Modular decomposition of metabolic reaction networks based on flux analysis and pathway projection

Jeongah Yoon¹, Yaguang Si², Ryan Nolan¹ and Kyongbum Lee¹,*

¹Department of Chemical and Biological Engineering, Tufts University, 4 Colby Street and ²Department of Biology, Tufts University, 163 Packard Avenue, Medford, MA 02155, USA

Received on February 22, 2007; revised on July 10, 2007; accepted on July 11, 2007

Advance Access publication July 27, 2007

Associate Editor: Olga Troyanskaya

ABSTRACT

Motivation: The rational decomposition of biochemical networks into sub-structures has emerged as a useful approach to study the design of these complex systems. A biochemical network is characterized by an inhomogeneous connectivity distribution, which gives rise to several organizational features, including modularity. To what extent the connectivity-based modules reflect the functional organization of the network remains to be further explored. In this work, we examine the influence of physiological perturbations on the modular organization of cellular metabolism.

Results: Modules were characterized for two model systems, liver and adipocyte primary metabolism, by applying an algorithm for top-down partition of directed graphs with non-uniform edge weights. The weights were set by the engagement of the corresponding reactions as expressed by the flux distribution. For the base case of the fasted rat liver, three modules were found, carrying out the following biochemical transformations: ketone body production, glucose synthesis and transamination. This basic organization was further modified when different flux distributions were applied that describe the liver’s metabolic response to whole body inflammation. For the fully mature adipocyte, only a single module was observed, integrating all of the major pathways needed for lipid storage. Weaker levels of integration between the pathways were found for the early stages of adipocyte differentiation. Our results underscore the inhomogeneous distribution of both connectivity and connection strengths, and suggest that global activity data such as the flux distribution can be used to study the organizational flexibility of cellular metabolism.

Contact: kyongbum.lee@tufts.edu

Supplementary information: Supplementary data are available at Bioinformatics online.

1 INTRODUCTION

A living system such as a cell routinely experiences a variety of perturbations that elicit multiple responses, such as a mutation, chemical challenge or change in the growth conditions (di Bernardo et al., 2005; Ideker et al., 2001; Segre et al., 2002). These perturbations may not only result in the activation or repression of a few specific biochemical pathways at the site of the perturbation, but also bring about global changes in gene expression or enzyme activity. These types of global responses cannot be attributed to the actions of one or a few individual genes, proteins or metabolites; rather, they arise from the concerted actions of networks of such molecules. Indeed, it has become increasingly apparent that the functions and features of a biochemical system are closely intertwined with its ‘wiring’, or structural layout (Stelling et al., 2002). Therefore, modeling frameworks are needed to characterize and analyze cellular biochemical networks as integrated systems. In recent years, graph theory based modeling has emerged as a useful approach to study the connectivity of cellular networks, especially metabolic networks. Important early findings include the scale-free character (Jeong et al., 2000) and small world property (Fell and Wagner, 2000), which clearly demonstrated the highly inhomogeneous distribution of connectivity in biochemical networks (Girvan and Newman, 2002). Previous studies have also suggested that more highly interconnected groups of cellular components, called modules, tend to share a common function (Ideker et al., 2001).

To date, observations on modularity have been generally based on a somewhat qualitative aspect of network connectivity, i.e. the presence of an expected biochemical interaction between a pair of components. An example is the assignment of a connection between a substrate and product metabolite based on reaction stoichiometry. Such an approach to connection assignments clearly limits the capacity of the graph model to consider the variable extents of biochemical interactions arising under different physiological conditions. One way to overcome this limitation is to introduce modifications that facilitate the integration of quantitative data on the distribution of biochemical activity. At least for unicellular organisms, optimization-based simulation tools such as flux balance analysis (FBA) (Schilling et al., 2000) have become available to calculate flux distributions for physiologically reasonable settings. Indeed, such in silico analyses have predicted global flux distributions in the Escherichia coli network that are clearly inhomogeneous and vary with changing environmental conditions (Almaas et al., 2004). A smaller-scale analysis conducted by our group using experimental data showed that the distribution of flux in a mammalian tissue such as the liver is also far from homogeneous (Nolan et al., 2006).

