A Cross Species Comparison of Metabolic Network Functions

Oliver Ebenhöh oliver.ebenhoeh@rz.hu-berlin.de
Thomas Handorf handorf@physik.hu-berlin.de
Reinhart Heinrich reinhart.heinrich@rz.hu-berlin.de
Theoretical Biophysics, Institute of Biology, Humboldt University Berlin, Germany

Abstract

We compare a large number of organisms with respect to their metabolic network functions. We measure such functions in terms of the synthesizing capacity of a network when it is provided with a few small chemical substances as external resources. We call this measure the scope and show that it is generally robust against structural alterations of the reaction network. Organisms can be separated into two groups, one with a small and one with a large scope. Networks with a high synthesizing capacity also show a high degree of robustness against structural changes, indicating that this network function has been a target in the evolutionary past of the corresponding organisms. A comparison between structural and functional similarities reveals that organisms with a similar structure do not necessarily show similar biological functions. The presented concepts allow for a systematic investigation of structure-function relationships of metabolic networks and may put forth valuable hints on the evolution of metabolic pathways.

Keywords: metabolic networks, phylogenetic analysis, scope, KEGG

1 Introduction

Whereas traditionally mathematical modelling of metabolic networks aims at simulating stationary states or time courses for the system variables with the system parameters used as input quantities, recently structural analyses gained much interest since in contrast to enzyme kinetic parameters the stoichiometric properties are often much better known [9]. The development of metabolic databases such as KEGG [8] allows for studying the structure of metabolism on a large scale and to compare the structures of the metabolic networks of an increasing number of organisms.

In previous work we performed structural analyses on subnetworks by applying evolutionary optimization using a limited set of generic reactions and could show that the stoichiometry of the central energy metabolism can be explained under the premise that contemporary networks ensure a high ATP production rate [4, 10]. We introduced the method of network expansion [3, 6] that allows the reconstruction of large scale metabolic networks. From this method arises the concept of the scopes of compounds. The scope defines the synthesizing capacity of a network when it is provided with a certain set of substrates and therefore provides a characterization of a biological function of a network. By linking the concept of scopes with our recently introduced concept of stoichiometric robustness characterizing the effect of structural changes on the function of metabolic networks [5], we are able to systematically analyse functional properties of many organism specific networks and how robust these functions are against changes of the network structures.

The goal of the present work is to assess similarities and differences of structural and functional properties of the metabolic networks of different organisms. We first analyse the combined metabolic network containing all reactions which occur in at least one organism. Functional properties are

characterized by calculating the scopes of many different compounds. We analyse the robustness of these functions against structural changes of the network. Secondly, we apply these methods to investigate the organism specific metabolic networks of 178 species. For all these organisms we identify a particular biological function, the synthesizing capacity of the metabolic network when a small number of simple chemical compounds is available in the environment. We further define a quantity representing the robustness of a network's function against gene deletions and compare the corresponding values for all organisms. Subsequently, these results are used to perform a comparison between differences in the network structures and dissimilarities in their functions.

2 Background

In the present work we analyse structural and functional properties of a large number of metabolic networks. For this purpose we apply the method of network expansion [3, 6]. This method is based on a process in which a series of metabolic networks increasing in size is constructed. The process depends on the specification of (a) the initial set of compounds (the seed) and (b) on the set of reactions (base set) which define the network that shall be analysed. The construction of the series of networks is performed by the following algorithm: 1. The process starts with one or more biochemical compounds acting as a seed of the expanding network. 2. Identification of those reactions from the base set which use as substrates only compounds already present in the current network. 3. Incorporation of the identified reactions and their products into the network. This results in the next generation of expanding networks. 4. Repetition of steps 2 and 3 until no further reactions can be identified for incorporation.

Upon completion of the process, the expanded network contains all compounds which can be synthesized from the seed using the reactions from the base set. This set of compounds is called the scope of the seed and the cardinality of the set defines the scope size. The scope describes the synthesizing capacity of a metabolic network which is provided with the seed compounds as external resource. Since not all metabolites can be produced from arbitrary seed compounds, the expansion process will in general not lead to a network containing all reactions from the base set.

Metabolic networks contain reactions which are almost irreversible and therefore predominantly take place only in one direction. For our calculations we assume that all reactions can take place both in forward and backward direction. We refrained from distinguishing between reversible and irreversible reactions, since the actual rates for the forward and backward directions depend on the physiological conditions, in particular on the concentrations of the initial substrates and end products.

