Barrier Trees for Metabolic Adjustment Landscapes

Christoph Flamm[1], Chris Hemmingsen[2], and Daniel Merkle[2]

[1] Institute for Theoretical Chemistry, University of Vienna, Austria
[2] Department for Mathematics and Computer Science, University of Southern Denmark, Denmark
daniel@imada.sdu.dk, xtof@tbi.univie.ac.at

Abstract

We construct and analyze a discrete fitness landscape called *metabolic adjustment landscape*, from sub-networks covered by different productive flux distributions of a metabolic network. The topological structure of this landscape, i.e., the local minima and saddle points, can be compactly represented as a hierarchical structure called *barrier tree*. The switching from one local optimal flux pattern to another one is accompanied by adjustment costs, since genes have to be turned on or off. This phenomenon gives raise to saddle points in the metabolic adjustment landscape. Our approach allows calculating the minimal cost pathway that connects any two local minima in the landscape. Furthermore, our method yields a detailed ordering which reactions have to be (de-)activated to switch from one flux distribution to another one with minimal adjustment costs. Such a mechanistic hypothesis can guide experimental verification. We apply our approaches to a network describing the central carbon metabolism of *E. coli*.

Introduction

In recent years information from high-throughput sequencing and metabolomics has been integrated into genome-scale high quality reconstructions of metabolic networks of various organisms (Thiele and Palsson, 2010). These networks are a valuable computational resource in areas such as metabolic engineering, white biotechnology or synthetic biology. In particular questions how changes in the genetic setup of organisms influence the distribution of mass fluxes through the metabolic network have shifted into the focus of research. Flux balance analysis (FBA) (Orth et al., 2010) is among the most popular computational techniques to calculate flux distributions for metabolic networks. Since FBA is an optimization method, the organismal phenotype is usually defined in the form of a biological objective function (Feist and Palsson, 2010) which is optimized under additional constraints that balance or impose bounds on the system.

Another set of fundamental question, very appealing to theorists, focus on how catalyzed reaction networks, i.e. metabolisms, evolve and how novel reaction chemistry emerges during the evolutionary process. In Forst et al. (2006) it was shown that the structure of metabolic networks at the level of the individual chemical reactions, contains enough information for the accurate reconstruction of the phylogenetic relationships. In Flamm et al. (2010) a graph grammar based computational framework for the co-evolution of the enzymes and the metabolic network was described and several scenarios of metabolic network evolution were analyzed. Complexity questions on how to find chemical motifs in chemical reaction networks were analyzed in Andersen et al. (2012). In Schuetz et al. (2012) a large data set of flux distributions measured with ^{13}C-labeling experiments was analyzed to unveil the principles that govern the distribution and change of metabolic fluxes in *E. coli* under varying conditions. A multi-objective optimization approach together with FBA was used since a combination of several competing objective functions turned out to be best suited for the analysis of the entire data set. Interestingly the study found that flux distribution in *E. coli* is governed by the principle of maximizing production only up to the degree where an easy switch of nutrients is still possible. This behavior seems to be a perfect adaption of *E. coli's* metabolism to a fluctuating environment where a sudden deficiency of a set of nutrients can be compensated by a switch to other ones accompanied by an easy restructuring of the original flux distribution to the new situation. This comes at the cost that production of compounds (e.g. for building up biomass) in the metabolic network can never be fully maximized to the theoretical limit within the metabolic network of *E. coli*.

While there are several sophisticated computational methods to assign one "optimal" flux distribution to a metabolic network (for a review see Lewis et al. (2012)), to the best of our knowledge, this is the first study that analyzes the entire variety of (optimal) flux distributions over a metabolic network with a fixed genetic setup, but varying activity for subsets of genes in the discrete landscape metaphor.

The brief outline of this paper is as follows: we will first introduce barrier trees for discrete fitness landscapes. Then we will explain different methods how to create metabolic adjustment landscapes, which we use for barrier tree ana-

lyzes. We expect the reader to be familiar with the concept of FBA, for an in depth introduction we refer to (Palsson, 2006). In the results section we will first present an artificial example to illustrate our approaches, and then analyze two metabolic adjustment landscapes of *E. coli*.

Barrier Trees

The switching between different productive flux distributions is accompanied by flux adjustment costs, since genes have to be regulated to achieve the change in the flux distribution. This raises saddle points which connect basins associated with optimal (productive) flux distributions. It seems therefore natural to apply the theory of discrete fitness landscapes to characterize the discrete landscape induced by flux adjustment costs and to get a deeper understanding of its topological and functional structure.

Formally, we define a landscape as a triple $(X; N; f)$ consisting of a set of configurations X, a topological structure N that determines the mutual accessibility of configurations, and a cost or "fitness" function $f : X \to \mathbb{R}$. In our case elements in X will be metabolic networks. The neighborhood relation N is typically defined by the "move set". In this contribution we will restrict ourselves to the simplest case in which the configuration space $(X; N)$ is a finite directed graph $G = (X; E)$ with vertex set X and edge set E. Here edges connect configurations that can be inter-converted by a single move. (If the move-set is symmetric, $(X; N)$ can also be represented as an undirected graph.) The fitness value of the lowest saddle point separating two local minima $x \in X$ and $y \in X$ is

$$\hat{f}[x, y] = \min_{\mathbf{p} \in \mathbb{P}_{xy}} \max_{z \in \mathbf{p}} f(z) \tag{1}$$

where $\mathbb{P}_{xy}$ is the set of all paths $\mathbf{p}$ connecting x and y by a series of consecutive operations from the move set.

