Flexible database-assisted graphical representation of metabolic networks for model comparison and the display of experimental data

Jana Tillack, Melanie Bende, Michael Rother, Maurice Scheer, Susanne Ulas, Dietmar Schomburg

Department of Bioinformatics and Biochemistry
Institute for Biochemistry, Biotechnology and Bioinformatics
Technische Universität Carolo-Wilhelmina zu Braunschweig
Langer Kamp 19B
38106 Braunschweig, Germany
j.tillack@tu-braunschweig.de
d.schomburg@tu-braunschweig.de

Abstract: Intracellular processes in living organisms are described by metabolic models. A visualization of metabolic models assists interpretation of data or analyzing results. We introduce the visualization tool DaViMM creating personalized graphical representations of metabolic networks for model comparison or the display of measurements or analyzing results. The tool is coupled to a relational database containing graphical network properties like coordinates, which ensure an intuitive network layout. A combination of DaViMM, the graphical database, and available biochemical databases enables an automated creation of metabolic network maps. The flexibility of this combination is demonstrated with some application examples.

1 Introduction

A genome-scale metabolic model comprises all information known about the metabolic interactions in an organism of interest [FHT+09]. Metabolic network maps are the graphical representation of metabolic models and support interpretation of data or analyzing results. In network maps, each substance and each reaction is visualized by a node and these nodes are connected via edges related to their metabolic role. In other words, all substances involved in a reaction as substrates or products are connected with this reaction by edges (Figure 1).

The benefits of the graphical representation of metabolic networks are obvious, but the process of its creation is often complex. Manual drawing of metabolic networks is time-consuming and user-dependent, but the networks are usually easy to handle and intuitive. Automatic drawing is fast and reproducible, but the resulting layout is often unstructured regarding biology and thus non-intuitive.

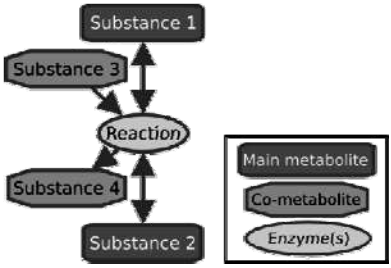

Figure 1: Nodes and edges combined to a metabolic reaction.

Different approaches try to combine the benefits of manual and automatic creation of metabolic networks [CBC+12]. Here, we present an automatic creation of metabolic network maps supported by a database containing positioning information in terms of coordinates for each node. The graphical information combined with biological information from databases like BRENDA (Braunschweig Enzyme Database, [SCP+12]) and BKM-react [LSS11] draw a complete and easy to handle metabolic network map for diverse visualization purposes.

The tool we are presenting produces personalized metabolic network maps based on a predefined database. The generated maps have a defined and intuitive layout and can be built in less than one minute. The tool is flexible and allows creating an effective visualization of genome-scale metabolic models, experimental data, or analyzing results in the context of metabolic networks.

2 Information Visualization in the Context of Metabolic Networks

2.1 Database - Basis for Visualization

To speed up the generation of metabolic network maps, all information needed for reproducible visualization is stored in a predefined database. This core database assisting the drawing of metabolic network maps has been developed in three main steps:

1. Manual creation of a genome-scale metabolic network map to get positioning, identity, and role of each node. This basic network has been created in Cytoscape [SMO+03] and the following information is defined manually within this network:
   Positioning information: coordinates
   Identity: ligand ID, EC number
   Role: enzyme/reaction, substrate/product, transporter

2. Storage of the graphical information in a MySQL database.

3. Addition of biochemical information to the database. This information is taken from BRENDA and BKM-react and based on the defined identifier. For the different node types the following information is added:

Enzyme: EC number, recommended name, enzyme to reaction allocation, enzyme to node allocation, enzyme to organism allocation

Reaction: reaction to pathway allocation

Substance: BRENDA ligand ID, recommended name, short name, synonyms, substance to reaction allocation with stoichiometric information

The resulting database contains all graphical and biochemical information for 1882 reactions of the metabolism and allows fast and reproducible map generation. The database content will continuously be extended.

