

Pathway identification by network pruning in the metabolic network of *Escherichia coli*

P. Gerlee*, L. Lizana and K. Sneppen

Center for Models of Life, Niels Bohr Institute, Blegdamsvej 17, 2100, Copenhagen, Denmark

Associate Editor: Dr. Jonathan Wren

ABSTRACT

Motivation: All metabolic networks contain metabolites which take part in many reactions, such as ATP and NAD, known as currency metabolites. These are often removed in the study of these networks, but no consensus exists on what actually constitutes a currency metabolite, and it is also unclear how these highly connected nodes contribute to the global structure of the network.

Results: In this paper we analyse how the *Escherichia coli* metabolic network responds to pruning in the form of sequential removal of metabolites with highest degree. As expected this leads to network fragmentation, but the process by which it occurs suggest modularity and long range correlations within the network. We find that the pruned networks contain longer paths than the random expectation, and that the paths that survive the pruning also exhibit a lower cost (no. of involved metabolites) compared to random paths in the full metabolic network. Finally we confirm that paths detected by pruning overlap with known metabolic pathways. We conclude that pruning reveals functional pathways in metabolic networks, where currency metabolites may be seen as ingredients in a well-balanced soup in which main metabolic production lines are immersed.

Contact: gerlee@nbi.dk

1 INTRODUCTION

The metabolic networks of many organisms have recently been mapped out and investigated, ranging from one of the simplest known organism *Mycoplasma pneumoniae* (Schuster et al., 2002) to *Escherichia coli* (Ma and Zeng, 2003), *Saccharomyces cerevisiae* (Herrgard et al., 2008) and the human metabolic network (Duarte et al., 2007). The general picture that emerges is that of a highly connected complex network of metabolites and reactions, sharing many features with other large-scale biological networks. For example the metabolites exhibit scale-free in- and out-degree distributions (Jeong et al., 2000), and the whole network also displays the small-world property (Fell and Wagner, 2000). Another typical feature is the so-called bow-tie architecture, which refers to the fact that the metabolites in the network can roughly be divided into three categories: the giant strong component (in which there is a path between any two metabolites), the substrate subset which feeds into the giant strong component and the product subset representing the output from the network (Ma and Zeng, 2003). It has also been shown that metabolic networks exhibit some degree of modularity, i.e. that they consist of subsets of segregated functions (Ravasz et al., 2002; Guimera and Nunes Amaral, 2005; Ma et al., 2004).

The analysis of metabolic networks is however hampered by the existence of so called "currency metabolites" such as ATP, NAD and protons, which take part in a large number of reactions. These few metabolites contribute significantly to the connectivity of the network, over-shadowing the actual pathway-like structure of the network. In order to focus on the actual properties of the network these metabolites are often removed when the network is analysed. However, no consensus exists on what constitutes a currency metabolites, and this has led to diverging opinions on the structure of metabolic networks. The most central metabolite in the *E. coli* network was for example glutamate in one study (Fell and Wagner, 2000), and pyruvate in another (Ma and Zeng, 2003). Considering different metabolites as currency, and thus removing a different set of nodes from the original network, they effectively analysed two different versions of the same network.

One way of classifying currency metabolites is by dividing all metabolites into internal and external (Schuster and Heinrich, 1996). External metabolites refer to those whose concentration is essentially constant such as ADP/ATP and water while internal have to be balanced with respect to production and consumption at steady state. In Schuster et al. (2002) the classification was done according to the degree of the metabolites, i.e. metabolites above a certain connectivity were considered external, and this lead to a fragmentation of the metabolic network into subnetworks that are biochemically meaningful. This is however only one possibility, and in Dandekar et al. (2003) a classification minimising the number of elementary modes was employed. This was done in order to reduce the number of elementary modes in metabolic networks, and consequently to avoid the combinatorial explosion which otherwise occurs in large networks. A systematic study of the removal of metabolites coupled with a modularity analysis was carried out by Holme and Huss (2007). They successively removed the metabolites with the highest degree until the modularity of the network reached a maximum. This maximum was found to be reached after removing between 1 and 6 currency metabolites.

