Published online 2 November 2015 *Nucleic Acids Research, 2016, Vol. 44, Database issue* ***D523–D526***
doi: 10.1093/nar/gkv1117

MetaNetX/MNXref – reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks

Sébastien Moretti[1,2], Olivier Martin[1], T. Van Du Tran[1], Alan Bridge[3], Anne Morgat[3,4] and Marco Pagni[1,*]

[1]Vital-IT group, SIB Swiss Institute of Bioinformatics, Lausanne 1015, Switzerland, [2]Department of Ecology and Evolution, Biophore, Evolutionary Bioinformatics group, University of Lausanne, Lausanne 1015, Switzerland, [3]Swiss-Prot Group, SIB Swiss Institute of Bioinformatics, Geneva 1206, Switzerland and [4]Equipe ERABLE, INRIA Grenoble Rhône-Alpes, Montbonnot Saint-Martin 38330, France

Received August 15, 2015; Revised September 25, 2015; Accepted October 11, 2015

ABSTRACT

MetaNetX is a repository of genome-scale metabolic networks (GSMNs) and biochemical pathways from a number of major resources imported into a common *namespace* of chemical compounds, reactions, cellular compartments—namely MNXref—and proteins. The MetaNetX.org website (http://www.metanetx.org/) provides access to these integrated data as well as a variety of tools that allow users to import their own GSMNs, map them to the MNXref reconciliation, and manipulate, compare, analyze, simulate (using flux balance analysis) and export the resulting GSMNs. MNXref and MetaNetX are regularly updated and freely available.

INTRODUCTION

A genome-scale metabolic network (GSMN), or stoichiometric model, describes the set of biochemical reactions which may occur in a given organism, as well as the requisite enzymes, and may also include information on sub-cellular compartments, transport reactions and transporters. By design GSMNs are focused on the metabolism of small molecular weight compounds when energy and mass conservation law can be applied, and are not suited to represent gene regulation or signaling pathways. In practice, a GSMN has a double purpose, as it is both a repository of knowledge about an organism's metabolism, and a model that can be simulated, using flux balance analysis (FBA). Such simulations can address different questions: (i) establish the essentiality of genes in specific growth conditions; (ii) reveal opportunities for metabolic engineering and optimization; (iii) suggest new drug targets (1). To permit simulations, a GSMN usually includes artificial reactions that describe the growth medium, a growth equation (which implies the composition of the biomass) and possibly hypothetical reactions not (yet) supported by experimental biology but required to make a model functional.

A relatively small number of high quality GSMNs have been published to date, essentially for model organisms, and are made available by a few dedicated databases (2–6). The development of such models requires significant human effort and curation, and the fully automated reconstruction of a GSMN from an annotated genome sequence remains a challenge (7,8). Such methods require the integration of high quality curated data covering the known biochemistry of a vast range of organisms, as well as methods that address the specific requirements of a functional GSMN, including the elemental balancing of individual reactions. These considerations form the major motivation for the development of the resource presented here.

MNXREF RECONCILIATION

The metabolite identifiers found in the early-published GSMNs were often specific to the individual groups developing and curating them, and did not generally reference the major databases of chemical compounds. In recent years there have been a few attempts to 'reconcile' the different nomenclatures of these compounds (9,10) including our own effort MNXref (11). The principles of the reconciliation algorithm used in MNXref can be summarized as follows:

(i) Reconciliation of common metabolites based on chemical structures;
(ii) Reconciliation of metabolites through shared chemical nomenclature;
(iii) Reconciliation of reactions through shared metabolites;
(iv) Identification of candidate reactions for reconciliation through shared cross-references;

*To whom correspondence should be addressed. Tel: +41 21 692 40 38; Fax: +41 21 692 40 55; Email: marco.pagni@isb-sib.ch

Table 1. Numbers of reconciled metabolites (a) and reactions (b) in MNXref 2.0, and mapped proteins (c), found in common between published GSMNs and major biochemical databases in MetaNetX.org

