This is a draft of the paper that has been accepted for publication in Journal of Computational Biology.

Please cite this paper as: M. Moškon, N. Zimic, M. Mraz, Journal of Computational Biology, 2018, DOI: 10.1089/cmb.2017.0209

Grohar: automated visualisation of genome-scale metabolic models and their pathways

Miha Moškon*†, Nikolaj Zimic†, Miha Mraz†

January 11, 2018

Abstract

Genome-scale metabolic models (GEMs) have become a powerful tool for the investigation of the entire metabolism of the organism in silico. These models are, however, often extremely hard to reconstruct and also difficult to apply to the selected problem. Visualisation of the GEM allows us to easier comprehend the model, to perform its graphical analysis, to find and correct the faulty relations, to identify the parts of the system with a designated function, etc. Even though several approaches for the automatic visualisation of GEMs have been proposed, metabolic maps are still manually drawn or at least require large amount of manual curation.

We present *Grohar*, a computational tool for automatic identification and visualisation of GEM (sub)networks and their metabolic fluxes. These (sub)networks can be specified directly by listing the metabolites of interest or indirectly by providing reference metabolic pathways from different sources, such as KEGG, SBML or Matlab file. These pathways are identified within the GEM using three different pathway alignment algorithms. Grohar also supports the visualisation of the model adjustments (e.g. activation or inhibition of metabolic reactions) after perturbations are induced. Grohar is freely available at https://bitbucket.org/mmoskon/grohar.

Key words: genome-scale metabolic models, visualisation of metabolic networks, flux balance analysis, pathway alignment, systems biology.

1 Introduction

Several computational tools for automatic or semi-automatic visualisation of genome-scale metabolic models (GEMs) have been developed in recent years. Manual and semi-automatic visualisation tools, such as Escher [1], aid the user in the manual construction of metabolic maps. On the other hand, most of the existent tools for automatic visualisation, such as MetDraw [2] and Cytoscape plugin CySBML [3], fail to produce an acceptable output when applied to large-scale GEMs. Paint4Net [4] allows the user to visualise only selected segments of the GEM, which have to be manually identified. None of these tools, however, do not account for large amount of manually curated and already visualised metabolic pathways. BiKEGG [5] aligns the GEM with an arbitrary KEGG pathway. The tool is, however, not focused to the visualisation of GEMs, but on the overlying of the metabolic fluxes over the KEGG maps. The alignment algorithms are based solely on the KEGG metabolite IDs, which are sometimes missing or misleading. In such cases manual curation is needed.

Here we present Grohar, a computational tool for automatic visualisation of GEMs. Grohar allows the automatic identification and visualisation of arbitrary (sub)networks of the GEM. Moreover, it allows the visualisation of the model perturbations and the alignment of the model with an arbitrary metabolic pathway or model. The alignment can be performed with other metabolic models in SBML or Matlab format or with an arbitrary pathway from the KEGG database [6]. The alignment algorithms are fully automatic and are based on several different properties of modelled metabolic reactions.

*To whom correspondence should be addressed.
†Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia

2 Features and implementation

The user needs to specify a Matlab (MAT) or SBML model in COBRA dialect [7]. The software allows the user to graphically analyse the model with three main functions, which are described in details in the remaining of the paper. All three functions visualise the obtained metabolic sub(network) as a directed bipartite graph, in which reactions and metabolites represent two different node types, and edges represent the relations between the reactions and the metabolites. Reaction directionalities and metabolic fluxes through the reactions are also included in the visualisation (note that flux balance analysis (FBA) [8] is always performed before the visualisation). User manual describing the installation, usage and functions of Grohar is available as Supplementary material.

2.1 Basic visualisation

Basic visualisation allows the user to visualise the neighbourhood of specific *metabolites*. The only mandatory parameter that the user has to specify is a set of metabolites of interest. Grohar will visualise the metabolic (sub)network surrounding, i.e. producing or consuming, these metabolites. The size of the neighbourhood around the metabolites of interest can be increased to also include non-adjacent nodes. The visualisation can also be limited to include solely the reactions producing or the reactions consuming the observed metabolites. The user can specify model *compartments* of interest to limit the size of obtained metabolic (sub)network. Reactions and metabolites that do not belong to the specified compartment will be omitted in the visualisation. If the user specifies two or more compartments, the visualisation will include only the transport reactions (and their corresponding metabolites) between the specified compartments. Model can be simplified before the visualisation. Grohar allows the user to omit the metabolites that are present in too many reactions and/or reactions with too many metabolites. Reactions that do not reflect minimal activity, i.e. metabolix flux through the reaction is lower than the user defined threshold, can also be omitted.

2.2 Visualisation of perturbations

The user can specify perturbations with the additional constraints that are imposed on the model. These can be used to, e.g., decrease the flux through a metabolic reaction, simulate the removal of a metabolite from the medium or to change an objective function. Grohar allows the user to visualise the metabolic reactions (and corresponding metabolites) that become active after the perturbations are introduced, the metabolic reactions that become inactive after the perturbations are introduced, or metabolic reactions that remain active with modified metabolic fluxes after the perturbations are introduced. In all three cases only a subset of metabolites and reactions that are specified by the *Basic visualisation* segment are visualised.