In this paper we explore pruning as a way to capture the backbone of pathways that forms the metabolic network in *E. coli*. As explained in the method section we prune the network by removing metabolites in order of their degree, starting with the nodes with the highest degree and removing several nodes in one step if they have the same degree. Removing nodes in the network gradually reduce the giant connected component (GC) of the network, as indeed shown in fig. 1 (Callaway et al., 2000; Albert et al., 2000). To compare this reduction with a proper null expectation we compare with the randomised version of the network, generated as described in the method section (Maslov and Sneppen, 2002). One notices that the real network breaks down faster than one should expect,

*to whom correspondence should be addressed

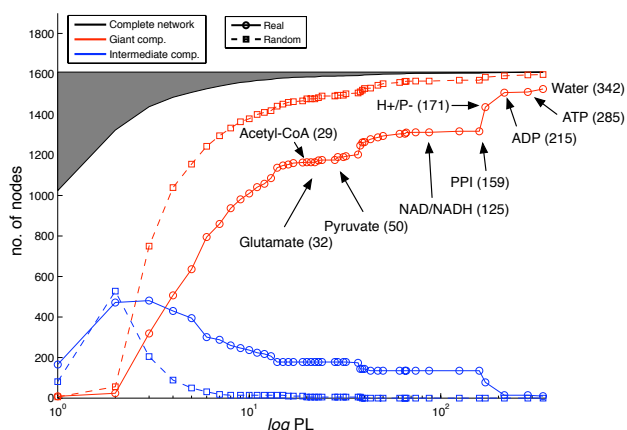


Fig. 1. The number of nodes in the giant component (red) and in intermediate components (blue), plotted as a function of the logarithm of the pruning level (PL). The arrows indicate at which pruning level the respective metabolites are removed. The shaded grey area shows the reduction in total network size during the pruning process.

suggesting that the metabolic network consists of modules which is glued together by currency metabolites, which is also evident from the larger fraction of nodes in components of intermediate size (components larger than 4 nodes but smaller than the GC). The modular features of the metabolic network reported by Holme and Huss (2007) are accordingly associated to relatively small clusters or pathways which are connected to the central metabolism through currency metabolites. A similar analysis was carried out for the giant strong component (GSC), which showed similar behaviour where the real network breaks down faster than its random counterpart (see fig. S4).

Using the metabolic network of *E. coli* as a model system we in this paper explore how the structural feature of pathways are associated to the use of “currency metabolites”. We will suggest that pruning can be used to identify biologically meaningful pathways in the network, and we will see that each such pathway tend to reuse the same currency metabolites in several of its reactions.

2 SYSTEM AND METHODS

The data for the *E. coli* metabolic network was downloaded from EcoCyc (28/10/08) (Karp, 2007) and was represented in its most general form as a directed bipartite graph with two types of nodes, metabolites and reactions. The network consists of 3311 nodes (1609 metabolite and 1702 reaction nodes) and 6658 edges between them. Reversible reactions are treated by representing each direction with a separate edge. The metabolite nodes exhibit broad connectivity distributions for both in- and outgoing links (Jeong et al., 2000) (see fig. S1), while the connectivity of the reaction nodes is distributed around a mean of approximately 2 in- and outgoing links (see fig. S2).

2.1 Pruning method

The pruning of the network was carried out by removing metabolite nodes in order of degree, starting with the nodes with the highest degree and removing

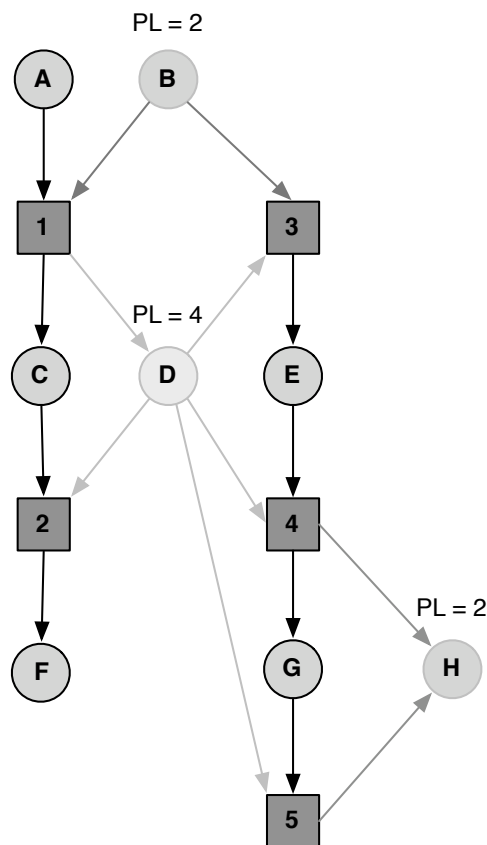


Fig. 2. Illustration of the pruning algorithm. This example network, where metabolites and reactions are visualised by circles and squares respectively, is unaffected until we have reached a pruning level of 4 at which the central node D is removed. The next step is the removal of nodes B and H at pruning level 2. The rest of the network remains intact as all remaining nodes have in- and out-degree one.