*To whom correspondence should be addressed.

© The Author 2007. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org
Appropriately modified, graph models should be useful in exploring issues related to organizational flexibility, i.e. whether the modular arrangement of a biochemical network is inherent to the network’s static layout, or instead depends on the functional state of the network. In the latter case, modularity analysis offers a powerful way to parse the behavior of complex biochemical networks by examining the relationship between the structural and functional building blocks. In turn, characterization of functionally relevant building blocks should help comparative analyses of similar modules across different species to identify mutually shared functions, associate a modular structure with a new function, and provide insight into the evolution of various network structures (Sharan and Ideker, 2006). Methods for the rational decomposition of biochemical networks should also facilitate the piecwise development of other detailed and mechanistic models that connect local properties, e.g. enzyme elasticities, to the global behavior of the overall system.

In this article, we present a framework for modularity analysis that takes as inputs both the connectivity information as well as biochemical activity distribution (connection diversity data). This framework extends a recently reported algorithm for top-down partitioning of directed and weighted networks (Yoon et al., 2006). Here, we combine this algorithm with stoichiometric vector space analysis and metabolic flux analysis (MFA) to examine the modularity of metabolic networks subjected to different physiological stimuli. The combined analysis procedure forms graphical substructures by evaluating a global index for the centrality of reactions based on the degrees of their biochemical engagement. In addition to the partition algorithm, we also describe a scoring method for selecting the most promising modular structures among candidate substructures by identifying positive relationships with recognizable metabolic functions. We develop both the partition and scoring methods against two relatively simple, but distinct model systems for which detailed, experimentally derived metabolic flux data were available. The first model described the impact of fasting and whole body inflammation on the energy metabolism of the rat liver (Lee et al., 2003). The second model represented the metabolic changes underlying de novo adipose tissue formation (Si et al., 2007).

2 METHODS

2.1 Experimental data

2.1.1 Liver perfusion Metabolite uptake or output (external flux) data were obtained using closed loop liver perfusion experiments. These livers were isolated from either burn or sham-burned rats. Liver perfusions were performed 0 (sham-burn), 1, 2, 3, 4 or 7 days after treatment. Rats were fasted for 24 h prior to the perfusions to deplete liver glycogen stores and synchronize the animals to a common initial nutritional state. Measurements on the perfusion medium consisted of assays on 25 major intermediates of liver central carbon metabolism: glucose, lactate, urea, oxygen, carbon dioxide, ketone bodies and amino acids. Experimental details on the metabolite measurements have been reported previously (Lee et al., 2003).

2.1.2 Adipocyte culture 3T3-L1 preadipocytes (ATCC, Manassas, VA, USA) were expanded in T-75 flasks, passed once, and seeded in 6-well plates whose wells had been previously coated with a collagen gel layer. The collagen gel mixture was prepared by mixing ice-cold type I rat tail collagen, 10× DMEM, and 7.5% NaHCO\textsubscript{3} at a ratio of 8.5:1.0:5, respectively. The preadipocytes were grown to confluence and differentiated using a standard protocol. Medium was replenished every other day through day 16 post-induction. Cell lysate and spent medium samples were collected at various times during the culture (days 4, 8, 12 and 16 post-induction) and analyzed for metabolite concentrations. Triglyceride (TG), total fatty acids, and free glycerol were measured in addition to all of the metabolites measured for the liver with exception of urea, oxygen and carbon dioxide.