We have retrieved all biochemical reactions from the KEGG database. These 5311 reactions connect 4587 compounds forming a combined network containing as subnetworks the specific metabolisms of 178 organisms. In section 3 we present results obtained from analysing the combined network. Section 4 focusses on the investigation of organism specific networks and in section 5 we present approaches how to analyse the relationship between structural and functional properties of these networks.

In all calculations of scopes we have assumed that water is always abundant in the environment and is available as a substrate for every reaction which is identified in step 2 and incorporated into the expanding network in step 3 of the algorithm.

3 Analysis of the Combined Network

3.1 Scopes of Compounds

For the combined network we have considered all seeds consisting of only one compound. These 4587 seeds result in 3345 different scopes which implies that there exist different seed compounds which lead to identical scopes. Since the scope comprises all compounds that can be synthesized from the

seed, two compounds resulting in identical scopes can be synthesized from one another, thus they are chemically interconvertible.

Figure 1 shows a histogram for the number of distinct scopes of a given size. It can be seen that the distribution of scope sizes is very non-uniform with the smallest scope size being 5 and the largest 2101. Scopes exist for all sizes between 5 and 32 while there are only very few scope sizes larger than 1500. The smallest scope of size 5 results from using water as the seed compound. The largest scope with size 2101 can be reached from four different seed compounds, adenosine 5'-phosphosulfate (APS), 3'-phosphoadenosine 5'-phosphosulfate (PAPS), dephospho-CoA, and UDP-6-sulfoquinovose.

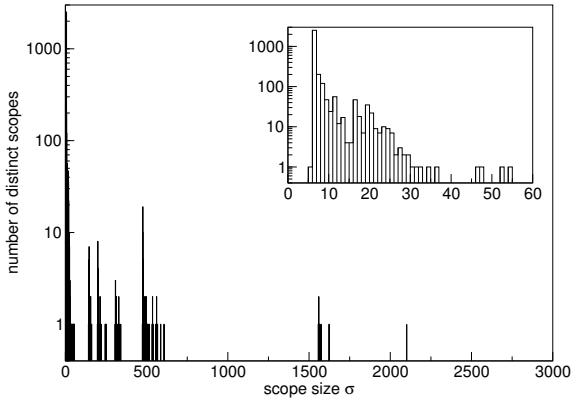


Figure 1: Histogram of scope sizes. It is shown how many distinct scopes have a certain size. The inset is a magnification for small scope sizes. All calculations include water as a seed compound. The smallest scope of size 5 corresponds to the scope of water (H_2O, O_2, O_2^-, H^+, H_2O_2).

Of special interest is the scope of size 1557. This scope can be reached from 103 different seed compounds, meaning that all these compounds are interconvertible. This group of compounds includes central cofactors such as ATP and GTP as well as the corresponding mono- and diphosphates and the nicotinamide dinucleotides NADH and NADPH. Further investigation reveals that this scope is a subset of the largest scope. This implies, for example, that ATP can be synthesized from APS while the opposite is not possible. This becomes apparent by considering the chemical composition of these two compounds: APS contains a sulfate group while ATP does not.

