

# Hierarchy of Metabolic Compounds Based on Their Synthesizing Capacity

Thomas Handorf<sup>1</sup>   Oliver Ebenhöf<sup>1</sup>   Daniel Kahn<sup>2</sup>   Reinhart Heinrich<sup>1</sup>  
handorf@physik.hu-berlin.de

<sup>1</sup> Theoreticla Biophysics, Institute of Biology, Humboldt University Berlin, Germany

<sup>2</sup> Institut National de la Recherche Agronomique, Toulouse, France

**Keywords:** metabolic networks, scope, scope hierarchy, structural analysis

## 1 Introduction

With the emergence of biochemical databases such as KEGG or Brenda, information about large-scale metabolic networks became easily accessible. Several approaches for the structural analysis of metabolic networks have been introduced, such as flux balance analysis, the concept of elementary flux modes or extreme pathways, as well as graph theoretical analyses. Recently, we introduced a new approach of network analysis based on the concept of scopes [1]. This method allows to determine all compounds which can in principle be synthesized from a given set of substrates using a specified set of biochemical reactions.

## 2 Method and Results

Starting with given metabolites, called seed compounds, the algorithm selects from a database in an iterative manner those reactions whose substrates are either part of the set of seed compounds or are products of reactions which were already selected in an earlier iteration step. This network expansion process ends when no further reactions fulfilling this condition can be found. All metabolites which can be produced by the resulting set of reaction form the scope of the seed compounds. Scopes describe therefore the synthesizing capacity of the corresponding seed compounds in a specified metabolic network.

In this work we confine ourselves on scopes of individual metabolites, which means that for each expansion process only one compound is used as seed. We systematically calculated the scopes of 4104 compounds included in the KEGG database. As mentioned in [1] different seed compounds may have identical scopes leaving only 2922 distinct scopes. A scope may be included in another scope. If the scope of compound  $B$ ,  $\Sigma(B)$ , is a proper subset of the scope of  $A$ ,  $\Sigma(A)$ , then  $\Sigma(A)$  is a superscope of  $\Sigma(B)$  and  $\Sigma(B)$  is a subscope of  $\Sigma(A)$ . There exists nesting in the sense that for three compounds  $A$ ,  $B$ ,  $C$  the relation  $\Sigma(C) \subset \Sigma(B) \subset \Sigma(A)$  holds. This inclusion relations can be compiled into a directed acyclic graph, where nodes represent the distinct scopes. Directed edges point from superscopes to their subscopes, where in case of nesting as described above the edge between node  $A$  and  $C$  is omitted.

Figure 1 shows the largest connected component of this graph for the analysed biochemical network. This component contains the most relevant information as the other nodes are either isolated or connected to only very few other nodes. Edges always point in downward direction. Therefore nodes near the bottom of the graph represent small scopes which only have a small number of subscopes, while larger scopes are situated rather at the top.

Remarkably, the nodes of the graph differ strongly with respect to both, their in and out degree. A node has a large out degree (mushroom shaped structures) if its scope contains many different

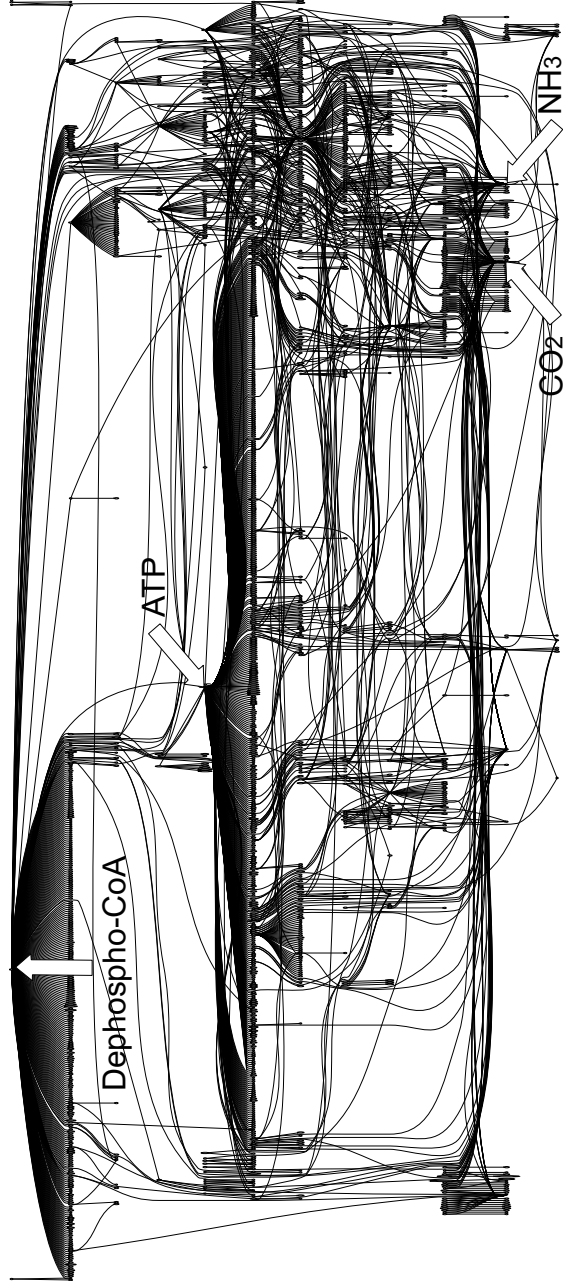


Figure 1: Scope Hierarchy of the KEGG biochemical network. Nodes represent scopes and edges point from superscopes to their subscopes. The nodes for the scopes of ATP, Dephospho-CoA, CO<sub>2</sub> and NH<sub>3</sub> are indicated. A large in degree means that the corresponding scope is a subscope of many different superscopes.

Systematic analysis of the graph reveals that 1191 out of 2922 nodes are isolated. Among the connected nodes 193 are sources (scopes that do not have superscopes) and 1279 are sinks (scopes that do not have subscopes). The nodes representing the scopes of ATP and Dephospho-CoA have the highest out-degrees (493 and 317) corresponding to the pronounced mushroom shaped structures at the top (Dephospho-CoA) and in the middle (ATP). The nodes for the scopes of CO<sub>2</sub> and NH<sub>3</sub> have the highest in-degrees (63 and 57) and both are situated near the bottom of the graph.

### 3 Discussion

The analysis of the scopes of single compounds reveals a hierarchy of metabolites which orders them according to their synthesizing capacity. This hierarchy also implies an ordering according to the complexity of the structure of the compounds, placing more complex compounds at the top while leaving more simple building blocks at the bottom.

The overall structure of the graph is strongly influenced by the chemical composition of the metabolites. Typically, complex compounds which are composed of several different subunits can be split up, implying that their scopes are generally large and are superscopes of the scopes of all their subunits. In contrast complex compounds can generally not be formed from a single type of subunit. However, complex compounds may have small scopes if the necessary reactions are not present. This fact is responsible for the large number of sink nodes in the graph (74% of the connected nodes are sinks).

So far our calculations concern the scopes of a metabolic network consisting of all known biochemical reactions. The same methods can directly be applied to the networks of specific organisms allowing for elucidation of their synthesizing capacities.

### References

- [1] Handorf, T., Ebenhöf, O., Heinrich, R., Expanding Metabolic Networks: Scopes of Compounds, Robustness, and Evolution, *J. Mol. Evol.*, 61:498-512, 2005