Saccharomyces cerevisiae iND750: A Fully Compartmentalized Genome-Scale Metabolic Model

Natalie C. Duarte, Markus J. Herrgård, and Bernhard Ø. Palsson
Department of Bioengineering, University of California, San Diego
9500 Gilman Drive, La Jolla, California, 92093-0412, USA

Abstract

Metabolic models are useful integration tools, allowing us to incorporate data from such diverse sources as gene expression profiles, proteomics studies, and genome annotations. Genome-scale constraint-based metabolic networks have been built for several microorganisms. The first genome-scale reconstruction of a eukaryotic organism, the yeast *Saccharomyces cerevisiae*, was recently completed by Förster and Famili [1].

As part of the iterative model building process, the contents of the Förster-Famili were used as a basis for the development of an expanded network, named *S. cerevisiae* iND750 [2]. There are three primary features that distinguish iND750 from the Förster-Famili model. First, iND750 is fully compartmentalized, accounting for eight cellular compartments (Table 1). A second distinguishing feature is that iND750 directly incorporates genomic, transcriptomic, and proteomic data as associations between genes, transcripts, proteins, and reactions (Figure 2). Finally, the reactions in iND750 have been formulated so that they are both elementally and charge-balanced, enabling a cell-wide proton balance.

To comprehensively evaluate iND750's performance, *in silico* phenotypic predictions for 682 gene deletion strains on seven media compositions were compared to *in vivo* results from two large-scale experiments [3,4]. In total, 4,154 *in silico* and *in vivo* gene deletion phenotypes were compared in this study (Table 2). In 82.6% of the cases, iND750's predictions agreed with the *in vivo* measurements.

Analysis of the failure modes (Figure 3) showed that false predictions were primarily caused by iND750's limited inclusion of cellular processes outside of metabolism. The analysis also led to the identification of gaps and inconsistencies in the body of information used for the reconstruction. In many cases, the false predictions led to direct suggestions of how to potentially improve the model or of specific experiments that could be performed to further improve our understanding of yeast metabolism.

What are constraint-based models?

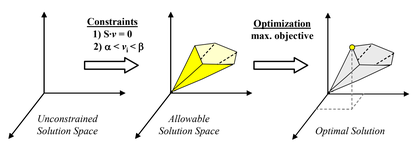

Figure 1: Principles of constraint-based modeling

Unlike kinetic models which find one solution to a system of equations, constraint-based models use physico-chemical constraints to eliminate solutions, leaving a set of feasible solutions defining the allowable solution space.

iND750 is an Expanded, Fully Compartmentalized Metabolic Model of *S. cerevisiae*

Table 1: Comparison of the Förster-Famili model and iND750. Nearly all of iND750's genes and metabolites have been inherited from the Förster-Famili model. The difference in reaction content is primarily due to compartmentalization.

	Förster-Famili	*S. cerevisiae* iND750
Genes	708	750 (94% conserved)
Metabolites	584	646 (90% conserved)
Reactions	842	1149 (56% conserved)*
Compartments	Cytosol Mitochondrion Extracellular Space	Cytosol Mitochondrion Extracellular Space Peroxisome Golgi Apparatus Endoplasmic Reticulum Nucleus Vacuole

* Reactions that only differed by protons and water molecules were also considered to be conserved.

Gene-Protein-Reaction Associations Represent the Logical Relationships between ORFs, Transcripts, Proteins, and Reactions

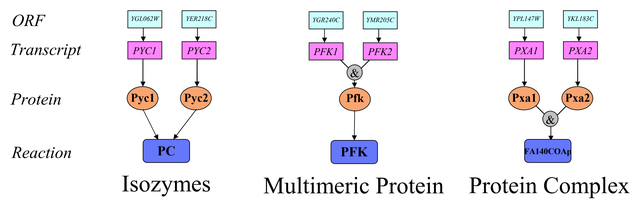

Figure 2: Gene-protein-reaction associations incorporate various types of "-omics" data

iND750 Correctly Predicts 3430/4154 Growth Phenotypes under Seven Media Conditions

The media conditions included in this study were aerobic growth on glucose minimal media (MMD) and on rich media with six different carbon sources: galactose (YPGal), glucose (YPD), glucose-ethanol-glycerol mixed media (YPDGE), glycerol (YPG), ethanol (YPE), and lactate (YPL) Predictions were also made for genes whose deletion strains could not be constructed (essential genes) and for those whose deletion lead to a slow growing strain on rich media (slow growth genes). The results were classified as true positive (experimentally and *in silico* viable), true negative (experimentally and *in silico* growth retarded), false positive (experimentally growth retarded, *in silico* viable), or false negative (experimentally viable, *in silico* growth retarded).