2.2 Liver and adipocyte models

Liver and adipocyte models of central carbon metabolism were constructed as follows. First, organism (rat or mouse) specific lists of enzyme-mediated reactions were collected from annotated genome databases generously provided by Ma and Zeng (2003). Second, stoichiometric information was added for each of the collected enzymes by cross-referencing their common names and enzyme commission names using the KEGG database (Kanehisa and Goto, 2000). Third, biochemistry textbooks and the published literature (Arias and Boyer, 2001; Digirolamo and Fried, 1987) were consulted to eliminate enzymes thought to be inactive in the rat liver during the fasted state or the mouse adipocyte during the fed state. Finally, the stoichiometric models were rendered into compound directed graphs, visualized using the MATLAB Bioinformatics toolbox (MathWorks, Natick, MA, USA), and corrected for missing steps and nonsensical dead ends. Reversible reactions flanked by irreversible reactions were assigned directionality so as to ensure unidirectional metabolic flux between the flanking reactions. The final liver model consisted of the following major pathways: gluconeogenesis, pentose phosphate pathway (PPP), urea cycle, amino acid catabolism, amino acid degradation, ketone body synthesis, β-oxidation and the TCA cycle. For the adipocyte model, we eliminated gluconeogenesis and urea cycle, and added glycolysis, glycerogenesis, lipogenesis and lipolysis. A full list of the enzymes included in the model and their reaction stoichiometry are included as Supplementary Material in Tables S1 (liver) and S2 (adipocyte).

2.3 Metabolic flux analysis

Intracellular fluxes were estimated by solving an optimization problem as described previously (Nolan et al., 2006):

$$\text{Minimize } \sum_{k} (v_{\text{obs}}^{k} - v_{\text{obs}}^{k})^2,$$

Subject to:

$$S \cdot v = 0 \quad (1)$$
$$G \cdot v \leq 0 \quad (2)$$

where the objective function minimizes the sum squared error between experimentally observed ($v_{\text{obs}}^{k}$) and predicted external fluxes ($v_{\text{obs}}^{k}$). Equation (1) expresses the pseudo steady-state balances around the major intracellular metabolites using a stoichiometric matrix $S (M \times R)$ and a flux vector $v (R \times 1)$. The number of balanced metabolites ($M$) and reactions ($R$) were 22 and 50 for the liver and 32 and 66 for the adipocyte model, respectively. Inequality (2) expresses the constraints derived from the Gibbs free energy change ($\Delta G$) form of the Second Law using a pathway $\Delta G$ matrix $G (P \times R)$. The number of pathways ($P$) were 217 and 582 for the liver and adipocyte model, respectively. For the liver, a unique set of fluxes was calculated for each of the six time points (days 0, 1, 2, 3, 4 and 7 post-burn), where the $v_{\text{obs}}^{k}$ for each time point represented the averaged result of at least four separate replicate perfusions (Lee et al., 2003). For the adipocyte, flux distributions were calculated for four different time points (days...
4, 8, 12 and 16 post-induction), where the $t_{g_{i}}$ for each time point was the averaged result of six independent replicate cultures.

2.4 Graph model and flux-based edge weight

We used a compound directed graph with metabolite vertices and reaction edges. This graph type intuitively represents the enzyme-mediated connectivity between metabolites, and accommodates assignment of edge weights that quantify the relative connection strengths. Here, the weights were determined by the metabolic flux distribution. The weight $W(x,y)$ of an edge from metabolite (vertex) $x$ to $y$ was calculated as follows (Fig. 1):

$$W(x,y) = \left( \sum_{f} s_{x,y} f_{j} \right)^{-1}, \quad \left\{ \begin{array}{l} n_{j} = \sum_{f} s_{x,y} x \in \{k\} (a) \\ n_{j} = 1 \text{ and } s_{x,y} = 1 (b) \end{array} \right.$$  

(3)

Each reaction $j$ has flux $f_{j}$ and metabolite $x$ as a reactant and $y$ as a product. If reaction $j$ has more than one reactant (case a), then $f_{j}$ is multiplied by the reactant's stoichiometric coefficient $s_{x,y}$ and divided by the sum $(n_{j})$ of the stoichiometric coefficients $(s_{x,y})$ of every reactant of reaction $j$. If there is only one reactant (case b), $n_{j}$ and $s_{x,y}$ are both simply set to one. If there are multiple reactions $j$ involving the reactant $x$ and product $y$, $W(x,y)$ is calculated by first summing the normalized fluxes $f_{j}/n_{j}$ and then taking the inverse of the sum.