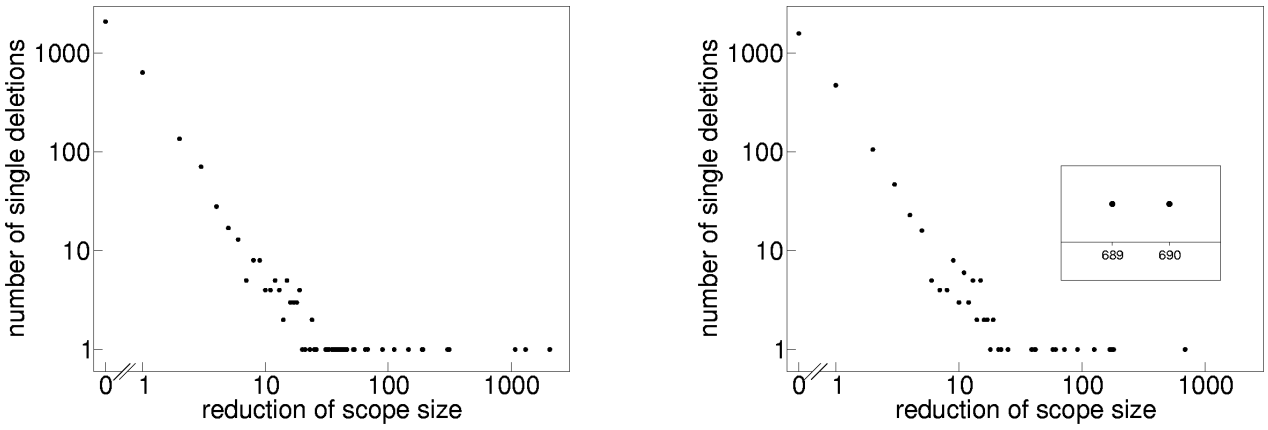
It is an intriguing question whether a network of size similar to the scope of APS can be obtained when starting the expansion process from a seed containing a small number of simple compounds which are present in the environment. Guided by the elements contained in APS, we select the following set of seed compounds: CO_2, NH_3, H_3PO_4 and H_2SO_4. Interestingly, the set of compounds which can be synthesized from these simple compounds is exactly the same as the set of compounds which can be produced from APS. We have tested whether the same scope can be reached when the carbon source CO_2 is replaced by CH_4. Surprisingly, the expansion results in a scope containing only 25, predominantly inorganic, compounds. A closer inspection of the expansion process shows that all reactions utilising methane require the presence of cofactors. However, these cofactors are rather complex molecules which have not been included in the expanding network at the stage in which they are needed and therefore the expansion process stops in an early stage explaining the small scope size.

3.2 Robustness of Scopes

Scopes represent functional modules in the sense that they comprise all compounds which can be synthesized from the seed substrates. Which compounds are contained in the scope depends on the set of available reactions (base set). In extension to our previous work [5], we analyse in the following the robustness of network functions against structural modifications. In particular, we investigate how the scopes are affected if the base set of reactions is changed. Such modifications can be caused by

deletions of genes coding for metabolic enzymes, resulting in a loss of the corresponding biochemical conversions. Stronger modifications result when considering specific metabolic networks of different organisms since each organism has its own base set of reactions which is a subset of the reactions found in the combined network. The organism specific networks, their functions and the robustness of these functions will be studied in the next two sections while in this section the robustness of the combined network is considered.

It is of particular interest how robust the results obtained in section 3.1 are against changes in this set of available reactions. Moreover, such analyses are of relevance since the KEGG database is not complete. A robustness of scopes against structural variations of the metabolic network also implies a robustness of this concept against possible modifications and extensions of the KEGG database which may be initiated for example by the availability of new and improved annotations of genomes.



(a) (b)

Figure 2: Robustness of scopes. Plotted is the number of single deletions resulting in a given reduction of the scope sizes of the set CO_2, NH_3, H_3PO_4, H_2SO_4 (a), and ATP (b).

We have calculated the effects of all single deletions from the base set of reactions on the scope sizes. In Figure 2, the resulting effects are depicted for the two scopes which result from using CO_2, NH_3, H_3PO_4 and H_2SO_4 as seed compounds (Figure 2a), and using ATP as the only seed compound (Figure 2b). Each plot shows in how many cases a single deletion reduces the scope size by a given number of compounds. The diagram reveals that in both cases the majority of such deletions does not affect the scope size at all. Most of the other deletions have only a small effect on the scope size. However, there are a few reactions whose deletion from the base set significantly reduce the scope sizes. We have analysed a large number of scopes of very different sizes and it turned out that this is a general property. In all cases there are many reactions whose removal have no or only minor effects on the scope size while the number of reactions whose deletions lead to a strong decrease of the scope size is small.

Since scopes characterize a synthesizing capacity of a network, those reactions whose deletions result in a drastic reduction of scope size are of vital importance for the functionality of the network. We are thus able to identify central reactions whose occurrences are essential for the network to perform certain biological functions. A detailed analysis reveals that the reaction whose removal has the strongest impact on the scope of the seed CO_2, NH_3, H_3PO_4 and H_2SO_4 (Figure 2a) is the reaction $CO_2 + H_2O_2 \rightleftharpoons \text{Oxalate} + O_2$ and its elimination results in a scope containing only 32 compounds corresponding to a reduction by 2069 compounds. In the case of the scope of ATP (Figure 2b), there are two reactions whose removal reduce the scope size by 689 and 690 compounds, respectively.