several nodes in one step if they have the same degree. In this process we treat in- and outgoing links the same, meaning that a node with in-degree 10 and out-degree 2 is removed in the same step as one with in-degree 2 and out-degree 10. The extent to which the network has been pruned we call the “pruning level” and refers to the upper limit of the metabolite degree still present in the network (i.e. at pruning level n nodes with at most degree $n - 1$ are present in the network). Note that reaction nodes are not removed in the pruning process due to their degree, only when they become disconnected from the network are they removed. An example of the application of the algorithm can be seen in fig. 2, which shows the effect of the pruning on a small example network. The first node (D) is removed at pruning level 4 and the network is further reduced at pruning level 2 when nodes B and H are removed. What remains in the network are the linear paths containing metabolites with in- and out-degree one (for a visual representation of the pruning process applied to the entire metabolic network please see fig. S3).

2.2 Defining backbone of pathways in EcoCyc

The manually curated pathways in the EcoCyc data set are defined in terms of reactions, and in order to analyse how these pathways are affected by pruning, and also compare our pruned paths with the pathways from the database, we need to extract the metabolites each pathway contains. Simply taking all metabolites connected to the reactions defining each pathway gives

a very broad definition of pathway, and as we are more concerned with capturing all reaction events along the pathway a reduction method was employed. To do this reduction for a given pathway we start with its complete list of metabolites. In order of degree we start removing metabolites from the pathway and accept the removal if the two following holds: (i) each reaction in the pathway has at least one in- and out-link and (ii) the pathway is connected (i.e. still consists of one connected component). As an example, if we apply the above algorithm to the two pathways defined by reactions (1,2) and (3,4,5) in fig. 2, the net result will be that the leftmost pathway will consist of metabolites (A,C,F) and the other one of (B,E,G,H).

2.3 Randomisation of network

Throughout the paper we compare the obtained results with random paths in the full network and the randomised version of the full network, which represent the expectation from a null model where only degree of nodes is preserved. The randomised version of the network is generated by degree-preserving rewiring of the original network (Maslov and Sneppen, 2002), here modified to preserve both in and out degrees of both reaction nodes and metabolites. In particular this implies that the scale-free degree distribution of the metabolites is preserved. Other requirements such as conservation of atomic balances could also be included in the randomisation procedure. This is, however, not considered here.

3 RESULTS

The pruning of the network was carried out by removing metabolite nodes in order of degree, starting with the nodes with the highest degree and removing several nodes in one step if they have the same degree. Inspired by the concept of pathways, we first consider the distribution of path lengths in the network. This was done by starting at a random metabolite node and in each node along the path follow one of the outgoing links at random (all the paths in the network could equally well be analysed exhaustively). This process was repeated until a dead-end (metabolite node with out-degree zero) or a previously visited node was reached, and the number of metabolite nodes visited in the path was recorded. This process was repeated 6×10^5 times for both the real network and a randomly rewired version of it, generated as described in methods section. Fig. 3 shows the thereby obtained path length distribution for the original network together with a pruned network at pruning level 3. For both pruning levels we also compare with path lengths obtained in their random counterparts. For the full network we see that the randomised version in general exhibits longer paths than the real network. In contrast, for pruning level 3 the real network exhibits longer paths. Accordingly, the paths detected at pruning level 3 represent significant topological structures which may have defined functional roles. This in accordance with the concept of metabolic pathways, lends support to the notion of metabolic networks as an assembly of directed paths carrying out specific catabolic and anabolic functions. Metabolic networks are “stringy”, in particular when viewed at low pruning levels.

The pruning of the network preserves paths which contribute significantly to the path length distribution. This begs the question if the paths revealed by the pruning of the network carry any other topological features or biological significance. We accordingly examine their topological features without any further biological information. Subsequently we will re-examine the found paths using all available biological information from annotated functional pathways and demonstrate that our simple pruning method in itself pinpoints known metabolic paths.

