MANCHESTER
1824

The University
of Manchester

A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology

Herrgård, M.J. and Swainston, N. and Dobson, P. and Dunn, W.B. and Arga, K.Y. and Arvas, M. and Blüthgen, N. and Borger, S. and Costenoble, R. and Heinemann, M. and Hucka, M. and Le Novère, N. and Li, P and Liebermeister, W. and Mo, M.L. and Oliveira, A.P. and Petranovic, D. and Pettifer, S. and Simeonidis, E. and Smallbone, K. and Spasić, I. and Weichart, D. and Brent, R. and Broomhead, D.S. and Westerhoff, H.V. and Kırdar, B.I. and Penttilä, M. and Klipp, E. and Palsson, B.Ø. and Sauer, U. and Oliver, S.G. and Mendes, P. and Nielsen, J. and Kell, D.B.

The University of Manchester

A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology

Markus J Herrgård[1,19,20], Neil Swainston[2,3,20], Paul Dobson[3,4], Warwick B Dunn[3,4], K Yalçin Arga[5], Mikko Arvas[6], Nils Blüthgen[3,7], Simon Borger[8], Roeland Costenoble[9], Matthias Heinemann[9], Michael Hucka[10], Nicolas Le Novère[11], Peter Li[2,3], Wolfram Liebermeister[8], Monica L Mo[1], Ana Paula Oliveira[12], Dina Petranovic[12,19], Stephen Pettifer[2,3], Evangelos Simeonidis[3,7], Kieran Smallbone[3,13], Irena Spasić[2,3], Dieter Weichart[3,4], Roger Brent[14], David S Broomhead[3,13], Hans V Westerhoff[3,7,15], Betül Kırdar[5], Merja Penttilä[6], Edda Klipp[8], Bernhard Ø Palsson[1], Uwe Sauer[9], Stephen G Oliver[3,16], Pedro Mendes[2,3,17], Jens Nielsen[12,18] & Douglas B Kell*[3,4]

Genomic data allow the large-scale manual or semi-automated assembly of metabolic network reconstructions, which provide highly curated organism-specific knowledge bases. Although several genome-scale network reconstructions describe *Saccharomyces cerevisiae* metabolism, they differ in scope and content, and use different terminologies to describe the same chemical entities. This makes comparisons between them difficult and underscores the desirability of a consolidated metabolic network that collects and formalizes the 'community knowledge' of yeast metabolism. We describe how we have produced a consensus metabolic network reconstruction for *S. cerevisiae*. In drafting it, we placed special emphasis on referencing molecules to persistent databases or using database-independent forms, such as SMILES or InChI strings, as this permits their chemical structure to be represented unambiguously and in a manner that permits automated reasoning. The reconstruction is readily available via a publicly accessible database and in the Systems Biology Markup Language (http://www.comp-sys-bio.org/yeastnet). It can be maintained as a resource that serves as a common denominator for studying the systems biology of yeast. Similar strategies should benefit communities studying genome-scale metabolic networks of other organisms.

Accurate representation of biochemical, metabolic and signaling networks by mathematical models is a central goal of integrative systems biology. This undertaking can be divided into four stages[1]. The first is a qualitative stage in which are listed all the reactions that are known to occur in the system or organism of interest; in the modern era, and especially for metabolic networks, these reaction lists are often derived in part from genomic annotations[2,3] with curation based on literature ('bibliomic') data[4]. A second stage, again qualitative, adds known effectors, whereas the third and fourth stages—essentially amounting to molecular enzymology—include the known kinetic rate equations and the values

of their parameters. Armed with such information, it is then possible to provide a stochastic or ordinary differential equation model of the entire metabolic network of interest. An attractive feature of metabolism, for the purposes of modeling, is that, in contrast to signaling pathways, metabolism is subject to direct thermodynamic and (in particular) stoichiometric constraints[3]. Our focus here is on the first two stages of the reconstruction process, especially as it pertains to the mapping of experimental metabolomics data onto metabolic network reconstructions.

Besides being an industrial workhorse for a variety of biotechnological products, *S. cerevisiae* is a highly developed model organism for biochemical, genetic, pharmacological and post-genomic studies[5]. It is especially attractive because of the availability of its genome sequence[6], a whole series of bar-coded deletion[7,8] and other[9] strains, extensive experimental 'omics data[10–14] and the ability to grow it for extended periods under highly controlled conditions[15]. The very active scientific community that works on *S. cerevisiae* has a history of collaborative research projects that have led to substantial advances in our understanding of eukaryotic biology[6,8,13,16,17]. Furthermore, yeast metabolic physiology has been the subject of intensive study and most of the components of the yeast metabolic network are relatively well characterized. Taken together, these factors make yeast metabolism an attractive topic to test a community approach to build models for systems biology.

Several groups[18–21] have reconstructed the metabolic network of yeast from genomic and literature data and made the reconstructions freely available. However, due to different approaches used to create them, as well as different interpretations of the literature, the existing reconstructions have many differences. Additionally, the naming of metabolites and enzymes in the existing reconstructions was, at best, inconsistent, and there were no systematic annotations of the chemical species in the form of links to external databases that store chemical compound information. This lack of model annotation complicated the use of the models for data analysis and integration. Members of the yeast systems biology community therefore recognized that a single 'consensus' reconstruction and annotation of the metabolic network was highly desirable as a starting point for further investigations.

