Reconstruction and Validation of RefRec: a Global Model for the Yeast Molecular Interaction Network

Tommi Aho[1], Henrikki Almusa[2], Jukka Matilainen[2], Antti Larjo[1], Pekka Ruusuvuori[1], Kaisa-Leena Aho[1], Thomas Wilhelm[3], Harri Lähdesmäki[1,4], Andreas Beyer[5], Manu Harju[1], Sharif Chowdhury[1], Kalle Leinonen[1], Christophe Roos[2], Olli Yli-Harja[1]

[1]Department of Signal Processing, Tampere University of Technology, Tampere, Finland; [2]Medicel Ltd., Espoo, Finland; [3]Institute of Food Research, Norwich, United Kingdom; [4]Department of Information and Computer Science, Helsinki University of Technology, Espoo, Finland; [5]Biotechnology Center, Technische Universität Dresden, Dresden, Germany.

Molecular interaction networks establish all cell biological processes. The networks are under intensive research that is facilitated by new high-throughput measurement techniques for the detection, quantification, and characterization of molecules and their physical interactions. For the common model organism yeast *Saccharomyces cerevisiae*, public databases store a significant part of the accumulated information and, on the way to better understanding of the cellular processes, there is a need to integrate this information into a consistent reconstruction of the molecular interaction network.

This work presents and validates RefRec, the most comprehensive molecular interaction network reconstruction currently available for yeast. The reconstruction integrates protein synthesis pathways, a metabolic network, and a protein-protein interaction network from major biological databases. The core of the reconstruction is based on a reference object approach in which genes, transcripts, and proteins are identified using their primary sequences. This enables their unambiguous identification and non-redundant integration. The obtained total number of different molecular species and their connecting interactions is ~67,000. In order to demonstrate the capacity of RefRec for functional predictions, it was used for simulating the gene knockout damage propagation in the molecular interaction network in ~590,000 experimentally validated mutant strains. Based on the simulation results, a statistical classifier was subsequently able to correctly predict the viability of most of the strains. The results also showed that the usage of all molecular domains in the reconstruction is important for accurate phenotype prediction.

In general, the findings demonstrate the benefits of global reconstructions of molecular interaction networks. With all the molecular species and their physical interactions explicitly modeled, our reconstruction is able to serve as a valuable resource in additional analyses involving objects from multiple molecular domains. For that purpose, RefRec will be freely available in the Systems Biology Markup Language format [1] in the BioModels database [2].

[1] Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, et al. (2003) The Systems Biology Markup Language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics*, 19: 524–531.

[2] Le Novère N., Bornstein B., Broicher A., Courtot M., Donizelli M., et al. (2006) BioModels Database: A Free, Centralized Database of Curated, Published, Quantitative Kinetic Models of Biochemical and Cellular Systems *Nucleic Acids Res.*, 34: D689-D691.