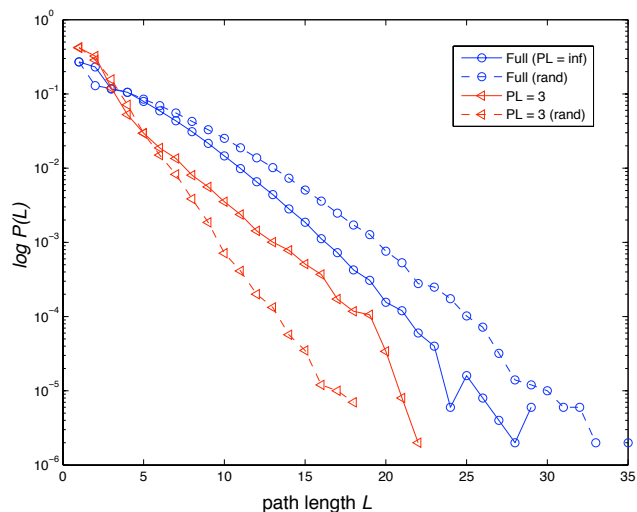


Fig. 3. The path length distribution for the full network (blue) and at pruning level 3 (red) plotted together with their random counterparts.

First we consider the topological features of the pathways highlighted by the pruning method. That is, we examine the neighbourhood of each of the found pathways, with respect to possible correlations between the metabolites which feed into the pathway. This neighbourhood of a given path is quantified in terms of the cost (Axelsen Bock et al., 2008), defined as the total number of different metabolites that contributes to the path, but are not produced by the path. That is the cost is defined as the number of in-links of all reaction nodes along the path minus the length of the path to compensate for the link being followed, see fig. 4a. If a metabolite enters a path as a substrate more than once it is only counted on its first occurrence. The cost of a path starting at node A and ending at B basically gives an estimate of how expensive in terms of needed cellular diversity it is to produce metabolite B from substrate A. This cost depends on the length of the path L . For example, the cost in the protein reaction network of *E. coli* grows approximately as $L/2$ (Axelsen Bock et al., 2008).

While investigated paths are identified from the pruned network it is important to stress that the cost is measured using the full metabolic network with all metabolites. The cost associated to these so called “pruned paths” (PP) was subsequently compared to the cost of randomly sampled paths (RP) from the full network. Fig. 4a show that the cost is significantly lower for the PP, suggesting that the PP have distinctively different structure than the RP sampled from the full network. If we calculate the cost differently and disregard the rule that each metabolite is counted only once the comparison between the selected and the random paths is quite different (see fig. 4b). The cost of the PP is in this case much closer to the random expectation. Therefore a major part of the reduction of metabolite diversity around a pruned path is associated to a systematic reuse of the same substrates.

We have established that the pruning process isolates low-cost paths in the network, but we do not know to what extent these paths resemble known metabolic pathways. In order to verify the

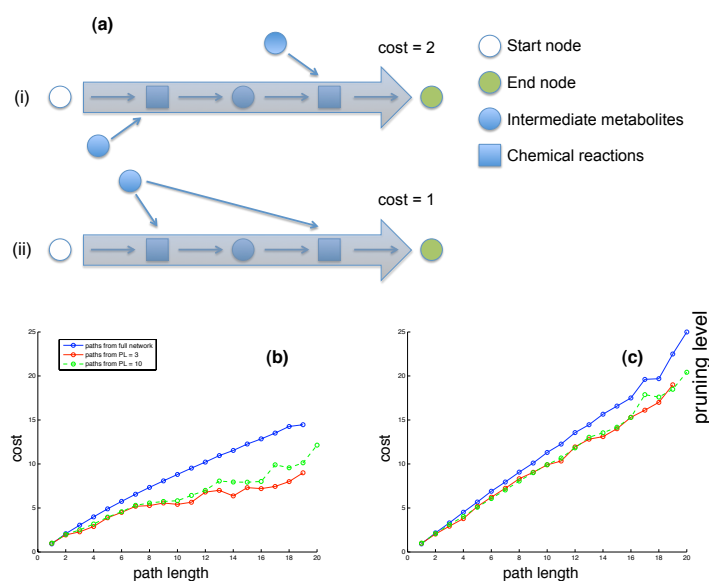


Fig. 4. The top panel (a) shows a schematic image of the cost measure between distant nodes in the network used to produce the graphs below. In (i) two different metabolites are required along the pathway and gives therefore a cost of 2, while in (ii) the same intermediate metabolite is used in several reactions resulting in a cost of unity. Panel (b) and (c) show the result of these two cost measures for paths found in the pruned network (red and green) and in the original network (blue). In (b) the repeated contribution of substrates is ignored, while in (c) every substrate along the path is included in the cost function.

correspondence we made use of the pathway-data in the EcoCyc database, and investigated (i) how known pathways are influenced by the pruning process, and (ii) to what extent the found PP indeed overlap with well known pathways.

