Knowledge-based generalization of metabolic networks: An applicational study

Anna Zhukova and David James Sherman

Inria / Université Bordeaux 1 / CNRS joint project-team MAGNOME,

351, cours de la Libération, F-33405 Talence Cedex, France

`{anna.zhukova, david.sherman}@inria.fr`

Genome-scale metabolic networks are complex systems that describe thousands of reactions participating in the organism's metabolism. During the process of genome network reconstruction these reactions are automatically inferred from pathway and reaction databases, and existing models for similar organisms, using genomic data [1]. The inferred draft network is then refined during several iterations of error detection, gap filling, analysis and improvement [2]. Although automatic tools for model inference and analysis are becoming more and more powerful, they may still miss some reactions or add those ones that should not belong to the network of the target organism. That is why the analysis by a human expert is needed during the network refinement process. However, being tailored for a computer simulation, and thus including all the reactions thought to participate in the organism's metabolism, genome-scale networks can be too complicated and detailed for a human. The errors may be hidden in the multitude of reactions.

To help a human expert understanding these detailed networks, we developed a method for knowledge-based generalization that focusses on the higher-level relationships in the network, while omitting the details [3]. The generalization process groups chemical species present in the network into semantically equivalent classes, based on their hierarchical relationships in the *ChEBI* ontology [4], and merges them into a generalized chemical species. For instance, *butyryl-CoA*, *hexanoyl-CoA* and *octanoyl-CoA* species can be generalized into *fatty acyl-CoA*. After the species generalization, reactions that share the same generalized reactants and the same generalized products, are factored together into a generalized reaction. This provides a higher-level view of the network.

In this poster, we show the application of this generalization method to the network of the yeast *Y. lypolitica* [5]. We analyze the generalized network, and illustrate how it can be used for easier error detection: We show the changes that both initial and generalized networks undergo if the catalyzing enzyme for some of the reactions is missing. The application of the generalization procedure also facilitates network comparison, which we show by comparing the generalized *Y. lypolitica* network to the networks of several other organisms.

1. I. Thiele and B. Ø. Palsson (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature protocols*, **5:**93–121.

2. N. Swainston, K. Smallbone, P. Mendes, D. B. Kell and N. W. Paton (2011) The SuBliMinaL Toolbox: automating steps in the reconstruction of metabolic networks. *Journal of integrative bioinformatics*, **8**:186.

3. A. Zhukova, D. J. Sherman (2013) Knowledge-based scaling of metabolic models. *To be submitted.*

4. P. de Matos, R. Alcantara, A. Dekker, M. Ennis, J. Hastings, K. Haug, I. Spiteri, S. Turner and C. Steinbeck (2010) Chemical Entities of Biological Interest: an update. *Nucleic Acids Research*, **38**:D249–D254.

5. N. Loira, T. Dulermo, J.-M. Nicaud and D. J. Sherman (2012) A genome-scale metabolic model of the lipid-accumulating yeast *Yarrowia lipolytica*. *BMC Systems Biology*, **6**:35.