2.3 Alignment and visualisation

Grohar allows the user to automatically identify a specified metabolic pathway within the model. The user can select an arbitrary metabolic pathway from KEGG [6], to which Grohar automatically connects (requires internet connection). Another option is to import a pathway from an SBML or MAT file. Alignment of the pathway with the GEM is performed with the calculation of similarities between the metabolic reactions. Grohar uses three different algorithms for the calculation of reaction similarities:

- *Topological similarity*: the same reaction should have similar topological properties in both sources, i.e. in the GEM as well as in the pathway. Grohar implements the extension of the MBPR method [9] to calculate the topological similarities. This method depicts the network of metabolic reactions as a directed graph, from which metabolites are omitted. Reactions from the GEM and from the pathway are then compared on the basis of their topological properties.

- *EC similarity*: the same reaction should be catalysed by the same enzymes in both sources. Enzymes can be identified by the Enzyme Commission (EC) numbers. These are obtained from the BiGG database [10] for SBML and MAT models and from the KEGG database

for KEGG pathways (requires internet connection). EC number comparison is performed according to [11].

- *Metabolite similarity*: the same reaction should have similar reactants and similar products in both sources. If KEGG pathway is used, KEGG metabolite identifiers are used in the comparison (similarly to BiKEGG described in [5]). KEGG metabolite identifiers are obtained from the BiGG database for SBML and MAT models. If SBML or MAT pathway is used, COBRA metabolite identifiers are used in the comparison.

The user can specify weights to be used with each of the algorithms. Different combinations of weights have to be used in different cases, e.g., the topological similarities can be misleading when the reactions are described with different accuracies in different sources.

After the similarities are calculated the reaction mappings from the metabolic pathway to GEM are identified with a greedy algorithm. A single reaction can map to many and/or many reactions can map to a single reaction (specified by the user). Further details describing the alignment and the parameters that can be set by the user are available in the Supplementary material. Aligned metabolic reactions are at the end visualised in a similar manner as described in Section 2.1.

Acknowledgement

This work has been partially supported by the scientific-research programme Pervasive Computing (P2-0359) financed by the Slovenian Research Agency in the years from 2009 to 2017.

References

[1] Z. A. King, A. Dräger, A. Ebrahim, N. Sonnenschein, N. E. Lewis, and B. O. Palsson, "Escher: A web application for building, sharing, and embedding data-rich visualizations of biological pathways," *PLOS Computational Biology*, vol. 11, no. 8, p. e1004321, 2015.

[2] P. A. Jensen and J. A. Papin, "MetDraw: automated visualization of genome-scale metabolic network reconstructions and high-throughput data," *Bioinformatics*, vol. 30, no. 9, pp. 1327–1328, 2014.

[3] M. König, A. Dräger, and H.-G. Holzhütter, "CySBML: a Cytoscape plugin for SBML," *Bioinformatics*, vol. 28, no. 18, pp. 2402–2403, Jul 2012.

[4] A. Kostromins and E. Stalidzans, "Paint4net: COBRA toolbox extension for visualization of stoichiometric models of metabolism," *Biosystems*, vol. 109, no. 2, pp. 233–239, 2012.

[5] O. Jamialahmadi, E. Motamedian, and S. Hashemi-Najafabadi, "BiKEGG: a COBRA toolbox extension for bridging the BiGG and KEGG databases," *Molecular BioSystems*, vol. 12, no. 11, pp. 3459–3466, 2016.

[6] M. Kanehisa, S. Goto, M. Hattori, K. F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, and M. Hirakawa, "From genomics to chemical genomics: new developments in KEGG," *Nucleic Acids Research*, vol. 1, no. 34, pp. D354–7, 2006.

[7] J. Schellenberger, R. Que, R. M. Fleming, I. Thiele, J. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, S. Rahmanian, J. Kang, D. Hyduke, and B. O. Palsson, "Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0," *Nature protocols*, vol. 6, no. 9, pp. 1290–1307, 2011.

[8] J. D. Orth, I. Thiele, and B. O. Palsson, "What is flux balance analysis?" *Nature Biotechnology*, vol. 28, no. 3, pp. 245–248, 2010.

[9] Y. Huang, C. Zhong, H. X. Lin, and J. Huang, "Aligning metabolic pathways exploiting binary relation of reactions," *PLoS One*, vol. 11, no. 12, p. e0168044, 2016.

[10] Z. A. King, J. Lu, A. Dräger, P. Miller, S. Federowicz, J. A. Lerman, A. Ebrahim, B. O. Palsson, and N. E. Lewis, "BiGG models: A platform for integrating, standardizing and sharing genome-scale models," *Nucleic Acids Research*, vol. 44, no. D1, pp. D515–D522, 2015.

[11] M. Heymans and A. K. Singh, "Deriving phylogenetic trees from the similarity analysis of metabolic pathways," *Bioinformatics*, vol. 19, no. Suppl 1, pp. i138–i146, 2003.