2.5 Network partition

Clusters of highly interconnected and active reactions were identified using an algorithm for top-down partition of directed and weighted graphs (Yoon et al., 2006). The algorithm consists of the following two steps, which are repeated iteratively until no more edges remain.

**Step 1:** Shortest paths through the network are calculated using Dijkstra's algorithm. These calculations take into account both edge directionality and weight, where the latter is derived from the flux distribution. The flux-derived weights were used to adjust the edge lengths between reactant and product metabolites such that a higher flux would correspond to a shorter length. These weights are calculated once initially and used throughout the iterations. In the limiting case of zero flux, the corresponding edge weight is infinite, and thus the edge is unavailable to any paths, reflecting a non-active component of the network. Such an edge was eliminated from the shortest path calculations.

**Step 2:** The edge-betweenness centrality (EBC) index is calculated for all edges based on the shortest paths obtained in step 1. The edge with the highest centrality index is removed, forming a new graph of the network with one less edge. It is possible to remove more than one edge during one iteration, if there are multiple edges with equally high EBC values. The EBC index measures the frequency with which an edge lies on the shortest paths of the network. The edges with highest EBC values are most likely to lie between subgraphs, rather than inside a subgraph (Newman and Girvan, 2004). Consequently, successively removing edges with the highest EBC values will eventually isolate subgraphs consisting of vertices that share connections only with other vertices in the same subgraph.

2.6 Projection and match scores

To guide in the selection of an optimal partition, we developed a scoring procedure that evaluates a candidate module based on its resemblance to a coherent metabolic pathway. These pathways were enumerated by applying the elementary flux mode (EFM) analysis to the parent network.

2.6.1 Projection score

Each subgraph (candidate module) was transformed into a $1 \times R$ binary reaction composition vector (RCV), where $R$ is the total number of reactions included in the network. An element was set to one, if the reactants and products of the corresponding reaction were present as metabolite nodes in the module; otherwise, an element was set to zero. In case a metabolite was solely present as either a reactant or product due to an edge removal, the reaction composition vector element corresponding to the removed reaction was set to zero. These module vectors were not computed for subgraphs consisting of a single node. The EFM vectors were also transformed into $1 \times R$ binary pathway inventory vectors (PIVs) by replacing all nonzero entries with one. A projection score was computed for every pairwise combination of a binary module vector and each of the 217 (liver) or 582 (fat cell) PIVs as follows:

$$\text{PS}^{(i)}_{j} = \left( \frac{\text{RCV}^{(i)} \cdot \text{PIV}_{j}}{N^{(i)}} \right), \quad i = 1, 2, \ldots, L^{(i)}, j = 1, 2, \ldots, m$$  

(4)

where $\text{PS}^{(i)}_{j}$, $\text{RCV}^{(i)}$ and $N^{(i)}$ were, respectively, the projection score, reaction composition vector and number of nodes in module $i$ at iteration number $k$, $L^{(i)}$ was the number of modules at $k$, PIV, was the $j$th PIV and $m$ was the total number of EFM vectors. The overall projection score of an iteration number $k$ was calculated by averaging the ‘best match’ projection scores of all modules at this iteration:

$$\text{PS}^{(k)} = \frac{\sum_{i=1}^{L^{(k)}} \max \text{ PS}^{(i)}_{j}}{L^{(k)}}$$  

(5)