If the fitness function is non-degenerate, i.e., two configurations have distinct fitness values, then there is a unique saddle point $s = s(x, y)$ connecting x and y characterized by $f(s) = \hat{f}[x, y]$. The extension to degenerate fitness functions is discussed in detail in Flamm et al. (2002). To each saddle point s there is a unique collection of configurations $B(s)$ that can be reached from s by a path along which the fitness value never exceeds $f(s)$. In other words, the configurations in $B(s)$ are mutually connected by paths that never go higher than $f(s)$. This property warrants calling $B(s)$ the *valley or basin below the saddle* s. Furthermore, suppose that $f(s) < f(s')$. Then there are two possibilities: if $s \in B(s')$ then $B(s) \subseteq B(s')$, i.e., the basin of s is a "sub-basin" of $B(s')$, or $s \notin B(s')$ in which case $B(s) \cap B(s') = \emptyset$, i.e., the valleys are disjoint. This property arranges the local minima and the saddle points in a unique hierarchical structure which is conveniently represented as a tree, termed *barrier tree*.

Landscapes of Metabolic Adjustment Networks

A metabolic (reaction) network, is usually represented as a hyper-graph, where the nodes indicate the set of chemical compounds that are connected by hyper-edges corresponding to the set of chemical reactions R. The power-set of R induces a whole series of "smaller" instances of metabolic networks, where a subset of reactions is removed from the original metabolic network. A *metabolic adjustment landscape* is a directed graph (X, E) with a vertex set (or configurations) X, where each vertex x corresponds to one of the metabolic network induced by an element of the power-set of R. The topological structure of the metabolic adjustment landscape is defined by the neighborhood function N, which determines how the "different" metabolic networks are connected via operations from the move set. In the simplest case the move set is defined by adding or removing exactly one reaction. An edge is labeled with the name of the reaction that has been added or removed (see right of Fig. 1). If the reaction is removed, there is a "-" in front of the name. Two vertices are connected by two edges going in opposite directions, if it is possible to go from one network to the other by adding/removing a reaction. The aforementioned configurations space is converted into a discrete landscape by assigning a fitness value $f(x)$ to each configuration x. However, some of the networks are not "viable" in the sense, that these networks cannot support flux between predefined input and output nodes. The viability of a network is decided by running a FBA with a predefined objective function. If the network is viable the assigned fitness value is the number of reactions in that network and infinity otherwise. The rational behind this fitness function is, that the expression of genes, to provide the corresponding chemical transformation in the network in the form of enzymes, is a costly process. In that sense the chosen fitness function quantifies the active or expressed portion of the genetic setup i.e. all possible enzymes encoded in the genome of an organism. More formally

$$f(x) = \begin{cases} |x| & \text{if network } x \text{ is viable} \\ \infty & \text{otherwise,} \end{cases} \tag{2}$$

where $|x|$ is the number of reactions (or hyper-edges) in a metabolic network $x \in X$.

Unrestricted and Restricted Landscapes

Furthermore, we distinguish two cases of metabolic adjustment landscapes. First, the *unrestricted* case, where any active reaction can be removed and any inactive reaction may be added. In other words two networks are connected if their symmetric difference consists of exactly one reaction. Since any reaction can be removed or added networks can be generated which cannot support flux between source and sink nodes. Hence, unviable networks are valid configurations in unrestricted metabolic adjustment landscapes. Second,

the *restricted* case, where adding and removing reactions is constrained by the following rationales: (i) reactions which would cut the last connection between source and sink nodes cannot be removed (this guarantees, that every network is viable) (ii) reactions can be inserted if all their reactants are produced by "other" reactions already in the network (iii) reactions can be removed if they do not disable following reactions. A reaction is disabled if no other reaction produces its reactants.

Fitness Functions for Networks in Metabolic Adjustment Landscapes

The standard fitness functions used for FBA are usually defined via a deviation to a given target flux, aiming at biomass production, or aiming at maximizing the production of one or more products. For a critical assessment of the assumptions made in FBA see Schuster et al. (2008). All these fitness functions do not account for the fact, that adding a reaction to a metabolic network induces costs (i.e., the genetic setup needs to be more complicated, as the corresponding enzymes need to be available). More formally this can be phrased as a Lemma.

Lemma 1. *Within the landscape of a metabolic adjustment network, any of the standard fitness functions for FBA (Schuetz et al., 2007) leads to a barrier tree with exactly one leaf node.*

Proof. A landscapes of metabolic networks has a barrier tree with exactly one leaf node (i.e., a barrier tree without any barriers) iff the landscape is (possibly weakly) unimodal. To show the unimodality, consider a flux balance analysis on a metabolic network x that reaches an optimal fitness score $f(x, v)$ with a certain optimal flux distribution v (the flux vector v assigns fluxes to all reactions). The core observation to be made is, that extending network x to a network x' can not lead to a worse fitness score, as the flux induced by v in the extended network reaches at least the fitness score $f(x, v)$ in the network x'. Adding all possible reactions successively will therefore always lead to the best possible fitness score (without the need to cross a barrier) and two strict local minima can not exist in the metabolic adjustment landscape. □

Based on the above observation that no barriers can appear if a standard fitness functions for a FBA is used to analyze metabolic adjustment landscape, we use as a natural motivation in creating a new fitness function, that adding a reaction needs to be penalized, i.e., the genetic setup is more complicated. We therefore defined the fitness function as given in Eqn. 2. Note, that this ensures that extending a metabolic network by a reaction leads to a worse fitness score, even if the added reaction is not used in an optimal flux in the extended network (where optimal is meant wrt. to the objective function used for the FBA) .