	MNXref	**BiGG** (2) 18 GSMNs	**BioCyc** (3) 19 GSMNs	**Path2Models** (4) 132 GSMNs	**The SEED** (18) 50 GSMNs	**YeastNet** (6) 1 GSMN
(a) Metabolites						
BiGG[a] (2) (version 2beta)	4039	3414	2610	2836	1829	1021
BioPath (19) (2010–05–03)	1313	649	875	943	567	427
ChEBI (20) (version 131)	46 477	8507	15 631	17 973	7108	4416
HMDB (21) (version 3.6)	42 542	1292	2525	3044	1054	714
KEGG (22) (version 75.1)	28 429	1958	3945	5356	1560	908
LIPIDMAPS (23) (2015–06–28)	40 719	412	1382	1587	280	252
MetaCyc (3) (version 19.1)	15 472	1835	5380	5637	1399	826
Reactome (24) (2015–07–13)	4576	1799	2539	2770	1467	1521
The SEED[a] (18) (2013–06–19)	16 280	2040	3120	4098	1551	678
UMBBD–EAWAG (12) (2014–06–30)	1395	206	347	588	150	67
UniPathway (25) (version 2015_03)	1113	692	874	928	657	393
(b) Reactions						
BiGG[a] (2) (version 2beta)	11 458	6055	3380	2580	1876	1730
BioPath (19) (2010–05–03)	1545	456	684	725	328	285
KEGG (22) (version 75.1)	9925	1335	3085	4309	877	528
MetaCyc (3) (version 19.1)	13 793	1419	5040	4220	828	549
Reactome (24) (2015–07–13)	23 592	4111	5848	5147	2849	2604
Rhea (16) (version 64)	32 256	5101	10 603	10 293	3190	2050
The SEED[a] (18) (2013–06–19)	13 260	3069	2980	3337	1738	932
UniPathway (25) (version 2015_03)	1994	1065	1435	1471	836	559
(c) Proteins						
UniProt (15) (version 2015_08)		11 142	15 670	76 293	27 154	912

[a]BiGG and The SEED distribute collections of metabolites and reactions that are not necessarily retrieved in one of their GSMN.

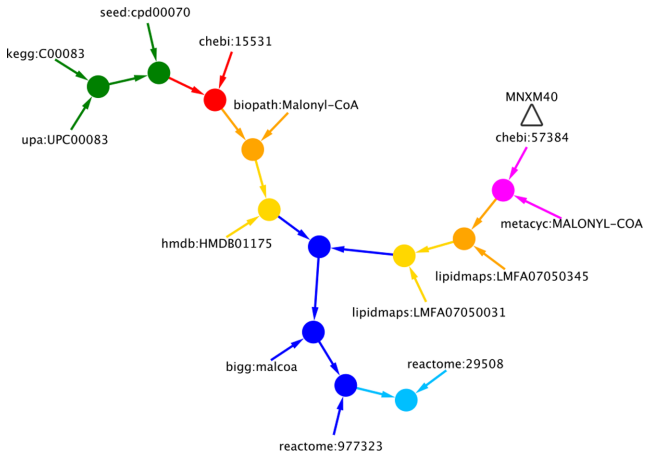

Figure 1. Evidences used to reconcile different chemical compounds for the metabolite malonyl-CoA (*MNXM40* in MNXref): magenta, using structure supplied by the source databases; red, using recomputed structure; orange, recomputed structure protonated at pH 7.3; yellow, recomputed structure protonated at pH 7.3 but ignoring the stereo layer of the InChI representation; green, using the cross-references supplied by the source databases; dark blue, based on compound primary names; light blue, based on compound synonym names. Triangle: in a post-processing step, chebi:57384 was chosen to best represent the targeted metabolite.

(v) Iterative reconciliation of metabolites through reaction context.

Figure 1 illustrates the reconciliation process using malonyl-CoA as an example; individual steps in the reconciliation are color-coded according to the type of evidence used. Table 1 summarizes the overlaps between the various sources of biochemical data and GSMNs according to the results of the MNXref reconciliation. The MNXref namespace is regularly updated with metabolite and reaction data from new resources; recent additions include the EAWAG-BBD/UMBBD pathway database (12).

In the construction and use of GSMNs, every reaction must be balanced with respect to elemental composition and charge; failure to balance reactions will lead to violations in mass conservation that can have detrimental effects on the downstream simulations. The case of protons is worthy of particular attention in this regard. Protons provide a means to balance chemical equations occurring in aqueous solution, but they are also responsible for creating membrane potentials whose dissipation is a major driving force in cell metabolism. In order to distinguish these two roles we have introduced separate identifiers for those protons transported across a membrane (*MNXM01* in MNXref), and those protons introduced for the purposes of balancing a reaction (*MNXM1* in MNXref). An artificial spontaneous reaction is then added to every compartment of the GSMN to permit the free exchange between transported and balanced protons (*MNXR01* in MNXref). In this way, the original properties of the GSMN are preserved.