2.6.2 Match score

The projection scoring generally identified more than one possible best match PIV for a given candidate module. Therefore, we computed a second match score to compare a candidate module against the set of reactions shared by these multiple best match PIVs. This set of reactions was designated as the consensus pathway (CPV) and was unique to each candidate module. The match score was calculated as follows:

$$\text{MS}^{(i)}_{j} = \frac{R - W^{(i)}_{j}}{R}, \quad i = 1, 2, \ldots, L^{(i)}$$  

(6)

where $\text{MS}^{(i)}_{j}$ was the match score of a candidate module $i$ at iteration number $k$, and $W^{(i)}_{j}$ was the number of mismatches between the module $i$ and the corresponding CPV. A mismatch occurs if a reaction is found in the module, but not the fragment vector or if a reaction is found in the fragment, but not the module vector. The overall match score of an iteration number $k$ was calculated by averaging the individual module match scores:

$$\text{MS}^{(k)} = \frac{\sum_{i=1}^{L^{(k)}} \text{ MS}^{(i)}_{j}}{L^{(k)}}$$  

(7)
3 RESULTS

3.1 Flux distribution and modularity

We first examined how the apparent modularity varies when the enzyme-mediated connections between metabolites are not assumed to be uniformly active, but instead weighted according to their corresponding reaction engagements. For brevity, detailed results are shown here only for the base cases of the fasted liver (time zero, sham-burn control) and fully differentiated adipocyte (day 12 post-induction). For the unweighted liver model, the main effect of the highest betweenness edge removals was to reduce the size of the parent network (Fig. 2a), as subgraphs were not observed until the second to last iteration (Fig. 2b). In contrast, with non-uniform edge weights, subgraphs were first observed at iteration 5 (Fig. 2d). The ‘best’ partition obtained at iteration 6 consisted of three subgraphs (Fig. 2e). The smallest subgraph consisted of the intermediates of the pentose phosphate pathway (PPP) and gluconeogenesis. A second subgraph consisted of several amino acids and their common deamination products in the TCA cycle. The third subgraph encompassed virtually all of the major metabolic pathways, including the malate cycle, TCA cycle, glycolysis, TG synthesis and the PPP.

Our results show that the partition of a graph based on a global measure of centrality clearly depends on

Fig. 2. Liver network at various algorithm iterations. Graphs show the parent network (a), the optimal partition without edge weights (b), partitions with flux weights at iterations 4 (c) and 5 (d), the optimal partition with flux weights at iteration 6 (e) and a partition with flux weights at iteration 8. See Figure 3 legend for the metabolite abbreviations. Thick arrows show edge(s) removed at the end of the iteration.