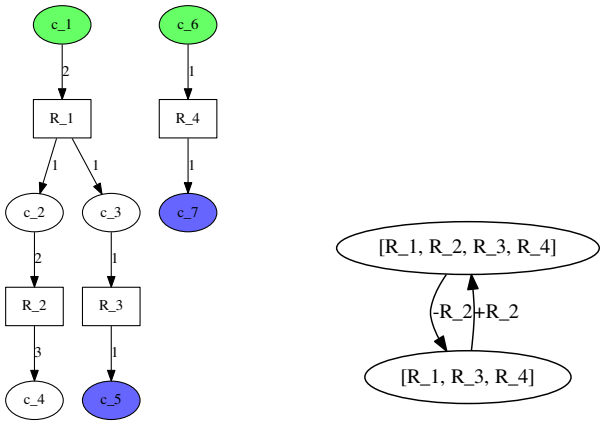

Figure 1: Toy example for a restricted metabolic adjustment landscape: The four given reactions are $R_1, \ldots, R_4$, educts are c_1 and c_6 (green circles), and the target compounds are c_5 and c_7 (blue circles). Left: the metabolic network consisting of all reactions. Right: the resulting restricted metabolic adjustment landscape.

Fig. 1 shows an illustrative example with four reactions $\mathcal{R} := \{R_1, \ldots, R_4\}$, the set of educts is $\{c_1, c_6\}$ (sources), and the set of products is $\{c_5, c_7\}$ (sinks). The restricted landscape has two nodes only, the fitness values of the two metabolic networks are 3 (in the network without reaction R_2) and 4 (in the network with R_2). The unrestricted landscape consists of $2^{|\mathcal{R}|} = 16$ metabolic networks. The additional 14 metabolic networks are all unviable, since removal of reactions R_1, R_3 or R_4 disconnects the last path between the source and the sink nodes, rendering a productive flux impossible. Therefore the fitness value of these networks is set to ∞.

Results

In the first part we will present results for an artificial example to illustrate our approach to infer a barrier tree from a metabolic adjustment landscape. In the second part we will apply our approach to the metabolic adjustment landscape derived from the main reactions in the central carbon metabolism of *E. coli*.

Artificial Example

We used a network of 23 compounds and 25 reactions as depicted in Fig. 2. Compound c_{17} is the only substrate (source) and compound c_5 is the only product (sink) compound to be produced in a quantity of 0.1. We computed the restricted metabolic adjustment landscape, which consists of 561 metabolic networks. This is relatively small in contrast to the unrestricted landscape which is comprised of 2^{25} networks, the majority of these networks are unviable. The restricted landscape was analyzed: it has 5 local minima, all of them reach a fitness value of 10 reactions. Note,

that any local minima can not have an unused reaction in an optimal flux distribution, as this reaction could be removed (leading to a better fitness in the metabolic adjustment landscape) while keeping the same fitness for the FBA of the reduced network. The barrier tree is shown in Fig. 3, the metabolic networks that correspond to minima 1 and 2 are depicted in Fig. 2 (the red nodes and edges, are only shown for illustration purposes and are not part of the solution networks which correspond to the local minima 1 and 2, these parts are shown in green). The barrier tree calculation allows backtracking the minimum cost path between any two minima in the barrier tree. Such a minimum cost path leading from minimum 1 to minimum 2 on the restricted metabolic adjustment landscape would (i) add the reactions with indices 10, 13, and 15, (ii) remove the reactions with indices 7, 8, and 12 (this became possible since step (i) introduced a parallel path connecting the source to the sink node keeping the network viable). This minimum cost path results in a barrier of 13 (denoted as B1 in Fig. 3) between the two local minima 1 and 2 since the saddle point network contains exactly 13 active reactions. The relations between the local minima and the minimum saddle points in the metabolic adjustment landscape can compactly be represented as a barrier tree Fig. 3. The internal tree nodes are the minimum saddle points between local minima which are located at the leaf nodes. For a change to a flux pattern that uses only the pathway via reaction indices 20, 19, ..., a barrier of height 20 needs to be crossed. This barrier is denoted as B2 in Fig. 3.

The Central Carbon Metabolism as a Restricted Landscape

The Central Carbon Metabolism (CCM) is a union of well known catabolic pathways, such as glycolysis and tricarboxylic acid cycle (TCA), and a minimal number of "interface reactions" to important anabolic pathways, found in all three kingdoms of life. The representation of the network we use originates from de Figueiredo et al. (2008). Their network has 37 reactions, provided that both directions of reversible reactions are counted separately. The target compound is glucose-6-phosphate (G6P). The network can "feed" on different substrates (source compounds) to achieve the production of G6P (target compound). Among them are Acetyl coenzyme A (AcCoA) created as degradation product of the fatty acids metabolism, the two amino acids Alanine (Ala) and Aspartic acid (Asp), or Pyruvate (Pyr), the simplest α-keto-acid derived as end-product of Glycolysis. We pruned unused reactions from the original network to increase computational speed. For this an elementary mode (EM) analysis (Papin et al., 2004) of the CCM with and without Isocitrate lyase EC 4.1.3.1 (ICL) and the Malate-Aspartate shuttle (MAS) using each substrate was conducted. (An elementary mode is a certain feasible flux distribution; all elementary modes can be created by combining the extreme pathways (EPs) of the network, which

