

Standardized network reconstruction of *E. coli* metabolism

Kieran Smallbone

Manchester Centre for Integrative Systems Biology
131 Princess Street, Manchester M1 7DN, UK
kieran.smallbone@manchester.ac.uk

Abstract

We have created a genome-scale network reconstruction of *Escherichia coli* metabolism. Existing reconstructions were improved in terms of annotation standards, to facilitate their subsequent use in dynamic modelling. The resultant network is available from EcoliNet (<http://ecoli.sf.net/>).

EcoliNet

The structure of metabolic networks can be determined by a reconstruction approach, using data from genome annotation, metabolic databases and chemical databases [1]. We built upon an existing reconstruction of the metabolic network of *E. coli* that was based on genomic and literature data (known as iJO1366, [2]). This model contains 1366 genes, 2251 metabolic reactions, and 1136 unique metabolites. Comparison to experimental data sets shows that it makes accurate phenotypic predictions of growth on different substrates and for gene knockout strains [2].

iJO1366 suffers from the use of non-standard names and is not annotated with methods that are machine-readable. The model was thus updated according to existing community-driven annotation standards [3]. The reconstruction is described and made available in Systems Biology Markup Language (SBML) (<http://sbml.org/>, [4]), an established community XML format for the mark-up of biochemical models that is understood by a large number of software applications. The network is available from EcoliNet (<http://ecoli.sf.net/>).

Annotation

The highly-annotated network is primarily assembled and provided as an SBML file. Specific model entities, such as species or reactions, are annotated using ontological terms. These annotations, encoded using the resource description framework (RDF) [5], provide the facility to assign definitive terms to individual components, allowing software to identify such components unambiguously and thus link model components to existing data resources [6]. Minimum Information Requested in the Annotation of Models (MIRIAM, [7]) –compliant annotations have been used to identify components unambiguously by associating them with

one or more terms from publicly available databases registered in MIRIAM resources [8]. Thus this network is entirely traceable and is presented in a computational framework.

Nine different databases are used to annotate entities in the network (see Table 1). The Systems Biology Ontology (SBO) [9] is also used to semantically discriminate between entity types. Eight different SBO terms are used to annotate entities in the network (see Table 2).

example	identifier	database
EcoliNet	562	taxonomy
EcoliNet	21988831	pubmed
cytoplasm	GO:0005737	obo.go
(-)-ureidoglycolate	C00603	kegg.compound
(-)-ureidoglycolate	CHEBI:57296	chebi
glgB	eco:b3432	kegg.genes
glgB	P07762	uniprot
1,4-alpha-glucan branching enzyme	2.4.1.18	ec-code
2-dehydro-3-deoxygalactonokinase	1555810845	isbn

Table 1: MIRIAM annotations used in the model.

example	SBO term	interpretation
cytoplasm	290	compartment
(-)-ureidoglycolate	247	metabolite
tRNA (Glu)	250	ribonucleic acid
glgB	252	enzyme
1,4-alpha-glucan branching enzyme	176	biochemical reaction
1,4-alpha-glucan transport	185	transport reaction
biomass objective function	397	modelling reaction
glgB → 1,4-alpha-glucan branching enzyme	460	catalyst

Table 2: SBO terms used in the model.

Use

We maintain the distinction between the E. coli GENome scale Network REconstruction (GENRE) [10] and its derived GENome scale Model (GEM) [11]. This is important to differentiate between the established biochemical knowledge included in a GENRE and the modelling assumptions required for analysis or simulation with a GEM. A GENRE serves as a structured knowledge base of established biochemical facts, while a GEM is a model which supplements the established biochemical information with additional (potentially hypothetical) information to enable computational simulation and analysis [12]. Reactions added to the GEM include the biomass objective function – a sink representing cellular growth – and hypothetical transporters.

Three versions of the network are made available:

- `<organism>_<version>.xml`, a GEM for use in flux analyses, provided in Flux Balance Constraints (FBC) format [13]
- `<organism>_<version>_cobra.xml`, the same GEM network, provided in Cobra format [14]
- `<organism>_<version>_recon.xml`, a GENRE containing only reactions for which there is experimental evidence

YeastNet

YeastNet is an annotated metabolic network of *Saccharomyces cerevisiae* S288c that is periodically updated by a team of collaborators from various research groups. It started on the shoulders of previous reconstructions of the yeast metabolic network that were published separately (iLL672 [15] and iMM904 [16]). However, due to the different approaches utilised, those earlier reconstructions had a significant number of differences. A community effort in 2007 resulted in a consensus network representation of yeast metabolism, reconciling the earlier results.

As of December 2012, six versions of the network have been released (see Table 3).

version	date	publications
1	February 2008	[3]
2	June 2009	–
3	October 2009	–
4	March 2010	[17]
5	September 2011	[12]
6	December 2012	–

Table 3: Development of YeastNet

The EcoliNet and YeastNet networks are structured identically to facilitate comparative studies. YeastNet is available from <http://yeast.sf.net/>.

Acknowledgements This work is deliverable 4.1 of the EU FP7 (KBBE) grant 289434 “BioPreDyn: New Bioinformatics Methods and Tools for Data-Driven Predictive Dynamic Modelling in Biotechnological Applications”.

