed currency metabolites (as determined by the degree threshold method). (B) The network connectivity, i.e. number of connections in the network, plotted against the percentage of removed metabolites.

Table S1. Table of the MC values visualized in Figure 5 (main article).

<table>
<thead>
<tr>
<th>Label</th>
<th>Network</th>
<th>WT</th>
<th>fis</th>
<th>hns</th>
<th>fis/hns</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EcoCyc</td>
<td>3.37</td>
<td>0.31</td>
<td>-0.60</td>
<td>-0.65</td>
</tr>
<tr>
<td>2</td>
<td>EcoCyc*</td>
<td>3.17</td>
<td>0.37</td>
<td>-0.53</td>
<td>-0.52</td>
</tr>
<tr>
<td>3</td>
<td>Intersection of EcoCyc and KEGG networks</td>
<td>3.03</td>
<td>0.26</td>
<td>0.54</td>
<td>-0.66</td>
</tr>
<tr>
<td>4</td>
<td>Intersection of EcoCyc and iAF1260\text{_man}</td>
<td>2.89</td>
<td>0.57</td>
<td>-0.40</td>
<td>-0.13</td>
</tr>
<tr>
<td>5</td>
<td>iAF1260\text{_man}</td>
<td>2.75</td>
<td>1.47</td>
<td>0.85</td>
<td>-0.05</td>
</tr>
<tr>
<td>6</td>
<td>iAF1260\text{_man} obtained from FBA (rich medium)</td>
<td>2.69</td>
<td>1.21</td>
<td>-1.11</td>
<td>-0.38</td>
</tr>
<tr>
<td>7</td>
<td>iAF1260\text{_man}</td>
<td>2.62</td>
<td>1.61</td>
<td>0.85</td>
<td>-0.065</td>
</tr>
<tr>
<td>8</td>
<td>EcoCyc**</td>
<td>2.55</td>
<td>0.22</td>
<td>0.33</td>
<td>0.65</td>
</tr>
<tr>
<td>9</td>
<td>KEGG*</td>
<td>2.57</td>
<td>0.76</td>
<td>1.59</td>
<td>1.32</td>
</tr>
<tr>
<td>10</td>
<td>iAF1260*</td>
<td>2.50</td>
<td>1.51</td>
<td>0.03</td>
<td>0.55</td>
</tr>
<tr>
<td>11</td>
<td>iAF1260\text{_deg}</td>
<td>2.40</td>
<td>1.52</td>
<td>0.16</td>
<td>0.66</td>
</tr>
<tr>
<td>12</td>
<td>iAF1260**\text{_man}</td>
<td>2.36</td>
<td>2.09</td>
<td>0.37</td>
<td>0.12</td>
</tr>
<tr>
<td>13</td>
<td>Flux-coupling network (fully coupled)</td>
<td>2.22</td>
<td>0.14</td>
<td>1.70</td>
<td>1.14</td>
</tr>
<tr>
<td>14</td>
<td>KEGG</td>
<td>2.20</td>
<td>0.61</td>
<td>1.31</td>
<td>1.40</td>
</tr>
<tr>
<td>15</td>
<td>Intersection of KEGG and iAF1260\text{_man}</td>
<td>1.80</td>
<td>1.43</td>
<td>-0.11</td>
<td>-0.46</td>
</tr>
<tr>
<td>16</td>
<td>Intersection of EcoCyc, KEGG and iAF1260\text{_man}</td>
<td>1.76</td>
<td>0.68</td>
<td>-0.26</td>
<td>-0.48</td>
</tr>
<tr>
<td>17</td>
<td>KEGG**</td>
<td>1.58</td>
<td>-0.17</td>
<td>-0.46</td>
<td>1.048</td>
</tr>
<tr>
<td>18</td>
<td>Flux-coupling network (fully and directionally coupled)</td>
<td>1.50</td>
<td>1.69</td>
<td>1.91</td>
<td>0.55</td>
</tr>
<tr>
<td>19</td>
<td>iAF1260\text{_deg}</td>
<td>1.37</td>
<td>1.62</td>
<td>0.14</td>
<td>-0.11</td>
</tr>
<tr>
<td>20</td>
<td>iAF1260**</td>
<td>1.20</td>
<td>1.17</td>
<td>0.42</td>
<td>1.17</td>
</tr>
<tr>
<td>21</td>
<td>Flux-coupling network (directionally coupled)</td>
<td>1.06</td>
<td>0.39</td>
<td>0.03</td>
<td>1.51</td>
</tr>
<tr>
<td>22</td>
<td>iAF1260</td>
<td>0.19</td>
<td>1.17</td>
<td>0.23</td>
<td>1.40</td>
</tr>
<tr>
<td>23</td>
<td>iAF1260*</td>
<td>0.11</td>
<td>1.14</td>
<td>0.13</td>
<td>1.41</td>
</tr>
</tbody>
</table>