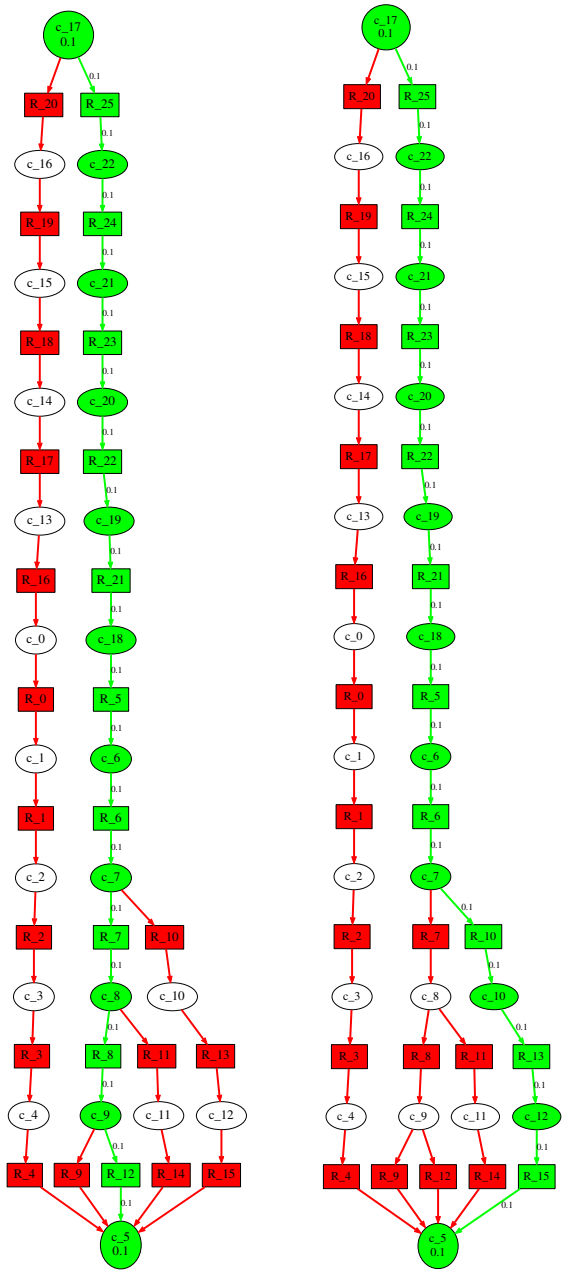

Figure 2: Minima 1 (left) and 2 (right) of the artificial example. Target is "c_5_cell". c_17_cell is a substrate. Note that the red nodes are not in the corresponding metabolic network, but are only depicted for illustration reasons.

are formally derived from the basis vectors of the null-space of the stoichiometric matrix of the network. The EPs are therefore a subset of the elementary modes.) From the set of all possible EMs it can be seen that some directions of reversible reactions are never used. We also removed the linear pathway for creation of "G6P" from Fructose-1,6-bisphosphate (F1,6PP) using the 3 enzymes Fructose-

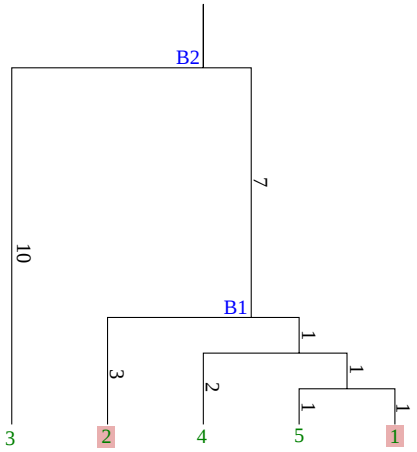

Figure 3: Barrier tree for the metabolic adjustment landscape of the artificial example: minima 1 and 2 (both reaching objective value 10, cmp. Fig. 2) are connected via the lowest saddle B1, which has an objective value of 13. The barrier B2 has objective value 20.

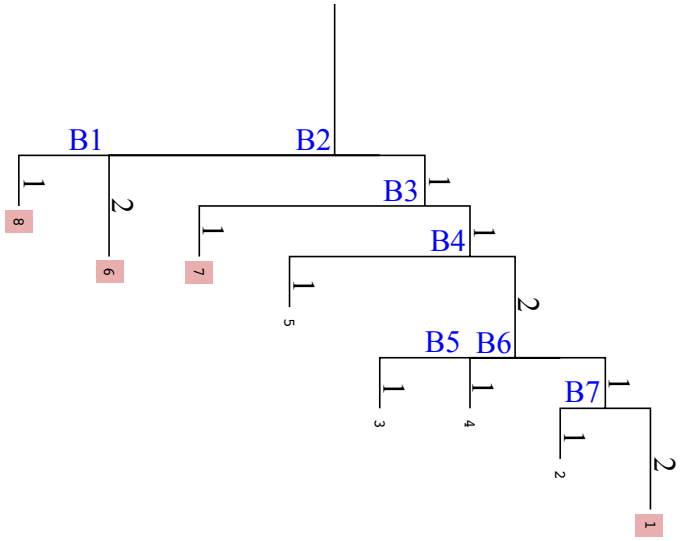

Figure 4: Barrier tree from the restricted metabolic adjustment landscape of CCM for "F1,6PP" production as objective function. Barriers are marked B1 to B7. Min 1: Use Asp. Min 2: Use Ala. Min 4: Use Asp, identical to Min 3, but with ME1_1 and ME1_2 blocking each other from being removed. Min 3: Use Ala. Min 5: Use Glu (utilize pathway from OG to MAL). Min 7: Use Glu (utilize pathway from OG to MAL). Min 6: Utilize AcCoA via ICL/MAS straight to OAA from Mal. Min 8: Utilize AcCoA via ICL/MAS detour over ME1_1 and PC fro Mal to OAA.