A crucial factor that enabled the building of a consensus network reconstruction is the ability to describe and exchange biochemical network

*A list of affiliations appears at the end of the paper.

models in a standard format, the Systems Biology Markup Language (SBML; http://www.sbml.org/)[22]. The SBML format is employed by most commonly used software applications for visualizing, simulating and analyzing biochemical networks, and also in pathway databases. SBML also provides the necessary standardized means ('Minimum Information Requested in the Annotation of biochemical Models' or MIRIAM[23]) to annotate models with information that is required to identify network components uniquely, including metabolites, proteins and genes. Representing the consensus metabolic network reconstruction in a MIRIAM-compatible SBML format allows widespread use of the reconstruction and assists in its continued curation, expansion and revision.

We developed this consensus reconstruction using a 'jamboree' approach—a large, focused work meeting, where we defined the protocol for the curation process as well as resolving the majority of discrepancies between the existing reconstructions. The jamboree event was followed by an extended process of curation of remaining discrepancies and careful annotation of components of the reconstructions by a smaller group of people. The overall goal of the effort was, by careful curation and comprehensive annotation of the network model and its components, to make the consensus reconstruction useful for the broadest possible set of users. The general reconstruction could then be used directly in bioinformatics applications aimed at integration of, for example, metabolomics and proteomics data or as a starting point for building predictive models using a number of different approaches[24,25], and for other purposes outlined below.

Here we describe how an initial 'community consensus' reconstruction of the yeast metabolic network was carried out. We make some further proposals for how this reconstruction of the yeast metabolic network may evolve as more information is acquired. We also discuss the possibility of using a similar approach to build consensus models of metabolic and other networks in other organisms.

Consensus reconstruction

As a starting point for the development of a consensus reconstruction, we chose two separately developed freely available metabolic network reconstructions, iMM904 (see http://www.cmb.dtu.dk/Forskning/ Software/models.aspx and http://gcrg.ucsd.edu/In_Silico_Organisms/ Yeast) and iLL672 (ref. 20), containing 904 and 672 yeast genes, respectively. We have also placed relevant files in SBML format on the website http://www.comp-sys-bio.org/yeastnet. Both of these reconstructions were derived from the first genome-scale metabolic network reconstruction for yeast iFF708 (ref. 18; for the basis of this terminology see ref. 26), but the process of curating the original reconstruction was substantially different for the two derived reconstructions. The iMM904 reconstruction has eight different compartments and was developed by curating and expanding an earlier reconstruction, iND750 (ref. 19). In contrast, the iLL672 reconstruction[20] was directly derived from iFF708 by extensively curating the reconstruction to improve the ability of the flux balance model derived from the reconstruction to predict gene deletion phenotypes[27]. It should be noted that yeast metabolic pathways in the Kyoto

Encyclopedia of Genes and Genomes (KEGG)[28] and the Saccharomyces Genome Database (SGD)[29,30] databases were used to establish the starting point for building the original iFF708 reconstruction and also for curating the iLL672 and iMM904 reconstructions. Hence, the information from early versions of these two reconstructions is included implicitly in the consensus reconstruction.

Due to the lack of common metabolite names and annotations, the comparison of the two starting-point reconstructions required first manually defining the correspondences between metabolites. After these had been assigned, the overall metabolite and reaction content of the two reconstructions could be compared (**Table 1**). The majority of metabolites (444) were found in both reconstructions, whereas 8 were found only in iLL672 and 269 only in iMM904. In terms of reactions, 566 were in both reconstructions, 177 were only in iLL672 and 836 only in iMM904. The large number of additional reactions in iMM904 is mostly due to the expanded number of compartments represented in this reconstruction.

The jamboree was held at The University of Manchester, UK, in April 2007. The comparison between the iLL672 and iMM904 reconstructions, proposed at a meeting of the Yeast Systems Biology Network (http://www. ysbn.eu/) in Helsinki, Finland, in June 2006, formed the starting point for the reconstruction (**Table 1**). The three-day event in Manchester concentrated on three separate areas: (i) defining standards for curation as well as for representation of the annotated reconstruction in SBML, (ii) annotating the metabolites with reference to external compound databases and (iii) resolving discrepancies between the reaction-metabolite sets in the two reconstructions. The presence of experts in fields such as yeast genetics and physiology, systems modeling, metabolomics, standards (SBML/ MIRIAM/metabolomics), and database or ontology development allowed the group to make good progress in all three areas. The annotation and curation was aided by a version of the B-Net database[31], and is provided in SBML form (**Supplementary Table 1** online). After the jamboree, a subgroup of the authors verified the curation and annotation, and resolved the remaining discrepancies between models. Below, we discuss some of the major components of the curation and annotation processes.