Fig. 3. Adipocyte network at various algorithm iterations. Graphs show the parent network (a), the optimal partition without edge weights at iteration 9 (b), the optimal partition with flux weights at iteration 4 (c) and a partition with flux weight at iteration 8 (d). Abbreviations: AACC, acetoacetyl-CoA; ACC, acetyl-CoA; ACAC, acetoacetate; AKA, 2-oxoadipate; AKG, 2-oxoglutarate; CIT, citrate; FUM, fumarate; F6P, fructose 6-P; G6P, glucose 6-P; GAP, glyceraldehyde 3-P; PEP, phosphoenolpyruvate; G, glycerone-P; MAL, malate; OAA, oxaloacetate; ORN, ornithine; PAL, palmitate; PYR, pyruvate; R5P, ribulose 5-P; SUCC, succinyl-CoA.
connection diversity. Here, we have used reaction flux estimates to quantify the relative engagements of the enzyme-mediated connections, and thereby reflect upon the modular decomposition process, the inhomogeneous distribution of connection strengths. The flux-based edge weights generally decreased the number of iterations needed to achieve both the optimal and complete partition (when all edges have been removed). Some reactions carry very low or zero flux, representing the inactive components. These edges are effectively removed at the start of the algorithm, thus reducing the number of subsequent edge removal operations. Moreover, the modular decomposition obtained in this way reflects the specificity of metabolic biochemistry with respect to cell type. For example, the liver plays a known role in maintaining whole body metabolic homeostasis during alternating periods of feeding and fasting. During fasting, the liver shifts its own fuel utilization from glucose to fatty acids, and releases glucose by carrying out gluconeogenesis. Depending on the severity and length of the fast, the liver may also utilize fatty acids and amino acids to synthesize ketone bodies. The partition shown in Figure 2e suggests that these tasks are accomplished through three sets of biochemical transformations: amino acid degradation, de novo synthesis of glucose and formation of ketone bodies. Each of these functions was not mapped to a standard biochemical pathway such as the PPP or urea cycle; rather, each involved piecewise combinations of several such pathways. This suggests that the modular analysis approach could provide a means to study another level of biochemical organization beyond the traditional notion of a ‘pathway’. One implication is that coordination of enzyme activities occurs not only within a particular pathway but also across pathways, which would confer the advantage of additional flexibility in terms of the types of biochemical transformation operations a cell may perform. Like the liver, the adipose tissue plays an important role in the regulation of whole body metabolism. In the fed state, which was simulated across pathways, which would confer the advantage of additional flexibility in terms of the types of biochemical transformation operations a cell may perform. Like the liver, the adipose tissue plays an important role in the regulation of whole body metabolism. In the fed state, which was simulated through a high glucose culture medium, the dominant function is to store lipids. The major substrates for lipogenesis are glucose and fatty acids, while amino acids are quantitatively less significant. The partition shown in Figure 2c clearly points to a separation between amino acid metabolism and the other major pathways. Moreover, the main module includes all of the biochemical transformations for lipogenesis, including glucose catabolism, fatty acid esterification and the generation of reducing co-factors through the PPP and malate cycle.

3.2 Reaction directionality and module interactions

A natural benefit of the directed graph representation is that it facilitates the tracing of carbon flow not only within, but also between modules. The carbon flow between modules at any given partition is simply deduced from the previous partition, which is illustrated here for the fasted-state liver model. The three modules of Figure 2e were obtained at iteration 6. To determine how these modules are connected, we refer to the partition obtained at iteration 5 (Fig. 2d). Visual inspection shows that modules for ketone body and glucose metabolism connect through the reaction converting oxaloacetate (OAA) to fumarate (FUM) (via aspartic acid, ASP). Similarly, an inspection of the partition at iteration 4 (Fig. 2c) shows that the module for amino acid degradation connects to the module for ketone body metabolism through the reaction converting succinyl-CoA (SUCC) to FUM. There is no direct connection between the modules for amino acid degradation and glucose metabolism, i.e. these modules interact mainly through their shared connections to the ketone body module, which encompasses portions of the TCA cycle. The layered interactions between modules observed here is supported by previous reports on the hierarchical organization of substructures in metabolic reaction networks (Holme et al., 2003).

3.3 Evaluation of module quality

When taking a top-down approach, an important question is when to stop the partition process. If the iteration number is too low (not enough partition operations have been performed), we cannot observe any modules. If the partition level is too high, most of the network has been decomposed into disconnected single nodes. The partitions obtained with our algorithm varied, in some cases significantly, with the number of performed iterations. While it is possible to manually inspect each partition and heuristically judge its biological relevance, a more systematic approach was preferred. Here, we used a projection-based scoring method to assess the similarity of a candidate module (i.e. subgraph) to a coherent pathway or pathway fragment. An inventory of pathways was established through an EFM analysis (Schuster et al., 2000) of the parent network. The assumption was that the EFMs and their linear combinations comprehensively describe the possible biochemical transformation supported by a metabolic network.