bisphosphatase EC 3.1.3.11 (FBP1), Phosphofructokinase EC 2.7.1.11 (PFKL) and Glucose-6-phosphate-isomerase EC 5.3.1.9 (PGI), as any solution capable of producing "G6P" would use "F1,6PP" and this pathway. After pruning the network contains 22 reactions, having "F1,6PP" as target compound. The pruned network is depicted in Fig. 7.

Barrier Tree We use our approaches to determine the restricted metabolic adjustment landscape for the network given in Fig. 7. For the FBA "F1,6PP" served as target compound and AcCoA, Ala, Asp and Pyr as individual source compounds. We used the tool FASIMU (Hoppe et al., 2011) to convert the reaction networks into the integer linear programs (ILPs), which were solved using IBM's commercial program IBM ILOG CPLEX (2012) (currently freely available for academic purposes). It took 2.2 seconds to build the restricted metabolic adjustment landscape, which contains 12853 vertices. It took 36 minutes for FASIMU and CPLEX to formulate and run all 12853 simulations, i.e. on average 5.95 simulations per second. The resulting barrier tree can be seen as Fig. 4. The barrier tree has 8 local minima, the flux distributions of minima 1, 7, and 8 are depicted in Fig. 5.

Minima Although the biological discussion of the results is out of the scope of this paper, it should be noted that the barrier tree nicely illustrates the shift from using the glyoxylate shunt in CCM (minima 8 (cmp. Fig. 5) and 6 (not depicted)) towards not using it (all other minima).

Barriers The leaf nodes of the barrier tree show the CCMs different abilities to produce "F1,6PP". The barriers are marked B1 to B7. Their fitness indicates the least amount of reactions a network needs to have when passing a barrier that is necessary to connect two metabolic networks.

Let M_7 (resp. M_8) be the network that minima 7 (resp. 8) represent. Imagine we want to modify M_7 such that it becomes M_8, while always being able to maintain "F1,6PP" production and minimizing the maximal fitness of all intermediate networks. The computed minimal cost path is as follows:

```
1                                  Minima 7     Score: 11
2  Add GPT_1                       Saddle B3    Score: 12
3  Remove ME1_1                                 Score: 11
4  Remove FH_1                                  Score: 10
5  Remove SDHA, SDHB, SDHC, SDHD_1              Score: 9
6  Remove SUCLG2, SUCLG1, SUCLA2_1              Score: 8
7  Remove OGDH, DLST, DLD          Minima 2     Score: 7
8  Add GOT1_1                      Saddle B7    Score: 8
9  Remove PC                                    Score: 7
10 Remove GPT_1                    Minima 1     Score: 6
11 Add CS                                       Score: 7
12 Add ACO1, ACO2_1                             Score: 8
13 Add ICL                                      Score: 9
14 Add SDHA, SDHB, SDHC, SDHD_1                 Score: 10
15 Add FH_1                                     Score: 11
16 Add MDH1, MDH2_1                             Score: 12
17 Add MAS                         Saddle B2    Score: 13
18 Remove GOT1_1                                Score: 12
19 Remove GLUD1_2                  Minima 6     Score: 11
20 Add ME1_1                                    Score: 12
21 Add PC                          Saddle B1    Score: 13
22 Remove MDH1, MDH2_1             Minima 8     Score: 12
```