Metabolite-naming conventions

The initial comparison made it very clear that the naming conventions used in the two models were completely different, such that it was difficult in some cases even for experts to know which chemical entities were being referred to. Moreover, some of the reactions involved 'generic' structures (molecules with R-groups or so-called 'Markush' structures), which are not effectively represented in stoichiometric metabolic models, while certain named entities represented 'composite' substances such as mixtures of different lipids or 'biomass'. Without standardized names, it is extremely hard to enable computer software to reason about the similarities and differences between different models[32–37]. This is even more problematic in the case of reconstructions of the larger human metabolic network[4,38].

However, as SBML allows one to annotate species such as metabolites with external references, we related them to molecules in the 'chemical entities of biological interest' (ChEBI)[39], KEGG[28] and PubChem[40] databases, and identified them precisely using database-independent representations of small molecules, such as 'simplified molecular input line entry system' (SMILES)[41] and international chemical identifier (InChI)[36,42] representations. We took advantage of this aspect of SBML to identify and annotate manually which chemical species were being described. In general, we searched these databases with the contents of the species' name attribute field in the SBML representation or by the chemical formula of the compound sought. The order of annotation was such that we annotated metabolite species using ChEBI identifiers and InChI strings, where possible. If these did not exist or could not be resolved, we used KEGG IDs— or, in two cases, Human Metabolome Database (HMDB)

Table 1 Comparison of starting-point reconstructions

	iMM904	iLL672	Common	iMM904 only	iLL672 only
Metabolites	713	452	444	269	8
Reactions	1,402	743	566[a]	836	177
Genes	904	659	646	258	13
Compartments	8	2	2	6	0

[a]Reaction comparisons were done by considering every reaction to be reversible and without taking into account water and extracellular or intracellular protons (explicitly accounted for in iMM904).

identifiers[43]—followed by PubChem IDs and finally PubMed references. This generated, for the first time, a representation that allows computational comparisons to be performed.

Because some individual molecules have multiple states (e.g., because of acid-base reactions), it would be desirable to use the chemical entities believed to be most common at the pH of the relevant compartment. However, in this version of the consensus reconstruction, all species are assumed to be in the form that corresponds to the most common protonation state at pH 7.2. Whenever possible, the metabolites were annotated with a database entry with the correct protonation state. However, in several cases, the databases only contained the metabolite in a neutral form or otherwise in an incorrect or incorrectly annotated protonation state.

Annotation of large-scale metabolic models in SBML

Although large-scale metabolic network reconstructions and models are now commonly represented in SBML, there has not thus far been a standard way to annotate these models. As part of the consensus reconstruction effort, we tried to develop such a standard that is compliant with MIRIAM[23]. Whereas the annotation of metabolites is quite straightforward, standardized annotation of the reaction content (molecules and reactions) of the reconstructed network proved to be more involved.

Where possible, we annotated reactions using literature references encoded as PubMed IDs, using the MIRIAM- and SBML-compliant "isDescribedBy" 'resource description framework' (RDF; see http://www.w3.org/TR/REC-rdf-syntax/) annotation tag. In addition, reaction annotations include modifiers (enzymes/enzyme complexes) where possible. If a given reaction can be catalyzed by two or more isozymes, we generated an individual reaction for each isozyme (or complex). We represented the formation of protein complexes by separate reactions. Proteins and genes were finally annotated by references to SGD[29] and UniProt[44]. In addition, we annotated cellular compartments using 'Gene Ontology' (GO) terms[45]. In all cases where annotations were used, the MIRIAM[23] web services (http://www.ebi.ac.uk/compneur-srv/miriam-main/mdb?section=ws) were consulted to ensure correct annotation. Examples of fully annotated species and reaction entries are shown in **Figure 1** and in **Supplementary Figure 1** online.

Contents of the consensus reconstructions

In all, the resulting consensus network consists of 2,153 species (1,168 metabolites, 832 genes, 888 proteins and 96 catalytic protein complexes) and 1,857 reactions (1,761 metabolic reactions and 96 complex formation reactions). Reactions and species can be localized to 15 compartments (**Table 2**), including membrane compartments. The network contains 664 distinct chemical entities (e.g., ATP present in the nucleus, cytoplasm, Golgi, mitochondrion, peroxisome and vacuole is classified as one chemi-

a

```
<species metaid="metaid_M_172" id="M_172" name="ATP" compartment="C_1"
sboTerm="SBO:0000299">
  <annotation>
    <in:inchi xmlns:in="http://biomodels.net/inchi" metaid="M_172_inchi">
InChI=1/C10H16N5O13P3/c11-8-5-9(13-2-12-8)15(3-14-5)10-7(17)6
(16)4(26-10)1-25-30(21,22)28-31(23,24)27-29(18,19)20/h2-4,6-7,10,16-17H,1H2,
(H,21,22)(H,23,24)(H2,11,12,13)(H2,18,19,20)/t4-,6-,7-,10-/m1/s1/f/h18-19,21
,23H,11H2
    </in:inchi>
    <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:bqbiol="http://biomodels.net/biology-qualifiers/">
      <rdf:Description rdf:about="#metaid_M_172">
        <bqbiol:is>
          <rdf:Bag>
            <rdf:li rdf:resource="urn:miriam:obo.chebi:CHEBI:15422"/>
            <rdf:li rdf:resource="#M_172_inchi"/>
          </rdf:Bag>
        </bqbiol:is>
      </rdf:Description>
    </rdf:RDF>
  </annotation>
</species>
```

b

Organism: taxonomy id 4932 *Saccharomyces cerevisiae*

β-D-glucose + ATP
CHEBI:15903   CHEBI:15422
SMILES OC[C@H]1O[C@@H](O)[C@H](O)[C@@H](O)[C@@H]1O

Gene name *HXK2*
Genetic locus YGL253W
SGD gene identifier S00003222
Uniprot protein identifier P04807
Reaction (SBOterm 0000176)
is "hexokinase"
is described, e.g., in PubMed paper #6394965
E.C. number 2.7.1.1