References

- [1] Palsson BØ, Thiele I: **A protocol for generating a high-quality genome-scale metabolic reconstruction.** *Nature Protoc* 2010, **5**:91–121. doi:10.1038/nprot.2009.203
- [2] Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, Palsson BØ: **A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism – 2011.** *Mol Syst Biol* 2011, **7**:535. doi:10.1038/msb.2011.65

- [3] Herrgård MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, Blüthgen N, Borger S, Costenoble R, Heinemann M, Hucka M, Le Novère N, Li P, Liebermeister W, Mo M, Oliveira AP, Petranovic D, Pettifer S, Simeonidis E, Smallbone K, Spasić I, Weichart D, Brent R, Broomhead DS, Westerhoff HV, Kirdar B, Penttilä M, Klipp E, Palsson BØ, Sauer U, Oliver SG, Mendes P, Nielsen J, Kell DB: **A consensus yeast metabolic network obtained from a community approach to systems biology.** *Nature Biotechnol* 2008, **26**:1155–1160. doi:10.1038/nbt1492
- [4] Hucka M, Finney A, Sauro H, Bolouri H, Doyle J, Kitano H, Arkin A, Bornstein B, Bray D, Cornish-Bowden A, Cuellar A, Dronov S, Gilles E, Ginkel M, Gor V, Goryanin I, Hedley W, Hodgman T, Hofmeyr J, Hunter P, Juty N, Kasberger J, Kremling A, Kummer U, Le Novère N, Loew L, Lucio D, Mendes P, Minch E, Mjolsness E, Nakayama Y, Nelson M, Nielsen P, Sakurada T, Schaff J, Shapiro B, Shimizu T, Spence H, Stelling J, Takahashi K, Tomita M, Wagner J, Wang J: **The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models.** *Bioinformatics* 2003, **19**:524–531. doi:10.1093/bioinformatics/btg015
- [5] Wang XS, Gorlitsky R, Almeida JS: **From XML to RDF: how semantic web technologies will change the design of ‘omic’ standards.** *Nature Biotechnol* 2005, **23**:1099–1103. doi:10.1038/nbt1139
- [6] Kell DB, Mendes P: **The markup is the model: reasoning about systems biology models in the Semantic Web era.** *J Theor Biol* 2008, **252**:538–543. doi:10.1016/j.jtbi.2007.10.023
- [7] Le Novère N, Finney A, Hucka M, Bhalla US, Campagne F, Collado-Vides J, Crampin EJ, Halstead M, Klipp E, Mendes P, Nielsen P, Sauro H, Shapiro B, Snoep JL, Spence HD, Wanner BL: **Minimum information requested in the annotation of biochemical models (MIRIAM).** *Nature Biotechnol* 2005, **23**:1509–1515. doi:10.1038/nbt1156
- [8] Laibe C, Le Novère N: **MIRIAM resources: tools to generate and resolve robust cross-references in Systems Biology.** *BMC Syst Biol* 2008, **252**:538–543. doi:10.1186/1752-0509-1-58
- [9] Courtot M., Juty N., Knüpfer C., Waltemath D., Zhukova A., Drger A., Dumontier M., Finney A., Golebiewski M., Hastings J., Hoops S., Keating S., Kell D.B., Kerrien S., Lawson J., Lister A., Lu J., Machne R., Mendes P., Pocock M., Rodriguez N., Villeger A., Wilkinson D.J., Wimalaratne S., Laibe C., Hucka M., Le Novère N.: **Controlled vocabularies and semantics in systems biology.** *Mol Syst Biol* 2011, **7**:538–543. doi:10.1038/msb.2011.77
- [10] Price ND, Reed JL, Palsson BØ: **Genome-scale models of microbial cells: evaluating the consequences of constraints.** *Nat Rev Microbiol* 2004, **2**:886–897. doi:10.1038/nrmicro1023
- [11] Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ: **Reconstruction of biochemical networks in microorganisms.** *Nat Rev Microbiol* 2008, **7**:129–143. doi:10.1038/nrmicro1949

- [12] Heavner BD, Smallbone K, Barker B, Mendes P, Walker LP: **Yeast 5 – an expanded reconstruction of the *Saccharomyces cerevisiae* metabolic network.** *BMC Syst Biol* 2012, **6**:55. doi:10.1186/1752-0509-6-55
- [13] Olivier BG, Bergmann FT: **Flux Balance Constraints, Version 1 Release 1.** Available from COMBINE. 2013.
- [14] Schellenberger J, Que R, Fleming RM, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, Kang J, Hyduke DR, Palsson BØ: **Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0.** *Nat Protoc* 2011, **6**:1290–1307. doi:10.1038/nprot.2011.308.4
- [15] Kuepfer L, Sauer U, Blank LM: **Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*.** *Genome Res* 2005, **15**:1421–1430. doi:10.1101/gr.3992505
- [16] Mo ML, Palsson BØ, Herrgård MJ: **Connecting extracellular metabolomic measurements to intracellular flux states in yeast.** *BMC Syst Biol* 2009, **3**:37. doi:10.1186/1752-0509-3-37
- [17] Dobson PD, Jameson D, Simeonidis E, Lanthaler K, Pir P, Lu C, Swainston N, Dunn WB, Fisher P, Hull D, Brown M, Oshota O, Stanford NJ, Kell DB, King RD, Oliver SG, Stevens RD, Mendes P: **Further developments towards a genome-scale metabolic model of yeast.** *BMC Syst Biol* 2010, **4**:145. doi:10.1186/1752-0509-4-145