Every pathway in the data set is specified by a set of reactions. When considering how such a pathway would be influenced by the pruning method applied to the full network, we first prune the pathways to its backbone of lowly connected metabolites as described in the method section. By working with these backbone pathways (BP), we secure that we are indeed discussing known pathways in the same terms as we possibly could extract pathways by the pruning method. We subsequently analysed how these BP are reduced when the network is pruned by measuring the fraction of metabolites still present in the backbone-pathway as a function of the pruning level. The result of this can be seen in fig. 5 and shows that the backbone of most pathways remain unaffected until we reach very low pruning levels. In fact at pruning level 10 and 3 (for which we measured the cost) the fraction of metabolites remaining on average for each pathways is 80% and 60% respectively (55% and 20% of the pathways are intact).

A direct visual comparison between known pathways from the database, and the pruned network is seen in fig. 6a, where we have mapped out five pathways from the database onto the network at pruning level 3. The stringy structure of the pruned network indeed seems to correspond well to the pathways defined in the database. The extent to which the PP recapture known pathways can be measured systematically, by examining how well the PP's

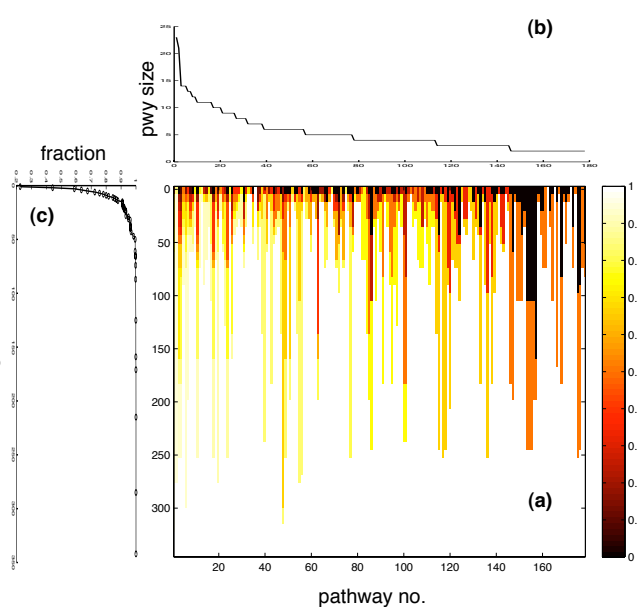


Fig. 5. The effect of the pruning algorithm on metabolic pathways as defined in the EcoCyc database. In (a) we have plotted the fraction of metabolites remaining in the pathway as a function of the pruning level. As shown in (b) the pathways are ordered with the largest one (containing the most metabolites) to the left and then in decreasing size. If we take the average over all pathways we see in (c) that the pathways are on average intact until a pruning level of approximately 50.

follow a single pathway in the database. For this investigation we define known pathways in terms of their reactions, and simply count (i) the number of different pathways that a given PP has an overlap with, and (ii) the number of reactions that the PP has in common with the pathway with which it has the largest overlap. Both of these measures naturally depend on the length of the walk that defines the PP. Figure 6b and c show the measures of pathway diversity and maximal overlap as function of the length of the PP. The PP's was extracted on pruning level 3 and, for comparison we also compare with random paths on the full metabolic network. As we can see random paths in the pruned network encounter fewer pathways (i.e. are more specific) and have a large overlap with an existing pathway. In contrast the paths that was generated by random walks on the full networks often use the currency metabolites to jump between known pathways, and therefore have little overlap with any of the known pathways that it visits. Overall, we conclude that paths extracted from pruned networks indeed isolates functionally important pathways of the network.