β-D-glucose-6-phosphate + ADP + proton
CHEBI: 17719   CHEBI: 16761   CHEBI: 24636
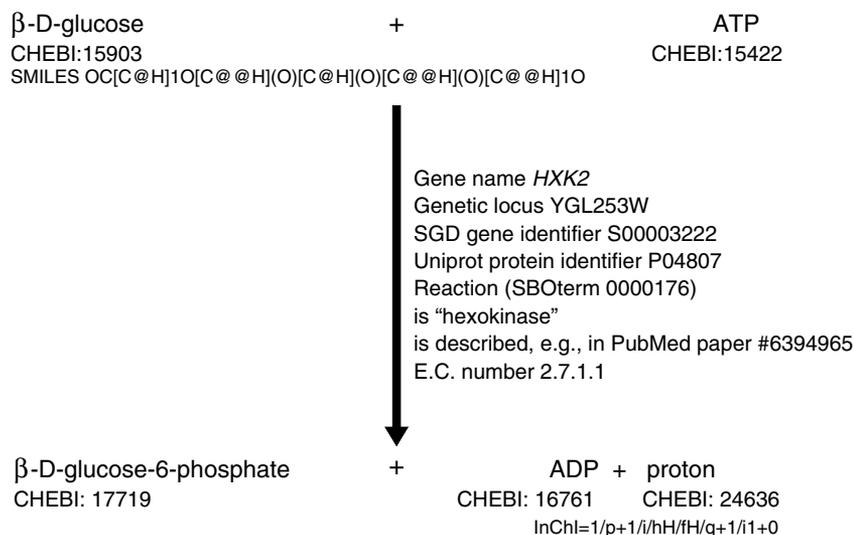   InChI=1/p+1/i/hH/fH/q+1/i1+0

Figure 1 An example of the SBML annotation of a metabolite species using the example of ATP, as used in the reconstruction of the consensus network, illustrating its use of the Systems Biology Ontology (http://www.ebi.ac.uk/sbo/) and its MIRIAM compliance. (**a**) Relevant parts of the SBML code. (**b**) An indication of the kinds of annotations included (for clarity, not all are shown).

cal species and Mg^{2+} liganding is ignored). Of these distinct chemical entities, 554 are annotated with ChEBI identifiers, 564 with InChI identifiers, 78 with KEGG identifiers, 10 with PubChem identifiers, 2 with HMDB identifiers and only 5 with PubMed references. In addition, 26 compounds are currently not annotated in this way. The majority of these are fatty acyl CoAs or acyl carrier proteins where the corresponding fatty acid is in public databases, but the fatty acyl CoA or acyl carrier protein is currently not deposited (but will be submitted to them).

The network includes 1,312 unique chemical transformations, of which 911 occur within a single compartment and the remaining 401 are transport reactions. The overall distribution of metabolites and reactions between the various compartments in the consensus network is given in **Table 2**. Enzyme Commission (EC) number and PubMed reference annotations are provided for 738 and 478 unique

Table 2 Summary of the consensus reconstruction by cellular compartment

Compartment	Reactions	Metabolites
Cytoplasm	835	590
Extracellular	15	158
Golgi	2	13
Mitochondrion	188	235
Nucleus	30	42
Endoplasmic reticulum (ER)	32	28
Vacuole	2	22
Peroxisome	77	80
Mitochondrial membrane	142	0
Plasma membrane	311	0
Peroxisomal membrane	44	0
ER membrane	17	0
Vacuolar membrane	35	0
Golgi membrane	5	0
Nuclear membrane	26	0

transformations in the network, respectively. Each reaction includes all of its cofactors (sometimes known as 'currency metabolites'), such as ATP, NADH and CoA. In addition, although we recognize that there is a certain arbitrariness about this, we have assigned pathway names for each reaction in the network.