To analyse potential cumulative effects of missing reactions, we examine how the simultaneous

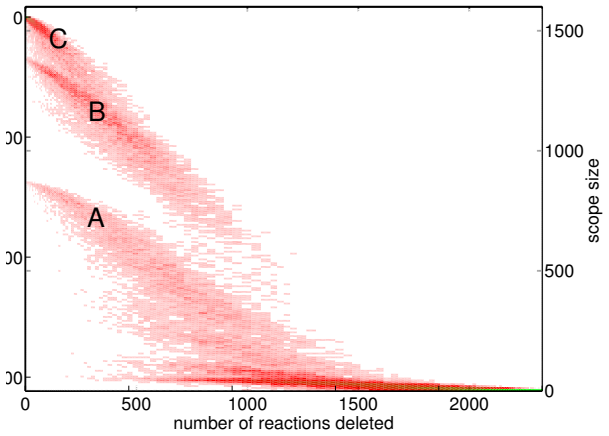


Figure 3: The effect of multiple deletions on the scope size of ATP. The shading indicates the probability that the scope has a certain size (y-axis) when randomly deleting a certain number of reactions (x-axis). For any chosen number of deleted reactions, 200 random samples were calculated.

removal of more than one reaction affects the scope sizes. Figure 3 shows how the size of the ATP scope decreases with increasing number of removed reactions. Specifically, a two-dimensional distribution is shown with the shading indicating the probability that a reduction of a given number of reactions results in a certain scope size.

There exist distinct domains in which these probabilities are high. This results from the fact that there are a few reactions which alone have a strong impact on the scope size. For example, domain A contains cases in which one of the two reactions reducing the scope size in single deletions by 689 or 690 compounds, has been removed. Similarly, domain B contains those cases in which at least one of a group of reactions has been removed, which reduce the scope size by about 190 (cf. Figure 2b). There exists a third domain (C) which contains those cases in which only reactions have been removed which do not have a strong effect on the scope size when considering single deletions. The existence of seperate domains and the fact that within each domain the scope sizes decrease uniformly with increased number of reduced reactions indicates that the scope size is critically influenced by only a small number of reactions, whereas it is robust against elimination of most reactions. The latter reactions influence the scope size mainly by reducing the number of available reactions.

4 Application to Organism Specific Metabolisms

4.1 Scopes of Compounds

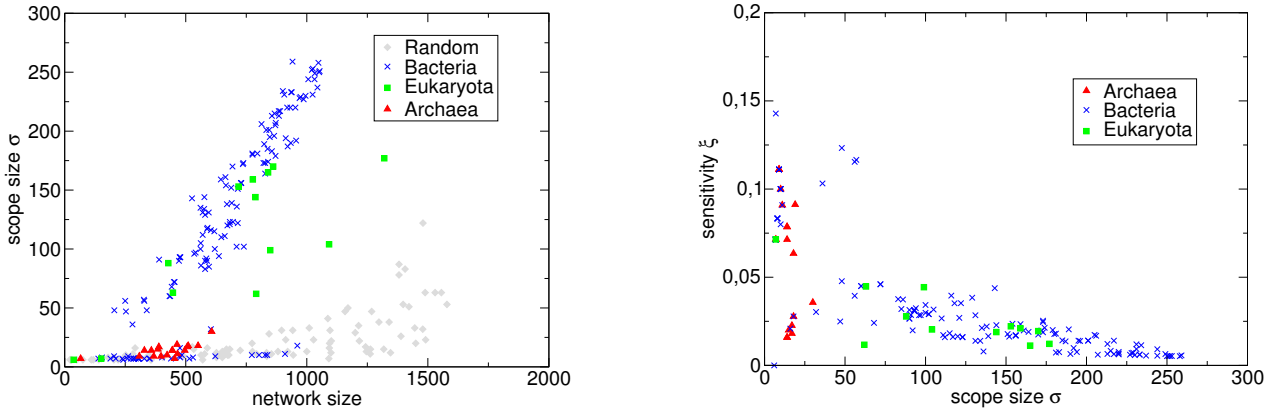
The concept of the scopes provides a suitable method to analyse functional properties of metabolic networks. We are interersted in the synthesizing capacities of different organisms when supplied with certain resources.