4 DISCUSSION

The pruning algorithm described in this paper was introduced in order to systematically investigate the importance of currency metabolites in metabolic networks. By subsequently removing nodes of high degree from the network, the metabolic network of *E. coli* reveals an underlying “stringy” topology. This understanding of metabolic network topology as an assembly of pathways is visible

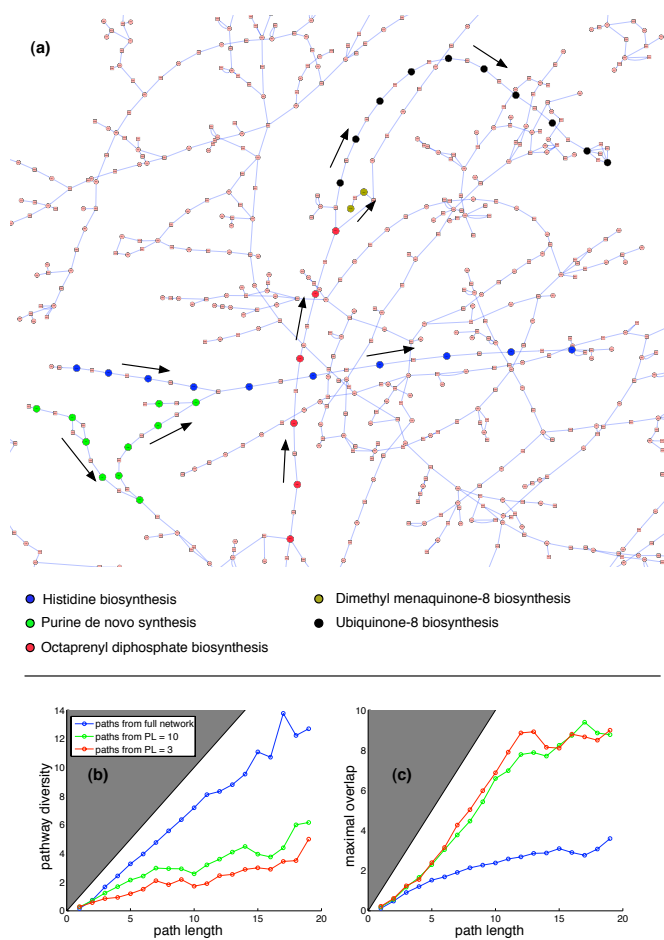


Fig. 6. Correspondence between paths in the pruned network and pathways identified from the EcoCyc database. In (a) a subset of the metabolic network at pruning level 3 is shown with a few pathways identified from the EcoCyc database extracted using our algorithm, marked out. The tendency for the stringy paths to overlap with real pathways is quantified in the two lower plots; (b) shows the number of EcoCyc pathways encountered for a random trajectory in the network as a function of the path length and (c) shows the size of the longest overlap with any pathway encountered by the random trajectory. The shaded areas indicate the upper bound of the two measures; in (b) a path of length L can at most encounter L different pathways, and in (c) a path of length L can at most have a maximal overlap equal to its own length.

in two topological features. First, we observed a cross-over in path length distribution for different pruning levels. The full network exhibits statistically shorter paths than its random counterpart, while at pruning level 3 the real network contains longer paths than its randomised counterpart. Accordingly there are on average longer paths in the metabolic network than expected from a standard null model. Second, the pathways found in the pruned network retained their specificity in the full non-pruned network, in the sense that they had lower “cost” than random pathways in the full network. Thus the pathways detected by pruning are indeed distinct in the sense that they can function and transfer biomass with relatively small

requirement on diversity of metabolites to fuel reactions along the path.

In order to assess if these results are specific to metabolic networks or if they also hold for the broader class of directed scale-free networks, we analysed the transcription networks of *E. coli* (Mangan and Alon, 2003) and yeast (Costanzo et al., 2001). Firstly the GC of these networks fragmented at pruning level of approximately 11 (as opposed to 5 for the metabolic network, see fig. S5), which suggests that the transcription networks are held together by nodes with higher connectivity, and not by string-like pathways as in the metabolic network. Further, the path length distribution was unchanged or even increased when the pruned transcription networks were randomised (see fig. S6), which is the opposite to the situation in the metabolic network. These facts suggest that these networks do not exhibit the same “stringy” topology as the metabolic network of *E. coli*.

Fig. 6 shows that the long string-like paths revealed by the pruning process correspond well to the known functional pathways listed in the EcoCyc-database. The pruning process therefore identifies biologically relevant subsets of the network, from a pure network topological perspective. We do not use any knowledge about the nature of particular metabolites, and do not employ more elaborate computational methods, such as flux balance analysis (Fell and Small, 1986; Förster et al., 2003). Naturally one needs to choose a reasonable level of pruning, otherwise pathways which contain metabolites with a high degree will be missed, but at least the pathways defined in the EcoCyc database seems fairly robust to the pruning process. The backbone of most pathways are unaffected until a pruning level of 50, and even at such low pruning levels as 3 the remaining fraction of metabolites in the pathways is approximately 60%. The pruning method also conserves less linear pathways such as CoA-synthesis pathway (lost at $PL = 2$) and the Pentose Phosphate pathway (first disrupted at $PL = 10$).