We have removed various reactions from the initial networks, especially where they contained Markush structures or ambiguities. This has led to the underrepresentation of lipids, where there are many combinatorial issues[46]. We anticipate that lipid pathways will be added in the future, but 'lipidomics' experiments will eventually be necessary to define the full complement of lipid species present in *S. cerevisiae*. In a similar vein, composite items such as 'biomass' are excluded. Although these are required for flux balance analysis, our purpose here is to provide the basic inventory of metabolites and network structure that
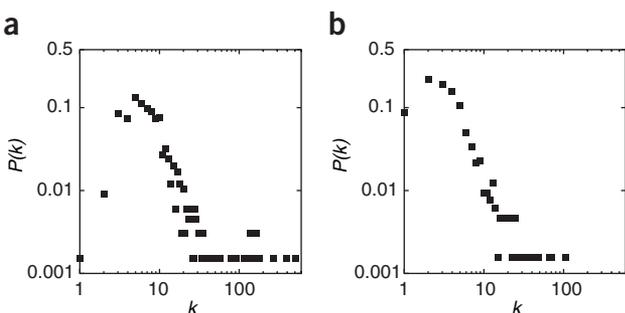
can be used, for example, to compare the network with experimental metabolomics data. This inventory can then form the basis for setting up flux balance models using different assumptions required for setting up these kinds of models, for example, assumptions on the biomass composition, reversibility of reactions and lumping of the reactions into fewer compartments.

Figure 2 depicts the degree distribution[47] of the complete metabolite network, and a version where the currency metabolites were ignored as described earlier[48]. The complete network (**Fig. 2a**) has an average clustering coefficient of 0.742, average node degree of 13.166, characteristic path length of 2.186 and betweenness centralization of 0.3897. The network without currency metabolites (**Fig. 2b**) has an average clustering coefficient of 0.421, average node degree of 5.138, characteristic path length of 4.178 and betweenness centralization of 0.2329. In the full network, the largest value for the shortest distance between any two metabolites ('diameter') is only 4 reaction steps, whereas it is 11 reaction steps (between dTTP and heme A) in the one without 'currency' metabolites. These statistics indicate that the currency metabolites should not be ignored as is sometimes done; without them the network is considerably less connected and several unconnected subnetworks appear, thus leaving some areas of metabolism unconnected from the rest. The center metabolite in the complete network is the proton, whereas in the smaller one it is coenzyme A. **Table 3** lists the top 15 most-connected metabolites of each network.

Dissemination and future curation of the reconstruction

An SBML-encoded version of the base model (with and without compartments) is available at http://www.comp-sys-bio.org/yeastnet. Specifically, the SBML representation of the model is made available under the Creative Commons Attribution-Share Alike 3.0 Unported License (http://www.creativecommons.org/). This is the preferred source for using the complete model with systems biology software. We have tested the SBML using various XML validators, and shown that it loads successfully into the COmplex PAthway SImulator (COPASI)[49] software. COPASI shows that there are 307 mass conservation relations, which were calculated from the stoichiometry matrix using the method of Vallabhajosyula[50], which is now standard in COPASI[49]. We have also loaded the model successfully into some versions of Cytoscape[51] and CellDesigner[52]. The SBML has been checked using libSBML[53,54] (see also http://sbml.org/software/libsbml/).

Recognizing that for many applications only subsets of this model are going to be relevant, we also make it available in an online database that facilitates searching the model. We used the database schema B-Net[31], which already supported all of the features required for our SBML model, including a structured mechanism for MIRIAM annotations. This B-Net representation of the model can be searched using synonyms and it also allows the user to navigate through the network, for example, going from a metabolite to all its reactions, then to the genes that encode the enzymes catalyzing those reactions and so forth. The database is also available at http://www.comp-sys-bio.org/yeastnet.

The B-Net database provides another important function as it is also the preferred means by which the community will be able to edit the model. It will thus be the primary source for the model. As there is no redundancy in the database, any change in any component immediately becomes global. For the time being, editing the model is limited to a few curators to ensure that the current standards are maintained. However, given the major benefits of community annotation[55,56], we have included at the database a mechanism that collects annotations from anyone who wishes to communicate corrections or additions to the model. These annotations will then be reviewed and incorporated into the model for future releases of new versions.

Figure 2 Degree distribution of the metabolic network. The metabolic reaction network was first summarized in a metabolite network, where metabolites are the nodes and one edge links two metabolites that co-occur in a reaction (in any role as substrates or products), as described[48]. For this analysis, transport steps were not considered nor were protein-protein binding reactions. (**a,b**) The figures plot the distribution of the degree of connectivity, $P(k)$, expressed as the fraction of metabolites that have k links out of the total number of metabolites plotted against the number of links (k) in the complete network (**a**) and in a network where the following metabolites were not considered (**b**): {water, proton, carbon dioxide, dioxygen, phosphate3–, diphosphate4–, ammonium, ATP, ADP, AMP, NAD+, NADH, NADP+, NADPH} (to be comparable with the analysis in ref. 48).

DISCUSSION

We have brought together a large segment of the community engaged in research involving genome-scale metabolic networks of yeast to create a consensus network that is freely available without restrictions and that can form the basis for future improvements. The SBML representation of the reconstruction is freely available under a Creative Commons License, and representations of the network were designed to facilitate future improvements.