In the discussion on the scope of the seed of several small compounds calculated for the combined networks in section 3.1 it became apparent that cofactors play an important role in the expansion process. Cofactors generally mediate special chemical transformations without actually being consumed. In particular, the pairs NAD^+/NADH and $NADP^+$/NADPH facilitate redox reactions by accepting/donating a pair of electrons. Similarly, the pair ATP/ADP is often involved in the transfer of a phosphate group from one compound to another and coenzyme A participates in the transfer of acyl groups. Since under normal conditions cofactors are already present in cells, they do not have to be especially snythesised but rather are available to the corresponding biochemical reactions. To assess functions of organism specific metabolic networks it is therefore realistic to assume that

the functionality of cofactors is always available. For the calculations in this section we modify the expansion process in the following way: during the identification of reactions in step 2 of the algorithm we consider the compounds NAD^+/NADH, $NADP^+$/NADPH, ATP/ADP and coenzyme A to be available as substrates if they appear in the corresponding reaction in their role as cofactors.

To analyse the synthesizing capacities of different organisms, we extract the organism specific networks from the KEGG database for 178 species and calculate for each network the scope of the seed CO_2, NH_3, phosphate, sulfate and glucose. Figure 4a shows the sizes of the scopes depending on the sizes of the networks where each data point represents a distinct organism. We indicate organisms from the three main groups eukaryotes, bacteria, and archaea by squares, crosses and triangles, respectively.

The organisms can clearly be divided into two clusters. The organisms within the cluster lying near the horizontal axis are characterized by very small scope sizes whereas those of the other cluster possess a significantly larger scope. Interestingly, all archaea are located in the lower cluster while all eukaryotes are found in the upper cluster. Bacteria, on the other hand, are present in both clusters.



(a) (b)

Figure 4: (a) Scope sizes for the seed CO_2, NH_3, phosphate, sulfate and glucose for all organism specific networks as a function of network size; (b) Sensitivities of organism specific networks depending on the scope size. Each data point represents one organism where eukayotes, bacteria and archaea are represented by squares, crosses and triangles, respectively. The diamonds in (a) represent scope sizes for random networks of various sizes.

These findings raise the question which characteristic features of the various metabolic networks are responsible for this clustering. In order to analyse this problem, we compare the networks with randomly generated networks of similar size. We constructed such networks by randomly selecting a certain number of reactions from the combined network which contains 5311 reactions. The scopes of these random networks are depicted by diamonds in Figure 4a. Their scope sizes are generally small and, surprisingly, the random networks show a very similar behaviour as the networks of the lower cluster. The small scope size of the random networks is understandable since it cannot be expected that a random selection of reactions from the combined network will result in a network which can make use of the given seed compounds. As a consequence, they should be distinguishable from real metabolic networks which are the products of the evolutionary development of the corresponding organisms. The fact that the networks within the lower cluster have a comparable scope size to random networks suggests that they are not functional regarding the selected set of seed compounds. This indicates that during the evolution of these organisms the synthesizing capacity regarding the seed compounds CO_2, NH_3, phosphate, sulfate and glucose was not an important selection criterion.

4.2 Robustness

Inspired by the analysis of robustness of scopes against structural changes of the underlying network (section 3.2), we investigate how robust the synthesizing capacities of different species are against structural modifications of their metabolic networks. In analogy to Figure 2, we calculate the effect of single reaction deletions on the scope of the seed CO_2, NH_3, phosphate, sulfate and glucose for the organism specific networks. To characterize the robustness of a metabolic network, we introduce the sensitivity ξ of a network, a value describing by what fraction the scope is reduced in average when deleting one reaction. The sensitivity is defined by

$$\xi = \frac{1}{\omega\sigma} \sum_r r\rho(r), \tag{1}$$

where σ denotes the original scope size and ω the number of reactions of the corresponding expanded network, and $\rho(r)$ is the number of reactions whose deletion results in a reduction of the scope size of r. It is evident that $0 \leq \xi \leq 1$ where $\xi = 0$ characterizes a completely insensitive, i. e. a completely robust network, for which all single reaction deletions will have no effect on the scope size, and $\xi = 1$ describes a maximally sensitive network, for which the removal of any reaction will result in the disappearance of the scope.

Figure 4b depicts the sensitivities depending on the scope size σ. Clearly, the robustness against deletions of reactions decreases with increasing scope size. In most cases, a removal of a reaction will in average result in a reduction of the scope size by less than 5%. There are a few organisms with a higher sensitivity. All these organisms, however, are characterized by a small scope size meaning that their synthesizing capacity regarding the compounds CO_2, NH_3, phosphate, sulfate and glucose is small.