The path never goes above 13 reactions. Although, based on

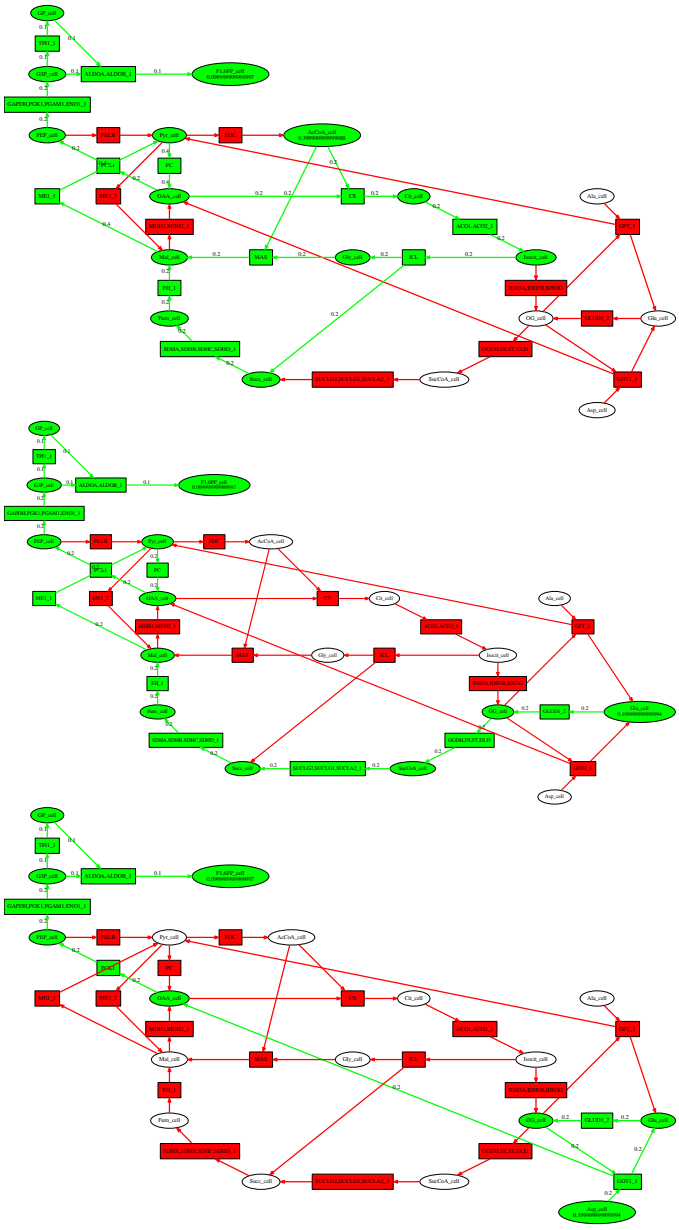

Figure 5: Minima 8 (top), 7 (middle), and 1 (bottom). Note, that the red nodes are not in the corresponding metabolic network, but are only depicted for illustration reasons.

the "closeness" of minima 7 and 8 in the barrier tree, one might think that connecting the two minima while not using more than 13 reactions would result in a short path, the example shows this is in general not trivial. It takes 22 steps of adding and removing a reaction. First "GPT_1" is added to enable a shift of substrate from "Glu" to "Ala". Using "Ala" means the 4 reactions along the path from "OG" to "OAA" are no longer needed. These are removed, as well as "ME_1", which is not needed either. Now "GOTH1_1" is added to switch from using substrate "Ala" to "Asp". This removes the need for "PC" and "GPT_1". Now all the reactions part of the glyoxylate cycle are added, 2 of which had been removed just some steps ago. Now a shift from using "Asp" to using "AcCoA" as a substrate can be made. This transition state is barrier "B2" and is shown in Fig. 6. "GOT1_2" and "GLUT1_2" is now no longer needed. Last thing needed is to use "ME1_1" and "PC" to reach "OAA" from "Mal", instead of using "MDH1,MDH2_1".

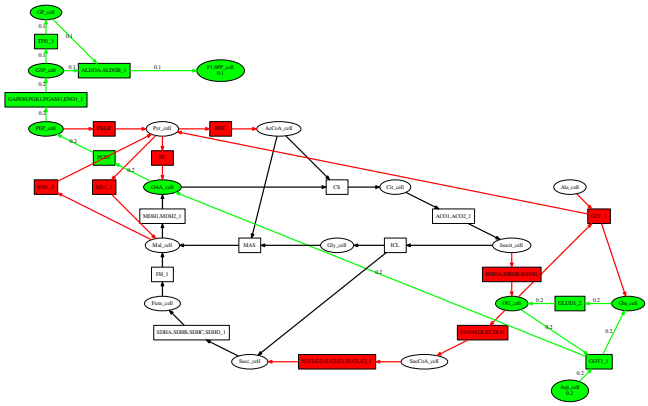

Figure 6: Barrier B2. Note, that the red nodes are not in the corresponding metabolic network, but are only depicted for illustration reasons. The black lines depict reactions that *are* in the metabolic network but are not used in the optimal flux distribution.

The Shift from Using Fatty Acids to Amino Acids as an Unrestricted Landscape

This section presents results that illustrate how our approach is used in order to analyze the shift between different given fluxes. Using this method instead of finding all sub-networks reduces the number of simulations, but also requires some sensible choices for the two networks. Here, we choose the networks such that both networks are subsets of the CCM and produce "F1,6PP", but one does it using fatty acids ("AcCoA"), and the other one does it using amino acids("Ala", "Glu", or "Asp"). For a biochemical discussion on the usage of fatty acids or amino acids to produce glucose see de Figueiredo et al. (2008). Both networks are shown in Fig. 7. Note, that in contrast to the restricted case, we disallow here the removal of reactions that appear in both networks (depicted as yellow hyper-edges in Fig. 7). We use the unrestricted transformation method in order to analyze the metabolic adjustment landscape. To transform the base networks topology to the target networks, 6 reactions must be removed and 6 reactions must be added. It took 0.6 seconds to build the resulting landscape, which contains $2^{12} = 4096$ vertices. It took 10 minutes for FASIMU and CPLEX to formulate and run all 4096 simulations (i.e., on average 6,83 simulations per second).

Barrier Tree The barrier tree for the metabolic adjustment landscape is depicted in Fig. 8. The tree has 4 minima that