Although annotation was semi-automated, a considerable element of manual annotation was still required, especially the parsing of the starting models. One of the biggest problems was the use of nonstandard and often arcane synonyms for referring to the same chemical entity. Several commentators have recognized the difficulties caused by synonyms[4,33,38]. For these purposes, we believe and strongly recommend that the best solution to this synonym problem is to reference chemical entities in persistent databases and with database-independent representations such as SMILES[41] and InChI[42]. Referencing the true chemical entity intended requires detailed consideration of its stereochemistry and the anomeric specificity of reactions in which it is involved, and not all databases have the required level of precision. We also recommend that these networks are first built in an assumption-free manner, and that extra features or assumptions that may be required for specific purposes (e.g., adding composite compounds for flux balance analyses) should only then be introduced and annotated. A further benefit of the jamboree approach is the access to experts necessary to annotate details such as the precise gene-protein relationships underlying specific reactions.

We believe the reconstruction presented here is currently the most comprehensive and consistent stoichiometric representation of yeast metabolism, from which predictive (sub)models, for example for genome-scale flux balance analysis, can be extracted and deployed. Presently, the reconstruction lacks information on effectors, reaction kinetics and parametrization. However, the basic framework of B-Net coupled to SBML models that can easily be populated with such data enables these to be added as they become available, and thus kinetic models that can be directly linked to the genome-scale metabolic network can be built. Some parameters are already available at the System for the Analysis of Biochemical Pathways–Reaction Kinetics (SABIO-RK) website (http://sabio.villa-bosch.de/SABIORK/).

Network reconstruction approaches have developed rapidly in recent years. When they reach the genome scale, they can be viewed as systems-level genome annotations[57]. Genome annotation is produced by a community-driven process to reach a consensus annotation that represents the state of knowledge about the genome of the target organism. Annotations are then updated based on new information and they serve as a common denominator for genome science studies of the target organism. The yeast metabolic reconstruction presented here represents an analogous process for systems biology studies of a target organism. With the successful achievement of the first consensus reconstruction, the systems biology community can look forward to similar two-dimensional annotation jamborees for other organisms.

The metabolite nomenclature proposed here will, we hope, become the standard terminology for metabolic models because the compounds themselves are essentially identical in all species. We believe that the semantically annotated reconstruction provided here will have special utility in a number of areas. First is the basic exploration of metabolic pathways and well-curated connections between gene products. Further, the reconstruction will allow the automated interpretation and visualization of metabolomics data as well as data on metabolic proteins, genes and transcripts. The network can form the basis of phenotype predictions, including product yield, in response to genetic and/or environmental perturbations using a variety of methods, including flux balance analysis and logical approaches[58]. It can also be used in metabolic flux estimation based on isotopomer data[59], for filling gaps in metabolic pathways and for exploring questions related to comparative metabolomics[60] and of metabolic pathway evolution. The widespread use of a consensus starting point will make both the comparison and the integration of such studies considerably easier.

Note added in proof: Nookaew et al.[61] have added useful knowledge of some of the lipid metabolism of baker's yeast.

Note: Supplementary information is available on the Nature Biotechnology website.

AUTHOR CONTRIBUTIONS
All authors conceived the idea of the consensus reconstruction, the majority were present during the jamboree itself and all contributed to the writing of, and approved, the manuscript.

Published online at http://www.nature.com/naturebiotechnology/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Kell, D.B. Metabolomics, modelling and machine learning in systems biology: towards an understanding of the languages of cells. The 2005 Theodor Bücher lecture. *FEBS J.* **273**, 873–894 (2006).

Table 3 Most connected nodes in the metabolite network

Complete metabolite network (as in Fig. 2a)			Abbreviated metabolite network (as in Fig. 2b)		
Metabolite	Degree[a]	Betweenness[b]	Metabolite	Degree[a]	Betweenness[b]
Proton	506	0.391	Coenzyme A	106	0.237
Water	390	0.226	L-glutamate^{1-}	71	0.232
ATP	268	0.099	Acetyl-CoA	66	0.065
Diphosphate^{4-}	181	0.062	2-oxoglutarate^{2-}	48	0.078
Phosphate^{3-}	178	0.041	Hydrogen peroxide	45	0.070
ADP	173	0.026	Pyruvate	38	0.070
NADPH	166	0.021	Glycine	31	0.041
NADP^{+}	166	0.021	S-adenosyl-L-methionine	28	0.063
NAD^{+}	143	0.018	Acetate	26	0.021
Carbon dioxide	139	0.020	S-adenosyl-L-homocysteine	25	0.039
NADH	139	0.017	(6S)-5,6,7,8-tetrahydrofolic acid	25	0.031
AMP	128	0.029	L-glutamine	25	0.025
Coenzyme A	119	0.021	Succinate^{2-}	24	0.029
Dioxygen	116	0.014	Acyl-carrier protein	23	0.013
Ammonium	92	0.011	L-cysteine	22	0.023

[a]The number of metabolites that co-occur in metabolic reactions. [b]The betweenness quantifies the number of paths between any two pairs of metabolites in the network that this one mediates (a global property).

2. Arakawa, K., Yamada, Y., Shinoda, K., Nakayama, Y. & Tomita, M. GEM System: automatic prototyping of cell-wide metabolic pathway models from genomes. *BMC Bioinformatics* **7**, 168 (2006).

3. Palsson, B.Ø. S*ystems Biology: Properties of Reconstructed Networks.* (Cambridge University Press, Cambridge; 2006).

4. Duarte, N.C. *et al.* Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. USA* **104**, 1777–1782 (2007).

5. Mager, W.H. & Winderickx, J. Yeast as a model for medical and medicinal research. *Trends Pharmacol. Sci.* **26**, 265–273 (2005).