Generally, the sensitivities of bacteria and eukaryotes show a similar behaviour. Archaea are characterized by a small scope size (see also Figure 4a) but show a large variety in their sensitivities. It can be concluded that for those organisms which are characterized by a large scope, meaning that their synthesizing capacity is high, this function is very robust against structural changes of their metabolic networks.

5 Similarity Analysis of Metabolic Networks

In the previous section it was shown that the sizes of the scope of the seed CO_2, NH_3, phosphate, sulfate and glucose varies significantly among different species. In order to analyse which structural properties of the metabolic networks are responsible for such vast differences in the scopes, we define two distance measures, one representing the structural dissimilarities of the metabolic networks of two species, the other measuring functional dissimilarities in terms of differences in their scopes.

5.1 Definition of the Distance Measures

We consider two different species X and X' with the metabolic networks N and N', respectively. These networks are defined by the biochemical reactions they contain. If we denote by $N \cup N'$ the network which contains all reactions appearing in at least one of the networks N or N', and analogously by $N \cap N'$ the network which contains only those reactions which appear in both networks N and N', then the function

$$d_S(X, X') = |N \cup N'| - |N \cap N'| \tag{2}$$

defines a distance between two species. Clearly, this distance is zero if and only if the networks of the two species are identical. The more reactions appear in only one of the two networks, the larger is this distance. Thus, the function d_S measures the structural dissimilarity of two metabolic networks. A related distance measure, the Jaccard coefficient, is obtained by calculating the quotient of the two

values $|N \cup N'|$ and $|N \cap N'|$ instead of their difference. It has, for example, been applied to extract phylogenetic modules of metabolic networks [11]. Closely related to our measure is the pathway content quantifying the similarity of metabolic pathways. Hong *et al.* [7] used this measure for the reconstruction of phylogenetic trees from the metabolic networks of different organisms. We choose Eq. (2) over the alternative measures, since it can be interpreted as an evolutionary distance between two networks with deletion and acquisition of enzymatic reactions as the fundamental evolutionary events.

To measure the distinctness of two metabolic networks in terms of a biological function, we define a second distance measure based on the concept of the scopes. Let S and S' denote the scopes of the seed compounds CO_2, NH_3, phosphate, sulfate and glucose resulting from the networks of species X and X', respectively. Since the scopes specify sets of metabolites, the function

$$d_F(X, X') = |S \cup S'| - |S \cap S'| \tag{3}$$

also defines a distance measure between two species. In analogy to Eq. (2), $S \cup S'$ denotes the set of metabolites which appear in at least one of the scopes S or S' and $S \cap S'$ is the set of compounds which are present in both scopes. Since the scopes describe a functional property of the underlying metabolic networks, the function d_F represents a measure of the functional dissimilarity of two species.

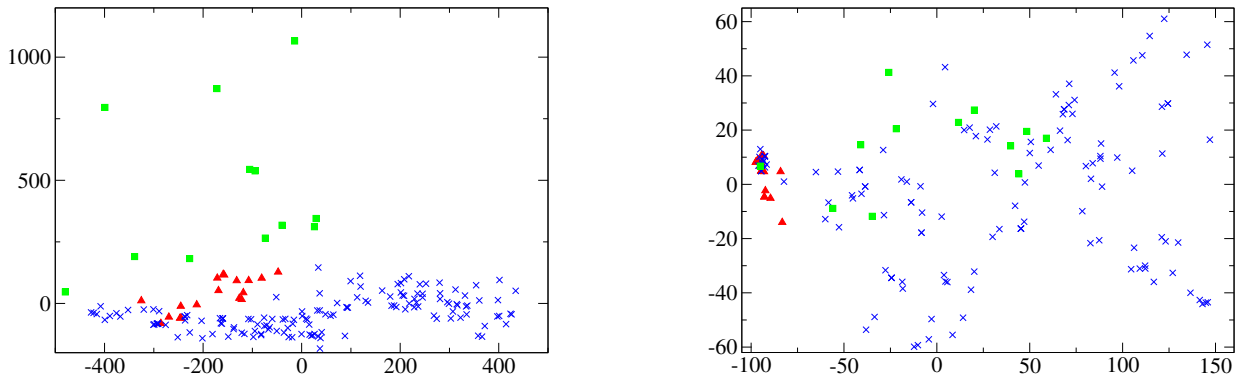
It is evident that if two species have identical metabolic networks, they also have identical scopes, thus

$$d_S(X, X') = 0 \quad \Rightarrow \quad d_F(X, X') = 0, \tag{4}$$

while the reverse implication is generally not true.