A similar technique to ours was applied by Croes et al. (2006) for identifying meaningful pathways in metabolic networks. They used a weighted metabolic network, where the weights represented the number of reactions each metabolite takes part in, and searched for paths with the lowest total weight. This simple technique was successful at distinguishing real pathways, and also highlights the importance of disregarding high connectivity metabolites when analysing these networks. A more elaborate approach for recovering metabolic pathways was used by Beasley and Planes (2007) where they also included a number of optimisation constraints related to biochemical requirements. These kind of approaches suffer from the combinatorial increase in the number of possible paths when considering large networks. This is avoided by our method since pruning heavily reduces the number of possible paths. On the other hand, unrealistic pathways may be identified, such as those in which no mass is transferred (de Figueiredo et al., 2009).

Our analysis suggests a grading of metabolites which softens the distinction between ordinary and currency metabolites. The high connectivity metabolites, such as water, ATP etc., facilitate the reactions that occur in the network, but do not define or specify the individual pathways. However lower ranking metabolites are also quite non-specific in regards to specifying pathways. This is indicated by the fact that one can remove all metabolites with a degree higher than 10 with a reduction of only 20 % in the backbones of known pathways (see fig. 5). Overall our analysis of metabolic networks suggest a reinterpretation of its scale-free

degree distribution into a graded specificity of involved metabolites. The backbone of the network is composed of metabolites that take part in relatively few reactions, and thereby define functional pathways. Of course these high degree metabolites are indispensable for the function of the overall metabolism, but the fact that they usually exist in high abundance means that they rather serve as a "soup" in which the actual low degree network is immersed.

The differentiated roles of high and low degree nodes in metabolism points to a difference between networks representing biochemical reactions, and networks representing information transmission such as the WWW (Broder et al., 2000) or the Internet (Siganos et al., 2003). In digital networks, the ability of individual hubs to differentiate between all their neighbours in the network open for a more complex role of these hubs than in the here studied metabolic case. In molecular networks, a given molecule only have a small repertoire of states, and therefore tend to treat all neighbours in similar way. In particular, in metabolic networks a metabolite can only be present or absent, and accordingly it convey the same coherent "message" to all reactions it participate in. Hubs in metabolic networks should accordingly be understood in a different light than hubs in networks where nodes have large intrinsic ability to differentiate. In metabolic networks the specificity is concentrated to the nodes with low degrees, a feature that is revealed in the network topology by the ability of low degree nodes to define the main pathways in the network.

5 CONCLUSION

In this paper we have investigated how the *E. coli* metabolic network responds to the systematic removal of metabolites in decreasing order of connectivity. The purpose of this method is two-fold; it serves both as a means of analysing the structure of the network, but also provides a way of identifying biologically relevant subsets of the network. The main conclusion of our study is that the high degree metabolites in the network serve as a background which facilitates the function of the biologically relevant pathways. In contrast to for example transcription networks these hubs can be removed without breaking up the network, and the removal instead reveals hidden structures.

This provides an alternative way of viewing metabolic pathways which possibly could lead to an increased understanding of their function and evolution. It could also be used to study other aspects of metabolism such as the interplay between transcription factors and metabolic regulation, or to compare the metabolic organisation of different organisms such as obligate parasites and free-living organisms.

Funding: This work was funded by the Danish National Research Foundation through the Center for Models of Life (CMOL), as well as by the Willum Kann Rasmussen foundation through BIONET.