6. Goffeau, A. *et al.* Life With 6000 genes. *Science* **274**, 546–567 (1996).

7. Winzeler, E.A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).

8. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).

9. Yen, K., Gitsham, P., Wishart, J., Oliver, S.G. & Zhang, N. An improved *tet*O promoter replacement system for regulating the expression of yeast genes. *Yeast* **20**, 1255–1262 (2003).

10. Hughes, T.R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).

11. Allen, J.K. *et al.* High-throughput characterisation of yeast mutants for functional genomics using metabolic footprinting. *Nat. Biotechnol.* **21**, 692–696 (2003).

12. Zhu, H. *et al.* Global analysis of protein activities using proteome chips. *Science* **293**, 2101–2105 (2001).

13. Castrillo, J.I. *et al.* Growth control of the eukaryote cell: a systems biology study in yeast. *J. Biol.* **6**, 4 (2007).

14. Delneri, D. *et al.* Identification and characterization of high-flux-control genes of yeast through competition analyses in continuous cultures. *Nat. Genet.* **40**, 113–117 (2008).

15. Wu, J., Zhang, N., Hayes, A., Panoutsopoulou, K. & Oliver, S.G. Global analysis of nutrient control of gene expression in *Saccharomyces cerevisiae* during growth and starvation. *Proc. Natl. Acad. Sci. USA* **101**, 3148–3153 (2004).

16. Oliver, S. A network approach to the systematic analysis of gene function. *Trends Genet.* **12**, 241–242 (1996).

17. Suter, B., Auerbach, D. & Stagljar, I. Yeast-based functional genomics and proteomics technologies: the first 15 years and beyond. *Biotechniques* **40**, 625–644 (2006).

18. Förster, J., Famili, I., Fu, P., Palsson, B.Ø. & Nielsen, J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* **13**, 244–253 (2003).

19. Duarte, N.C., Herrgard, M.J. & Palsson, B.Ø. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.* **14**, 1298–1309 (2004).

20. Kuepfer, L., Sauer, U. & Blank, L.M. Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Res.* **15**, 1421–1430 (2005).

21. Caspi, R. *et al.* MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* **34**, D511–D516 (2006).

22. Hucka, M. *et al.* The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531 (2003).

23. Le Novère, N. *et al.* Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat. Biotechnol.* **23**, 1509–1515 (2005).

24. Çakir, T. *et al.* Integration of metabolome data with metabolic networks reveals reporter reactions. *Mol. Syst. Biol.* **2**, 50 (2006).

25. Kümmel, A., Panke, S. & Heinemann, M. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Mol. Syst. Biol.* **2**, 2006.0034 (2006).

26. Reed, J.L., Vo, T.D., Schilling, C.H. & Palsson, B.Ø. An expanded genome-scale model of *Escherichia coli* K12 (iJR904 GSM/GPR). *Genome Biol* **4**, R54 (2003).

27. Förster, J., Famili, I., Palsson, B.Ø. & Nielsen, J. Large-scale evaluation of *in silico* deletions in *Saccharomyces cerevisiae*. *OMICS* **7**, 193–202 (2003).

28. Kanehisa, M. *et al.* From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* **34**, D354–D357 (2006).

29. Nash, R. *et al.* Expanded protein information at SGD: new pages and proteome browser. *Nucleic Acids Res.* **35**, D468–D471 (2007).

30. Caspi, R. *et al.* The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* **36**, D623–D631 (2008).

31. Li, X.J. *et al.* in Metabolic profiling: its role in biomarker discovery and gene function analysis. (eds. Harrigan, G.G. & Goodacre, R.) 293–309 (Kluwer Academic Publishers, Boston, 2003).

32. Goble, C. & Wroe, C. The Montagues and the Capulets. *Comp. Funct. Genomics* **5**, 623–632 (2004).

33. Ananiadou, S., Kell, D.B. & Tsujii, J. Text mining and its potential applications in systems biology. *Trends Biotechnol.* **24**, 571–579 (2006).

34. Poolman, M.G., Bonde, B.K., Gevorgyan, A., Patel, H.H. & Fell, D.A. Challenges to be faced in the reconstruction of metabolic networks from public databases. *Syst. Biol. (Stevenage)* **153**, 379–384 (2006).

35. Spasić, I. *et al.* Facilitating the development of controlled vocabularies for metabolomics with text mining. *BMC Bioinformatics* **9**, S5 (2008).

36. Williams, A.J. Internet-based tools for communication and collaboration in chemistry. *Drug Discov. Today* **13**, 502–506 (2008).

37. Williams, A.J. A perspective of publicly accessible/open-access chemistry databases. *Drug Discov. Today* **13**, 495–501 (2008).

38. Ma, H. *et al.* The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol. Syst. Biol.* **3**, 135 (2007).

39. Brooksbank, C., Cameron, G. & Thornton, J. The European Bioinformatics Institute's data resources: towards systems biology. *Nucleic Acids Res.* **33**, D46–D53 (2005).