5.2 Structural and Functional Dissimilarities

Based on the metabolic networks of 178 species and the corresponding scopes of the seed compounds CO_2, NH_3, phosphate, sulfate and glucose, we calculated the distances d_S and d_F between all pairs of organisms. Figure 5a graphically represents the structural distances d_S between all these organisms.



(a) (b)

Figure 5: Structural (a) and functional (b) distances d_S and d_F, respectively, between 178 organisms. The plots have been generated using a classical two dimensional scaling algorithm. The numbers on the axes are arbitrary. The euclidean distance between two points in the plot resemble approximately the true distances d_S (a) and d_F (b). Eukayotes, bacteria and archaea are represented by squares, crosses and triangles, respectively. The metabolic networks for the 178 species have been retrieved from the KEGG database. The functional distances have been calculated based on the scopes of the seed compounds CO_2, NH_3, phosphate, sulfate and glucose.

The plot has been generated by a classical two dimensional scaling method [2]. This algorithm ensures that the distances as they appear in the graph are approximately equal to the true distances d_S.

The three groups archaea, bacteria and eukaryotes are surprisingly well distinguishable. All bacteria are located near the bottom of the graph. The extension of this group is rather large, allowing for the conclusion that there exists a large diversity in the metabolic network design among bacteria. The archaea are located close to each other, indicating that the structural variability in their metabolic networks is very low. Most of the archaea are located above the group of bacteria while there are a few which are found within the group of bacteria. Considering the lateral gene transfer which is assumed to take place not only among bacteria but also between bacteria and archaea [1], it is not surprising that some metabolic networks of archaea are not distinguishable from some networks of bacteria. Two strains of *Tropheryma whipplei*, Actinobacteria with a reduced genome, as well as *Pyrococcus abyssi* and *Pyrococcus horikoshii*, two closely related hyper-thermophilic archaea, are such examples. Most eukaryotes are clearly separable from the other two groups. They are located predominantly in the upper two thirds of the graph. Clearly, the distances among eukaryotes are generally larger than between bacteria and archaea, indicating large differences in the metabolic structures. There are three eukaryotes whose network structures seem to be closer to bacteria or archaea than to the other eukaryotes (located near the bottom left corner of the graph). A closer inspection shows that these three are unicellular eukaryotes, the two intracellular parasites *Plasmodium falciparum* and *Encephalitozoon cuniculi* as well as the red algae *Cyanidioschyzon merolae* which possesses a small, compact genome. All of those species are likely to have lost many enzymes during their evolutionary past, explaining why their metabolic networks are not clearly distinguishable from prokaryotes.

In Figure 5b, functional distances d_F between all considered organisms are represented. One cluster of organisms with a very small spatial extension located near the left edge of the graph can be identified. All archaea are included in this cluster, indicating that the metabolic function represented by the scopes is very conserved within this group of organisms. This behaviour can be explained since all archaea have been identified to possess a very small scope size (see Figure 4a). The other organisms not included in this cluster cover the largest part of the graph and a separation between bacteria and eukaryotes is not detectable. This is not surprising considering that the scope represents a functional property. Therefore one can expect that rather those organisms living in similar environments will be located close to one another than organisms with a similar network structure.

6 Discussion

We have applied the method of network expansion to characterize metabolic networks by scopes of compounds. These scopes define sets of metabolites which can be synthesized from seed compounds and therefore represent functional properties of the networks.

In the analysis of the combined metabolic network which contains all reactions which were identified in at least one organism, we found that exactly the same ensemble of metabolites can be synthesized from either exclusively using the rather complex substrate APS or from using the simple inorganic substances CO_2, NH_3, H_3PO_4 and H_2SO_4. We further showed that the scopes are very robust against structural changes of the underlying network and provided evidence that the scope size is critically influenced by only a small number of reactions.