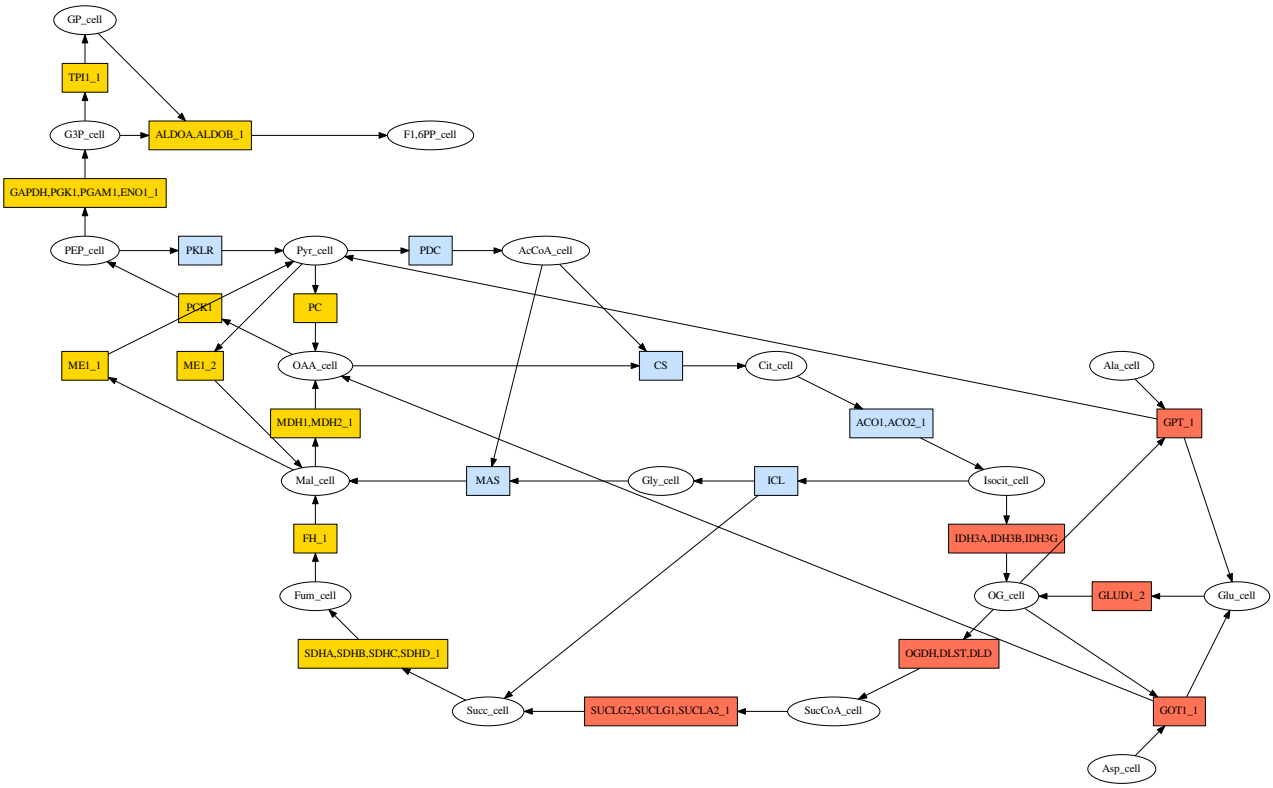

Figure 7: Two networks with an overlap that use different source compounds to analyze substrate switching: colored in yellow are reactions common to both networks, colored in blue are the additional reactions specific for the network utilizing fatty acids, and reactions colored in red are specific for the network utilizing amino acids.

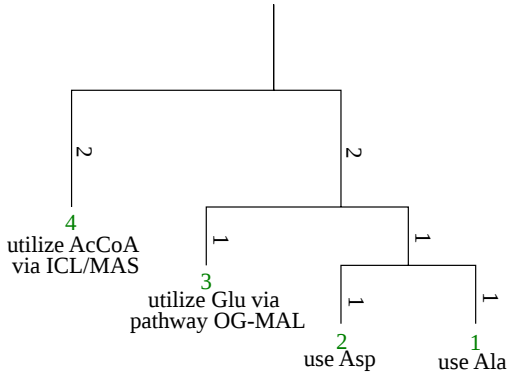

Figure 8: Barrier tree illustrating the shift from utilizing amino acids (minima 1, 2, and 3) to utilizing fatty acids (minima 4).

nicely illustrates the usage of amino acids (minima 1-3) or fatty acids (minimum 4) as source compounds to produce the target compound "F1,6PP". While changing the flux pattern between different amino acids requires passing only a small barrier of 1-2 additional reactions, switching to a flux pattern that utilizes fatty acids, in contrast, requires to pass a rather high barrier of additional 4 reactions.

Minima 2 of the four minima are depicted in Fig. 9. Reactions that have been removed from the network are marked red, and the flux distribution is shown in green. Each minimum uses different substrates. Minima 1 and 2 show optimal use of both "Ala" and "Asp". They both avoid using the metabolic pathway from "OG" to "MAL", and also do not use "ME1_2". The reason for the large amount of unused "black" reactions is, that they are not allowed to be removed since they are present in the intersection of the two networks. Minima 3 uses "Glu" as source compound and requires the usage of part of the TCA cycle (from "OG" to "MAL") to connect to the target compound "F1,6PP". The barrier tree again suggests that using "AcCoA" as the only substrate requires more active enzymes than in the case of "Ala" or "Asp". In other words the amino acids are a "much cheaper" resource to produce "F1,6PP" from then "AcCoA".