This publication will not deal with the database generation in detail, but using the information stored in the database to build metabolic network maps for various applications.

2.2 Visualizing Metabolic Networks

Based on the introduced database personalized metabolic network can be visualized. For this purpose, the tool DaViMM has been developed in the programming language Python. The general workflow of network generation is presented in Figure 2.

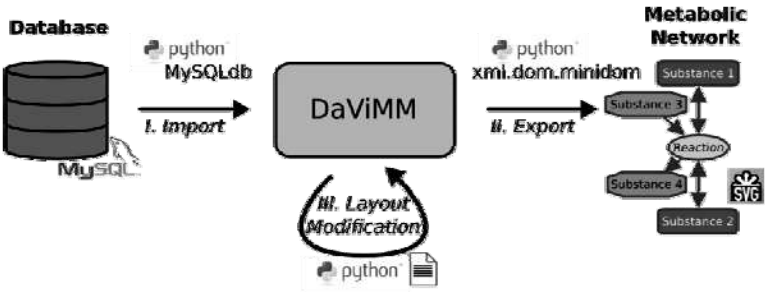

Figure 2: Workflow of metabolic network generation.

First, the information included in the database about nodes and edges is imported. From that point on, the nodes and edges can be written into a file in vector image file format (svg) composing a metabolic network.

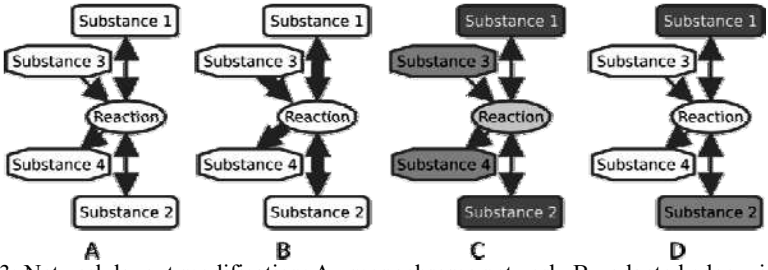

Figure 3: Network layout modification: A - monochrome network, B - adapted edge width, C - node colors depending on the node kind, D - node colors depending on data.

Additionally, the visual properties of nodes and edges can be changed before export via import of a data file or definition of the layout in prepared configuration files (Figure 3).

The opportunity of network layout modification allows to personalize the networks for presentations or publications and to visualize various kinds of data in the context of a network. The following sections will give an insight in how flexible the metabolic network maps can be applied.

2.3 Metabolic Network Maps integrated into BRENDA

One application for the network visualization is the database BRENDA. BRENDA contains information about enzymes and ligands involved in enzymatic conversions.

An integration of metabolic network maps into the BRENDA website enhanced the classification of e.g. an enzyme. The network maps characterize the surrounding of an enzyme or a substance of interest, they allocate enzymes to pathways via their location, they allocate enzymes to organisms via highlighting of nodes, and they allow visualizing taxonomic information for an organism using multiple colors.

The basic metabolic network maps generated for BRENDA are an overview map containing one pseudo node for each pathway and pathway maps containing reactions and substances. The pathways in the overview map are classified in four main groups - carbohydrate and lipid metabolism, nucleotide and amino acid metabolism, energy metabolism, and secondary metabolism (Figure 4). The pathway names are always shown on the map or will appear on mouseover depending on the size of the pathways.