We have compared the metabolic networks of 178 organisms with respect to one particular biological function, their synthesizing capacities from the compounds CO_2, NH_3, phosphate, sulfate and glucose. The results indicate that the organisms can be divided into two groups. Organisms of one group are characterized by a small scope size with the size being rather independent of the size of the network. The other group contains organisms with a high synthesizing capacity and within this group the scope size generally increases with increasing network size. We claim that for organisms appearing in the first group, a large synthesizing capacity from the selected seed compounds was not an important target during their evolutionary history. This view is supported by our analysis that the

corresponding scope sizes cannot be distinguished from those of randomly generated networks. Further, we see biological support in the fact that all archaea are located in this group. Archaea usually live in extreme environments like the deep sea and in hot springs. In such habitats, glucose will not be an abundant substrate and therefore it cannot be expected that such organisms have adopted a high synthesizing capacity which relies on glucose as external resource. The second group of organisms contains many eukaryotes, species characterized by a high variability in their network structure. We claim that the species of this group have undergone in their evolutionary past a selective pressure towards a high synthesizing capacity. If this claim is true, it can be conjectured that these organisms have, during their evolution, developed mechanisms to protect such an important function against random mutations. This can indeed be observed when analysing the robustness of this network function against structural alterations. In general, those species with a large scope show a more robust snythesising capacity than those with a small scope.

To provide a method to analyse in what respect changes in the structure of the metabolic network inflict changes in the biological functions, we have defined a structural and a functional distance measure and calculated the distances between all pairs of the 178 organisms. While with the structural measure the three domains of life are to some extent separable, this is not the case for the functional measure. Since biolgical functions of organisms can be viewed as the result of an evolutionary adaptation process, it can be expected that scopes, representing such functions, to some extent reflect the environmental conditions a species has experienced during its evolutionary past.

Our results lead to the following conjecture: Specialized organisms, especially those living in extreme habitats, will generally be characterized by a low synthesizing capacity while generalists, which are able to prosper in a wide variety of habitats, will show a higher capacity and different organisms living under similar environmental conditions will also possess similar metabolic functions. This hypothesis can in principle be tested by categorising the types of cellular environments, for example by the abundant chemicals and minerals, and comparing these categories with the synthesizing capacities of the corresponding species. We expect that such analyses will be part of future applications of our presented methods.

References

[1] Aravind, L., Tatusov, R.L., Wolf, Y.I., Walker, D.R., and Koonin, E.V., Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles, *Trends Genet.*, 14(11):442–444, 1998.

[2] Cox, T.F. and Cox, M.A.A., *Multidimensional Scaling*, Chapman and Hall, 1994.

[3] Ebenhöh, O., Handorf, T. and Heinrich, R., Structural Analysis of Expanding Metabolic Networks, *Genome Informatics*, 15(1):35–45, 2004.

[4] Ebenhöh, O. and Heinrich, R., Evolutionary optimization of metabolic pathways. Theoretical reconstruction of the stoichiometry of ATP and NADH producing systems, *Bull. Math. Biol.*, 63:21–55, 2001.

[5] Ebenhöh, O. and Heinrich, R., Stoichiometric Design of Metabolic Networks: Multifunctionality, Clusters, Optimization, Weak and Strong Robustness, *Bull. Math. Biol.*, 65:323–357, 2003.

[6] Handorf, T., Ebenhöh, O. and Heinrich, R., Expanding Metabolic Networks: Scopes of Compounds, Robustness and Evolution, *J. Mol. Evol.*, in press, 2005.

[7] Hong, S.H., Kim, T.Y., and Lee, S.Y., Phylogenetic analysis based on genome-scale metabolic pathway reaction content, *Appl. Microbiol. Biotechnol.*, 65:203–210, 2004.

[8] Kanehisa, M., Goto, S., Kawashima, S. and Nakaya, A. The KEGG databases at GenomeNet, *Nucleic Acids Res.*, 30(1):42–46, 2002.

[9] Schilling, C. H., Schuster, S., Palsson, B. O., and Heinrich, R., Metabolic pathway analysis: Basic concepts and scientific applications in the post-genomic era, *Biotechnol. Prog.*, 15:296–303, 1999.

[10] Stephani, A., Nuño, J. C. and Heinrich, R., Optimal stoichiometric design of ATP-producing systems as determined by an evolutionary algorithm, *J. Theor. Biol.*, 199:45–61, 1999.

[11] Yamada, T., Goto, S. and Kanehisa, M., Extraction of Phylogenetic Network Modules from Prokaryote Metabolic Pathways, *Genome Informatics*, 15(1):249–258, 2004.