Figure 4 shows that a given subgraph vector (reaction composition vector, RCV) could be projected onto several different PIVs with equally high scores. Multiple best projections were observed for both the liver and adipocyte models, where the multiplicity resulted from the sharing of reaction sequences between different EFMs. For very large networks, it would be impractical to investigate each of these candidate projections in detail. Therefore, the shared reactions of the best projections were extracted as the common pathway fragment (CPF), whose similarity with the candidate module was then assessed by calculating the match score. Table 1 lists the CPFs and the associated biochemical functions of the corresponding modules. A CPF is only a piecewise contiguous portion of an EFM, and therefore does not satisfy the non-decomposability property of an EFM. Nevertheless, calculation of the CPF facilitated the systematic association of a candidate module with a characteristic biochemical function. This method of association should be especially useful when computational limitations preclude a complete enumeration of all EFMs, and there is only partial knowledge about the possible array of functions of a metabolic network.

Figure 5a and b shows representative plots of the overall projection (PS) and match scores (MS), respectively, for the liver (control and day 2 post-burn) and adipocyte (day 4 and 12 post-induction). For both PS and MS, a higher score generally implies a better match with a feasible pathway (as defined by an EFM). Overall, the PS decreases and the MS increases with the iteration number. The decrease in PS was interrupted when...
a partition generated one or more new subgraphs. A biphasic behavior was also noted for the MS curve. The initial decrease in MS reflects edge removals that do not lead to the formation of a subgraph. Once an edge removal isolates one or more subgraphs, there is a sharp decrease in the lengths of the consensus pathway fragments (CPFs). Larger modules, and hence larger CPFs have higher chances of mismatches. Thus, the MS eventually levels off and rises as the later iterations generate smaller sub-graphs and more isolated nodes.

Table 1. Characteristic module function(s)

<table>
<thead>
<tr>
<th>Modules(*)</th>
<th>CPF reactions</th>
<th>Function(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liver</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1(9)</td>
<td>2–4,14–16,18–21, 35–36</td>
<td>KB</td>
</tr>
<tr>
<td>2(5)</td>
<td>2–4, 6–7</td>
<td>G</td>
</tr>
<tr>
<td>3(7)</td>
<td>1–4, 13–14</td>
<td>TA</td>
</tr>
<tr>
<td>Adipocyte</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1(25)</td>
<td>1–6, 8–10,12–16, 30</td>
<td>L</td>
</tr>
<tr>
<td>2(2)</td>
<td>27 and 39</td>
<td>AA</td>
</tr>
</tbody>
</table>

*Number of single nodes are shown in parentheses. Abbreviations: KB, ketone body metabolism; G, glucose metabolism; TA, transamination; L, lipogenesis and AA, amino acid metabolism.

Fig. 5. Average projection and match scores of the liver (a) and adipocyte (b) models at various iteration numbers. Dashed boxes indicate iteration numbers corresponding to the optimal partitions (see accompanying test).

3.4 Physiological perturbations

Finally, we investigated how the physiological perturbations experienced by a metabolic network, and the resultant changes to the connection diversity, could lead to a reorganization of its modules. For the rat liver, changes to the connection diversity were calculated as adjustments to the flux distributions on days 1, 2, 3, 4 and 7 following a burn injury (Lee et al., 2003). For the adipocyte, flux distributions were calculated for days 4, 8, 12 and 16 following the application of an adipogenic induction cocktail. Table 2 summarizes the algorithm outputs for the various liver and adipocyte time point flux distributions. In the case of the liver, two distinct partitions were found for the early (days 1, 2 and 3 post-burn) and late (days 4 and 7) stages of the hypermetabolic response. Interestingly, the number and composition of the reaction modules for the later stage were essentially identical to the sham-burn control case (day 0). This suggests that the hypermetabolic response was dissipating, and liver metabolism was returning to its basal state, consistent with prior reports on the first-week time course of burn injury in rats (Lee et al., 2003). The optimal partitions of days 1, 2 and 3 post-burn were nearly identical (Supplementary Material, Fig. S1) and consisted of four unevenly sized subgraphs.
Liver 0 Fast 3 KB, G, TA related hepatic hypermetabolism (Lee *et al*., 2007). Measurements on intracellular TG (*Si et al.*, 2007) showed continued increase in lipid loading through day 16, which is presumably reflected in the completion of the malate cycle. The modular decomposition of metabolic reaction networks 4 CONCLUSIONS In this article we have extended the use of graph-based models to include quantitative data on connectivity distribution in the characterization of modular structures in metabolic networks. A methodologically significant aspect of our analysis is the projection-based scoring used to evaluate the candidate modules. The scoring is based on the fractional overlap between a given subgraph and an inventory of biochemical transformation routes supported by the parent network. As the inventory was derived by calculating the elementary flux modes of the network, partition selection favored the formation of sub-graphs that resembled coherent metabolic pathways with defined inputs and outputs.