Conclusions

We introduced a systematic approach to characterize the flux landscapes of a metabolic network. The genetic setup of the metabolic networks is the always same (forming a supernetwork); embedded are different optimal flux distributions for varying substrate usage and/or target compound productions. Switching between different flux distributions is ac-

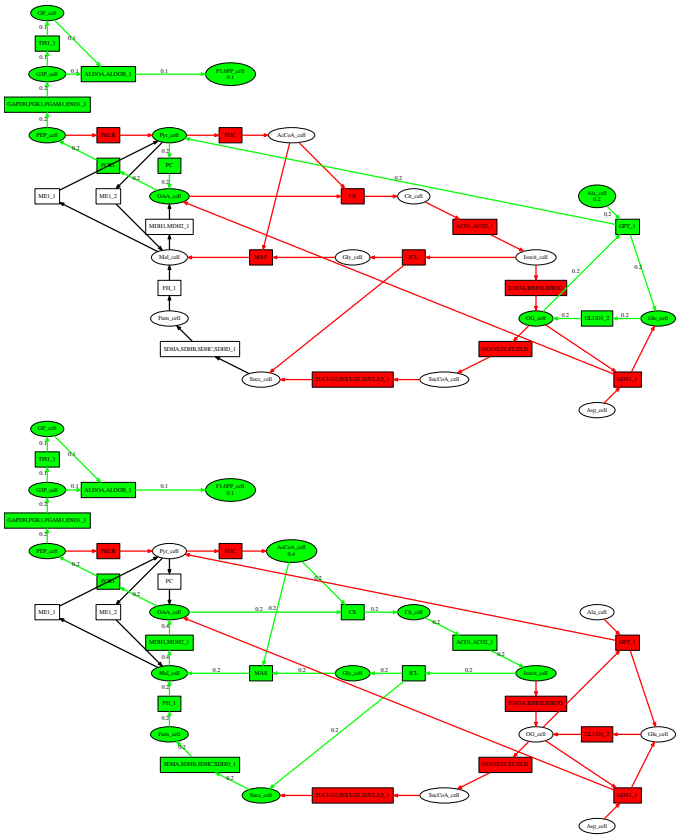

Figure 9: Minima 1 (top): Only uses substrate "Ala" in an optimal fashion by avoiding the metabolic pathway from "OG" to "MAL". Minima 4 (bottom): Only uses AcCoA as substrate and utilizes the glyoxylate shunt by activating the reactions "ICL" and "MAS". Note, that the red nodes are not in the corresponding metabolic network, but are only depicted for illustration reasons. Black edges are unused existing edges in the optimal flux distribution.

companied by adjustment costs since inactive genes have to be activated and active genes have to be deactivated. From the networks induced by the subset of active reactions (genes) a discrete landscape can be constructed, which we termed *metabolic adjustment landscape*. This landscape is analyzed in terms of local minima and connecting saddle points, and can be efficiently visualized in a hierarchical structure called *barrier tree*. The analysis allows us to find the cost for changing from one optimal flux pattern to another. Furthermore, for the first time, we can calculate in mechanistic detail how this minimal cost pathway looks like, in particular in which order the reactions have to be (de-)activated to achieve the change in the flux distribution. This mechanistic hypothesis can be tested by experimental approaches.

Acknowledgments

This work was supported by the Danish Council for Independent Research, Natural Sciences.

References

Andersen, J. L., Flamm, C., Merkle, D., and Stadler, P. F. (2012). Maximizing output and recognizing autocatalysis in chemical reaction networks is NP-complete. *J Sys Chem*, 3(1).

de Figueiredo, L. F., Schuster, S., Kaleta, C., and Fell, D. A. (2008). Can sugars be produced from fatty acids? A test case for pathway analysis tools. *Bioinformatics*, 24(22):2615–2621.

Feist, A. M. and Palsson, B. Ø. (2010). The biomass objective function. *Curr Opin Microbiology*, 13:344–349.

Flamm, C., Hofacker, I. L., Stadler, P. F., and Wolfinger, M. T. (2002). Barrier trees of degenerate landscapes. *Z Phys Chem*, 216:155–173.

Flamm, C., Ullrich, A., Ekker, H., Mann, M., Högerl, D., Rohrschneider, M., Sauer, S., Scheuermann, G., Klemm, K., Hofacker, I. L., and Stadler, P. F. (2010). Evolution of metabolic networks: a computational frame-work. *J Systems Chem*, 1(4).

Forst, C. V., Flamm, C., Hofacker, I. L., and Stadler, P. F. (2006). Algebraic comparison of metabolic networks, phylogenetic inference, and metabolic innovation. *BMC Bioinformatics*, 7(67).

Hoppe, A., Hoffmann, S., Gerasch, A., Gille, C., and Holzhütter, H.-G. (2011). FASIMU: flexible software for flux-balance computation series in large metabolic networks. *BMC Bioinformatics*, 12:28.

IBM ILOG CPLEX (2012). High-performance mathematical programming engine. v12.5, `http://www.ibm.com/software/integration/optimization/cplex`.

Lewis, N. E., Nagarajan, H., and Palsson, B. Ø. (2012). Constraining the metabolic genotype-phenotype relationship using a phylogeny of *in silico* methods. *Nature Reviews Microbiology*, 10:291–305.

Orth, J. D., Thiele, I., and Palsson, B. Ø. (2010). What is flux balance analysis? *Nature Biotechnology*, 28(3):245–248.

Palsson, B. (2006). *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press, first edition.

Papin, J. A., Stelling, J., Price, N. D., Klamt, S., Schuster, S., and Palsson, B. Ø. (2004). Comparison of network-based pathway analysis methods. *Trends Biotechnol.*, 22(8):400–405.

Schuetz, R., Kuepfer, L., and Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Molecular Systems Biology*, 3:119.

Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M., and Sauer, U. (2012). Multidimensional optimality of microbial metabolism. *Science*, 336(6081):601–604.

Schuster, S., Pfeiffer, T., and Fell, D. A. (2008). Is maximization of molar yield in metabolic networks favoured by evolution? *J Theor Biology*, 252:497–504.

Thiele, I. and Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, 5:93–121.