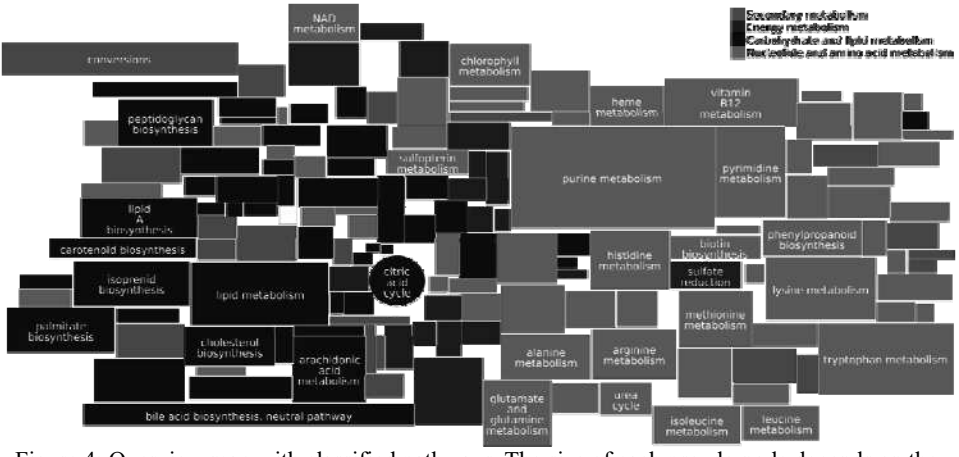

Figure 4: Overview map with classified pathways. The size of each pseudo node depends on the number of nodes located in the pathway.

Starting with this overview map, all single pathways can be selected. In each pathway the substances and enzymes are linked to the BRENDA website via the node identifier. Due to the complex interactions in a metabolic network, edges may combine nodes

located in different pathways. For pathway visualization external substance nodes linked to internal enzyme nodes will be shifted to the pathway boundary (Figure 5).

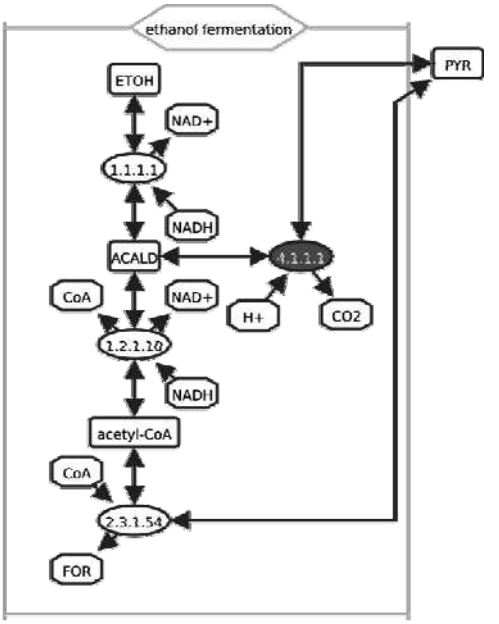

Figure 5: Pathway map with highlighted enzyme.

2.4 Data Visualization in the Context of Metabolic Networks

Data visualization for publication or presentation is often hand-made. To support and speed up the visualization process the network maps, always known from the sections before, can be modified to visualize data in the context of a metabolic network.

To change network properties depending on information stored in data files, different methods have been developed and are presented in detail.

Highlight of nodes of a list: The first method simply highlights all nodes included in the input file by filling the node with one defined color. Therefore, the name used in the input file will be compared with the synonyms stored for each element of the metabolic pathway.

Highlight of models: A second method highlights all reaction nodes included in a model in text format. Additionally, models can directly be compared within one map with another model or with database knowledge. Database knowledge means all enzymes known for an organism or the taxonomic information of an organism from BRENDA. In contrast to the method before, matching takes place based on all substrates and products involved in a reaction.

Visualization of data: Most of the other methods assign measurement values to the nodes. To use these methods, the data files should contain pairs of node name and measurement value. The nodes will be filled with a color depending on the measurement value and the defined colors. Two different types of data visualization will be distinguished.

1. If data only include positive values, the nodes will be colored using one gradient. An example for this kind of visualization is the display of metabolite concentrations e.g. from a metabolome analysis (Figure 6A).
2. For coloring of nodes with data, which can be negative as well as positive, two gradients will be combined. This method will be applied to visualize fold changes, e.g. for gene expression or enzyme activity (Figure 6B).

Visualization of flux rates: The last method assigns analyzing results to the edges by adapting the width of each edge based on e.g. results of flux balance or ^{13}C flux analysis (Figure 6C).