In conjunction with the partition algorithm, we used the scoring method to examine the organizational flexibility of liver and adipocyte metabolic networks by characterizing their modularity at various times following a major physiological perturbation. Here, organizational flexibility refers to how the different perturbations give rise to different groupings of active reactions. The results clearly depended on both the cell type and the physiological setting, which, respectively, determined the connectivity distribution and the connection engagements. Importantly, the modules calculated using the metabolic flux weights showed better agreement with known biochemistry compared to the modules derived solely based on reaction stoichiometry. For example, the unweighted partition in Figure 2b fails to connect 2-oxoglutarate (AKG) and SUCC, which are substrate and product, respectively, of a reaction in the TCA cycle. The partition in Figure 3b scatters several mitochondrial metabolites across the three different subgraphs, and does not reflect the expected biochemical closeness of the TCA and malate cycle intermediates. This type of partition can mask the involvement of a key metabolite in multiple pathways, in this case the participation of citrate (CIT) in both the TCA and malate cycles. These results suggest that biologically meaningful refinement of the modular structures may be obtained by taking into account the extents of biochemical interactions between the metabolites.

Taken together, our results suggest that the functional organization of a metabolic network depends both on its connectivity, which is fixed by the stoichiometry of its component enzymes, as well as the diversity of its connections, which reflects its overall activity state and is determined by the physiological stimuli addressing the network. For example, if an edge lies on many different shortest paths due to the stoichiometry of the corresponding reaction, then this edge will almost certainly have a high betweenness centrality regardless of the flux distribution. Conversely, if there are two or more alternative reaction paths between two metabolites, then the shortest path is determined by their relative activities, irrespective of the number of reaction steps in each path. In conclusion, the perturbation-dependent reconfiguration of modules points to a flexible design, which would allow a metabolic network to perform a variety of biochemical transformation operations using a finite set of reactions. Whether this design principle is common to metabolic reaction networks of other cells or organisms or other types of biochemical networks remains to be explored through

<table>
<thead>
<tr>
<th>Group (days)</th>
<th>Condition</th>
<th>Modules</th>
<th>Function(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liver</td>
<td>0</td>
<td>Fast</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>1, 2 and 3</td>
<td>Fast + Burn</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>4 and 7</td>
<td>Fast + Burn</td>
<td>3</td>
</tr>
<tr>
<td>Adipocyte</td>
<td>4</td>
<td>Young</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>Differentiated</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>12 and 16</td>
<td>Mature</td>
<td>2</td>
</tr>
</tbody>
</table>

*Refer to Section 2.1 for description of treatment groups. Abbreviations: TA1, transamination for anaplerosis; TA2, transamination for urea production and M, malate cycle; see also Table 1 footnote.
future work. Prospectively, the approach presented here could be used to associate reaction module configurations with particular types of physiological perturbations, and thereby obtain further insight into the functional design of biochemical networks.

ACKNOWLEDGEMENTS
The authors gratefully acknowledge Prof. Anselm Blumer in the Department of Computer Science at Tufts University for his help in implementing the edge betweenness centrality algorithm. This work was in part funded by the NIH grant no. 1 R21DK67228.

Conflict of Interest: none declared.

REFERENCES