REFERENCES

- Albert, R., Jeong, H., Barabasi, A., 2000. Error and attack tolerance of complex networks. *Nature* 406, 378–382.
- Axelsen Bock, J., Krishna, S., Sneppen, K., 2008. The cost and capacity of signaling in the *escherichia coli* protein reaction networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008 (01), P01018–.
- Beasley, J. E., Planes, F. J., 2007. Recovering metabolic pathways via optimization. *Bioinformatics* 23, 92–98.
- Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., Wiener, J., 2000. Graph structure in the web. *Comput. Netw.* 33 (1-6), 309–320.
- Callaway, D. S., Newman, M. E. J., Strogatz, S. H., Watts, D. J., 2000. Network robustness and fragility: Percolation on random graphs. *Phys. Rev. Lett.* 85 (25), 5468–5471.
- Costanzo, M. C., Crawford, M. E., Hirschman, J. E., Kranz, J. E., Olsen, P., Robertson, L. S., Skrzypek, M. S., Braun, B. R., Hopkins, K. L., Kondu, P., Lengieza, C., Lew-Smith, J. E., Tillberg, M., Garrels, J. I., 2001. Ypdtm, pombeptdm and wormptdm: model organism volumes of the bioknowledgetm library, an integrated resource for protein information. *Nucl. Acids Res.* 29, 75–79.
- Croes, D., Couche, F., Wodak, S., van Helden, J., 2006. Inferring meaningful pathways in weighted metabolic networks. *J Mol Biol* 356 (1), 222–236.
- Dandekar, T., Moldenhauer, F., Bulik, S., Bertram, H., Schuster, S., 2003. A method for classifying metabolites in topological pathway analyses based on minimization of pathway number. *Biosystems* 70 (3), 255–270.
- de Figueiredo, L. F., Schuster, S., Kaleta, C., Fell, D. A., 2009. Can sugars be produced from fatty acids? a test case for pathway analysis tools. *Bioinformatics* 25, 152–158.
- Duarte, N. C., Becker, S. A., Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., Srivas, R., Palsson, B. ., 2007. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences* 104 (6), 1777–1782.
- Fell, D., Small, J., 1986. Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochem. J.* 238 (3), 781–786.
- Fell, D. A., Wagner, A., 2000. The small world of metabolism. *Nat Biotech* 18 (11), 1121–1122.
- Förster, J., Famili, I., Fu, P., Palsson, B. O., Nielsen, J., 2003. Genome-scale reconstruction of the *saccharomyces cerevisiae* metabolic networks. *Genome Res.* 13, 244–253.
- Guimera, R., Nunes Amaral, L. A., 2005. Functional cartography of complex metabolic networks. *Nature* 433 (7028), 895–900.
- Herrgard, M. J., Swainston, N., Dobson, P., Dunn, W. B., Arga, K. Y., Arvas, M., Buthgen, N., Borger, S., Costenoble, R., Heinemann, M., Hucka, M., Le Novere, N., Li, P., Liebermeister, W., Mo, M. L., Oliveira, A. P., Petranovic, D., Pettifer, S., Simeonidis, E., Smallbone, K., Spasie, I., Weichart, D., Brent, R., Broomhead, D. S., Westerhoff, H. V., Kurdar, B., Penttila, M., Klipp, E., Palsson, B. O., Sauer, U., Oliver, S. G., Mendes, P., Nielsen, J., Kell, D. B., 2008. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotech* 26 (10), 1155–1160.
- Holme, P., Huss, M., 2007. Currency and commodity metabolites: Their identification and relation to the modularity of metabolic networks. *IET Systems Biology* 1, 280–285.
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z., Barabasi, A.-L., 2000. The large-scale organization of metabolic networks. *Nature* 407, 651–654.

- Karp, P., 2007. Multidimensional annotation of the *escherichia coli* k-12 genome. *Nucl. Acids Res.* 35 (22), 7577–7590.
- Ma, H.-W., Zeng, A.-P., 2003. The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* 19 (11), 1423–1430.
- Ma, H.-W., Zhao, X.-M., Yuan, Y.-J., Zeng, A.-P., 2004. Decomposition of metabolic network into functional modules based on the global connectivity structure of reaction graph. *Bioinformatics* 20 (12), 1870–1876.
- Mangan, S., Alon, U., 2003. Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences of the United States of America* 100, 11980–11985.
- Maslov, S., Sneppen, K., 2002. Specificity and stability in topology of protein networks. *Science* 296, 910–913.
- Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., Barabási, A. L., 2002. Hierarchical organization of modularity in metabolic networks. *Science* 297, 1551–1555.
- Schuster, S., Heinrich, R., 1996. *The Regulation of Cellular Systems*. Springer Verlag.
- Schuster, S., Pfeiffer, T., Moldenhauer, F. Koch, I., Dandekar, T., 2002. Exploring the pathway structure of metabolism: decomposition into subnetworks and application to *mycoplasma pneumoniae*. *Bioinformatics* 18 (2), 351–361.
- Siganos, G., Faloutsos, M., Faloutsos, P., Faloutsos, C., 2003. Powerlaws and the AS-level internet topology. *ACM/IEEE Transactions on Networking* 11, 514–524.