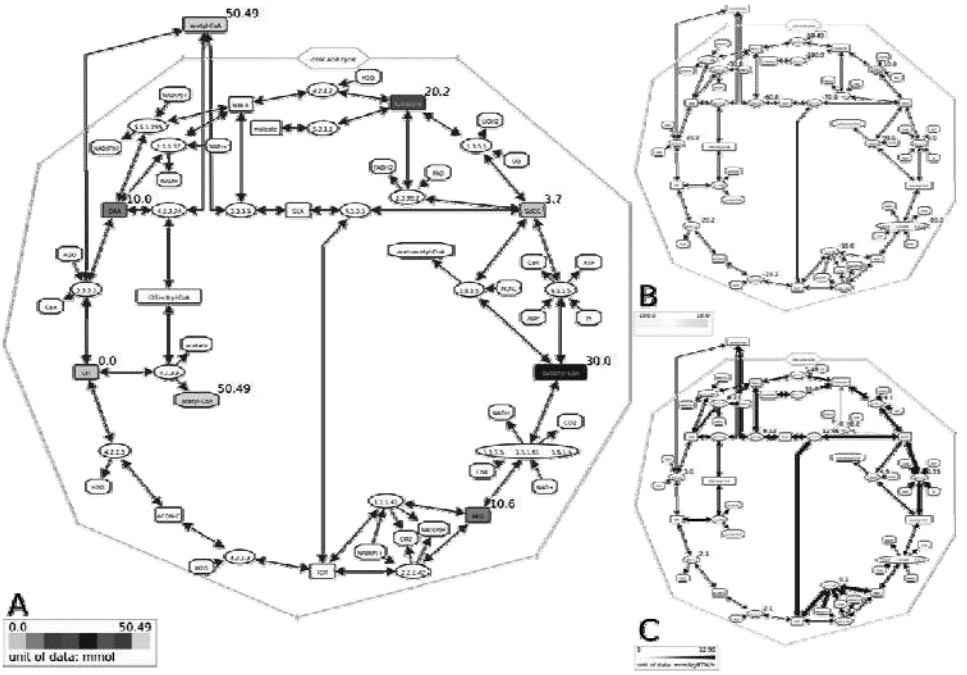

Figure 6: Different methods of visualizing data and analyzing results: A - absolute metabolite concentrations, B - fold changes, C - flux rates.

2.5 Network Maps Supporting Modeling Projects

All different visualization applications introduced so far can be combined to support different steps of metabolic modeling. A detailed description of the steps occuring during genome-scale stoichiometric modeling can be found in [TP10]. In the following selected

steps will be discussed regarding their potential to be supported by metabolic network maps.

Model generation: An overview of all known enzymes for the organism of interest is a good starting point of the modeling procedure. This information is available in databases and can be visualized using the introduced methods. In addition to the organism of interest itself, visualization of members of its taxonomic tree will give information about enzymes active in similar organisms which might also be relevant for the organism of interest (Figure 7).

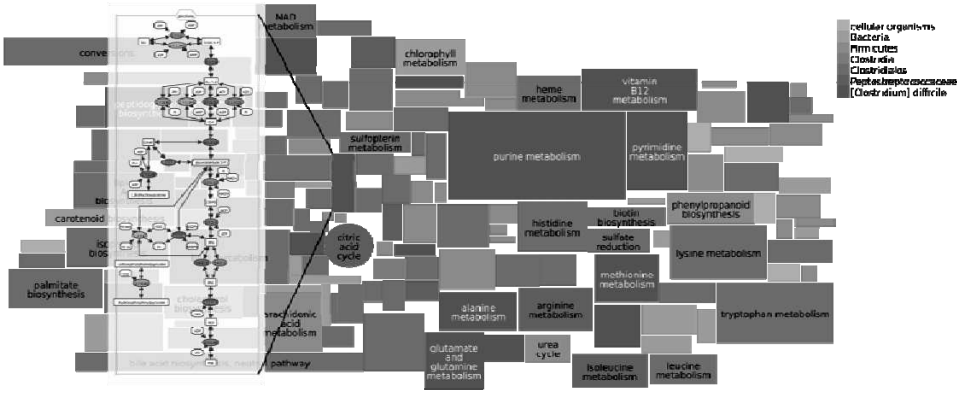

Figure 7: Taxonomic information from BRENDA for the organism of interest. Different taxonomic levels are visualized by the colors of pseudo nodes and nodes.