40. Wheeler, D.L. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **35**, D5–D12 (2007).

41. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **28**, 31–36 (1988).

42. Coles, S.J., Day, N.E., Murray-Rust, P., Rzepa, H.S. & Zhang, Y. Enhancement of the chemical semantic web through the use of InChI identifiers. *Org. Biomol. Chem.* **3**, 1832–1834 (2005).

43. Wishart, D.S. *et al.* HMDB: the Human Metabolome Database. *Nucleic Acids Res.* **35**, D521–D526 (2007).

44. The UniProt Consortium. The universal protein resource (UniProt). *Nucleic Acids Res.* **36**, D190–D195 (2008).

45. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

46. Sud, M. *et al.* LMSD: LIPID MAPS structure database. *Nucleic Acids Res.* **35**, D527–D532 (2007).

47. Barabási, A.-L. & Oltvai, Z.N. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004).

48. Wagner, A. & Fell, D.A. The small world inside large metabolic networks. *Proc. R. Soc. Lond., B, Biol. Sci.* **268**, 1803–1810 (2001).

49. Hoops, S. *et al.* COPASI: a COmplex PAthway SImulator. *Bioinformatics* **22**, 3067–3074 (2006).

50. Vallabhajosyula, R.R., Chickarmane, V. & Sauro, H.M. Conservation analysis of large biochemical networks. *Bioinformatics* **22**, 346–353 (2006).

51. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

52. Funahashi, A., Tanimura, N., Morohashi, M. & Kitano, H. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *BIOSILICO* **1**, 159–162 (2003).

53. Li, P., Oinn, T., Soiland, S. & Kell, D.B. Automated manipulation of systems biology models using libSBML within Taverna workflows. *Bioinformatics* **24**, 287–289 (2008).

54. Bornstein, B.J., Keating, S.M., Jouraku, A. & Hucka, M. LibSBML: an API library for SBML. *Bioinformatics* **24**, 880–881 (2008).

55. Surowiecki, J. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few* (Abacus, London, 2004).

56. Tapscott, D. & Williams, A. *Wikinomics: How Mass Collaboration Changes Everything* (New Paradigm, Toronto, 2007).

57. Palsson, B. Two-dimensional annotation of genomes. *Nat. Biotechnol.* **22**, 1218–1219 (2004).

58. Whelan, K.E. & King, R.D. Using a logical model to predict the growth of yeast. *BMC Bioinformatics* **9**, 97 (2008).

59. Blank, L.M., Kuepfer, L. & Sauer, U. Large-scale ^{13}C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biol.* **6**, R49 (2005).

60. Kell, D.B. Systems biology, metabolic modelling and metabolomics in drug discovery and development. *Drug Discov. Today* **11**, 1085–1092 (2006).

61. Nookaew, I. *et al.* The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. *BMC Syst. Biol.* **2**, 71 (2008).

[1]Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0412, USA. [2]School of Computer Science, The University of Manchester, Oxford Rd., Manchester M13 9PL, UK. [3]The Manchester Centre for Integrative Systems Biology, Manchester Interdisciplinary Biocentre, The University of Manchester, 131 Princess St., Manchester M1 7DN, UK. [4]School of Chemistry, The University of Manchester, Manchester M13 9PL, UK. [5]Department of Chemical Engineering, Boğaziçi University, Bebek 34342, Istanbul, Turkey. [6]VTT Biotechnology Espoo, PO Box 1500, FIN-02044, Finland. [7]School of Chemical Engineering and Analytical Science, The University of Manchester, UK. [8]Max-Planck-Institut für Molekulare Genetik, Ihnestrasse 73, 14195 Berlin, Germany. [9]Institut für Molekulare Systembiologie, ETH Zurich Wolfgang-Pauli-Str. 16, 8093 Zürich, Switzerland. [10]Control and Dynamical Systems, California Institute of Technology, Pasadena, California 91125, USA. [11]Computational Neurobiology, EMBL-EBI, Wellcome-Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. [12]Center for Microbial Biotechnology, Department of Systems Biology, Technical University of Denmark, Building 223, DK-2800 Kgs. Lyngby, Denmark. [13]School of Mathematics, The University of Manchester, Manchester M13 9PL, UK. [14]The Molecular Sciences Institute, 2168 Shattuck Avenue, Berkeley, California 94704, USA. [15]Department of Molecular Cell Physiology, Vrije Universiteit, de Boelelaan 1085, 1081 HV Amsterdam, The Netherlands. [16]Department of Biochemistry, University of Cambridge, Sanger Building, 80 Tennis Court Road, Cambridge CB2 1GA, UK. [17]Virginia Bioinformatics Institute, Virginia Tech, Washington St. 0477, Blacksburg, Virginia 24061, USA. [18]Department of Chemical and Biological Engineering, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden. [19]Present addresses: Synthetic Genomics, Inc., 11149 N. Torrey Pines Rd., La Jolla, California 92037, USA (M.J.H.) and Department of Chemical and Biological Engineering, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden (D.P.). [20]These authors contributed equally to this work. Correspondence should be addressed to D.B.K. (dbk@manchester.ac.uk).