Table 2: Overall results of the comparison between *in silico* and *in vivo* gene deletion studies. The number of phenotypes per condition is shown in parentheses.

	Essential (118)	*Slow (83)*	*MMD (564)*	*YPGal (564)*	*YPD (565)*	*YPDGE (565)*	*YPG (565)*	*YPE (565)*	*YPL (565)*	*All (4154)*
True positive	0	0	439	476	474	465	466	461	466	3247
False positive	81	67	74	69	73	64	62	60	61	611
True negative	37	16	35	7	3	17	23	23	22	183
False negative	0	0	16	12	15	19	14	21	16	113
Correct rate	**31.4**	**19.3**	**84.0**	**85.6**	**84.4**	**85.3**	**86.5**	**85.7**	**86.4**	**82.6**

Analysis of Failure Modes Revealed that False Predictions were Primarily Caused by iND750's Limited Inclusion of Cellular Processes Outside of Metabolism

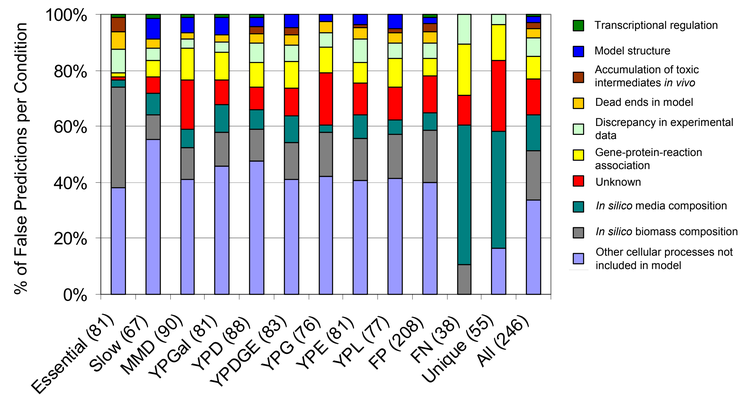

Figure 3: Breakdown of the false predictions by source. Results are shown for each experimental condition including essential genes (Essential) and slow growth genes (Slow) on rich media. In addition, the distributions of the sources of false predictions are shown separately for false positive (FP), false negative (FN), and unique false predictions (Unique).

Conclusions

- Multi-compartmental metabolic models of eukaryotic cells with elementally and charge-balanced reactions can be successfully built.
- Gene-protein-reaction associations allow for the integration of diverse data sets.
- Genome-scale models can be used to compute growth phenotypes of organisms with altered genotypes in various media conditions.
- Analysis of the failure modes can be used to identify gaps or inconsistencies in our knowledge base that require further experimental investigation.

References

1. Forster, J., et al., *Genome-Scale Reconstruction of the Saccharomyces cerevisiae Metabolic Network.* Genome Research, 2003. **13**(2): p. 244-53.
2. Duarte, N.C., M.J. Herrgård, and B.O. Palsson, *Saccharomyces cerevisiae iND750: A fully compartmentalized genome-scale metabolic model that correctly predicts 3430/4514 growth phenotypes under seven media conditions.* submitted.
3. Giaever, G., et al., *Functional profiling of the Saccharomyces cerevisiae genome.* Nature, 2002. **418**(6896): p. 387-91.
4. Steinmetz, L.M., et al., *Systematic screen for human disease genes in yeast.* Nat Genet, 2002. **22**: p. 22.

For more information, please visit http://systemsbiology.ucsd.edu

Acknowledgments

Support for this work was provided by grants from the NSF (MCB98-73384 and BES98-14092) and the NIH (GM57089).