Besides database knowledge, measurements are needed for model generation. All kinds of omics data can be visualized in the context of a metabolic network map compared to Figure 6.

Model verification: Nearly all of the introduced visualization methods can support model verification. The first approach is to highlight all enzymatic reactions included in the model in a metabolic network map, e.g. to find gaps. Furthermore, the model reactions can be compared with database information (Figure 8), all reactions of another model, or measurements (Figure 6).

Finally, the model is used to generate hypotheses about the organisms' behaviour. Therefore, mathematical algorithms, e.g. Flux-Balance Analysis, are applied to the network. The analyzing results in terms of flux ratios can also be visualized in the network context (Figure 6C).

3 Conclusion

The visualization of metabolic networks helps to understand the complex interactions in an organism. Taking the network maps as basis for the visualization of measurements or analyzing results supports data interpretation.

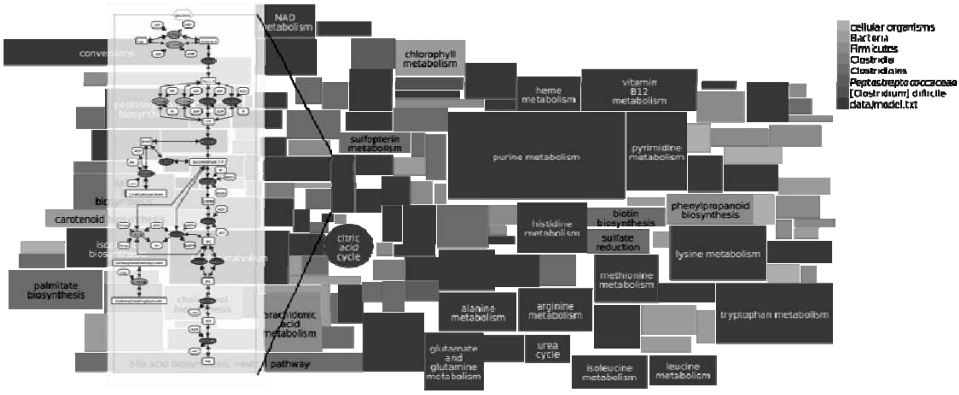

Figure 8: Comparison of the model and taxonomic information from BRENDA for the modeled organism. Different taxonomic levels (blue) and the model reactions (grey) are visualized by the colors of pseudo nodes and nodes.

The challenge in metabolic network visualization is to find the balance between time consuming manual creation and an intuitive network layout. The introduced tools and databases allow building network maps automatically. The layout is defined, but the appearance can be modified individually. For this reason, personalized and colorful network maps for presentations and publications as well as for integration into websites can be built with less effort and time.

By being programmed in Python, the visualization tools run on all common operating systems. The source code is available on request.

References

[CBC+12] W. Copeland, B. Bartley, D. Chandran, M. Galdzicki, K. Kim, S. Sleight, C. Maranas, and H. Sauro. Computational tools for metabolic engineering. Metabolic Engineering, 14(3):270–280, May 2012.

[FHT+09] A. Feist, M. Herrgard, I. Thiele, J. Reed, and B. Palsson. Reconstruction of biochemical networks in microorganisms. Nat Rev Microbiol, 7(2):129–143, 2009.

[LSS11] M. Lang, M. Stelzer, and D. Schomburg. BKM-react, an integrated biochemical reaction database. BMC biochemistry, 12(1):42, 2011.

[SCP+12] I. Schomburg, A. Chang, S. Placzek, C. S¨ ohngen, M. Rother, M. Lang, C. Munaretto, S. Ulas, M. Stelzer, A. Grote, M. Scheer, and D. Schomburg. BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. Nucleic acids research, 41(November 2012):764–772, 2012.

[SMO+03] P. Shannon, A. Markiel, O. Ozier, N. Baliga, J. Wang, D. Ramage, N. Amina, B. Schwikowski, and T. Ideker. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research, 13(11):2498–2504, November 2003.

[TP10] I. Thiele and B. Palsson. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc, 5(1):93–121, 2010.