METANETX REPOSITORY AND TOOLS

MetaNetX.org (13) is a website that provides free access to the MNXref reconciliation data and a collection of published GSMNs and biochemical pathways mapped onto MNXref. The website also allows users to upload, manipulate, analyze or modify their own GSMNs and export them in SBML or in our own tab-delimited format. MetaNetX.org also offers a selection of tools for analyses including network structure, FBA or nested pattern methods (14).

Gene names have been widely used in published GSMNs to describe the protein complexes that act as enzymes or transporters. Gene nomenclature is, however, essentially organism specific, if not dependent on a particular genome assembly. In MetaNetX we use UniProt accession numbers (15) to identify gene products: it greatly facilitates the inter organisms comparison of GSMNs from different sources.

Although the MNXref reconciliation algorithm is essentially automated, the compilation of the MetaNetX repository requires some manual intervention and a certain number of editorial choices. This includes definition of an accepted list of species and strains that includes important model organisms. Preference is given to the most comprehensive GSMNs from external sources that use accepted standard formats and have sufficient protein coverage (full acknowledgement is given to these external sources). We are closely collaborating with Rhea (16), which is a database of manually curated biochemical reactions, as part of the ongoing effort to further improve the quality of annotation of our resource.

CONCLUSION

The www.metanetx.org resource provides a comprehensive suite of tools for the analysis of genome-scale metabolic models, based on a single integrated namespace of metabolites and metabolic reactions that integrates the most widely used biochemical databases and model repositories – MNXref. The reconciliation process used in MNXref greatly simplifies the development and analysis of genome-scale metabolic models, allowing users to concentrate on model analysis rather than the time-consuming problem of identifier mapping. Future developments will include the provision of tools and the integration of new resources such as the SwissLipids knowledgebase (17), which provides lipid structures and curated data on enzymatic reactions.

ACKNOWLEDGEMENTS

Computation and maintenance of the MetaNetX.org server are provided by the Vital-IT center for high-performance computing of the SIB Swiss Institute of Bioinformatics (http://www.vital-it.ch). We thank Ioannis Xenarios and Joerg Stelling for support and feedback.

FUNDING

Swiss Initiative for Systems Biology [SystemsX.ch projects MetaNetX, HostPathX and SyBIT] evaluated by the Swiss National Science Foundation; Swiss Federal Government through the Federal Office of Education and Science. Funding for open access charge: SIB Swiss Institute of Bioinformatics.

Conflict of interest statement. None declared.

REFERENCES

1. Bordbar,A., Monk,J.M., King,Z.A. and Palsson,B.O. (2014) Constraint-based models predict metabolic and associated cellular functions. *Nat. Rev. Genet.*, **15**, 107–120.
2. Schellenberger,J., Park,J.O., Conrad,T.M. and Palsson,B.O. (2010) BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, **11**, 213.
3. Caspi,R., Altman,T., Billington,R., Dreher,K., Foerster,H., Fulcher,C.A., Holland,T.A., Keseler,I.M., Kothari,A., Kubo,A. *et al.* (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucl. Acids Res.*, **42**, D459–D471.
4. Büchel,F., Rodriguez,N., Swainston,N., Wrzodek,C., Czauderna,T., Keller,R., Mittag,F., Schubert,M., Glont,M., Golebiewski,M. *et al.* (2013) Path2Models: large-scale generation of computational models from biochemical pathway maps. *BMC Syst. Biol.*, **7**, 116.
5. Henry,C.S., DeJongh,M., Best,A.A., Frybarger,P.M., Linsay,B. and Stevens,R.L. (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotechnol.*, **28**, 977–982.
6. Kim,H., Shin,J., Kim,E., Kim,H., Hwang,S., Shim,J.E. and Lee,I. (2014) YeastNet v3: a public database of data-specific and integrated functional gene networks for Saccharomyces cerevisiae. *Nucleic Acids Res.*, **42**, D731–D736.
7. Aziz,R.K., Bartels,D., Best,A.A., DeJongh,M., Disz,T., Edwards,R.A., Formsma,K., Gerdes,S., Glass,E.M., Kubal,M. *et al.* (2008) The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*, **9**, 75.
8. Karp,P.D., Paley,S.M., Krummenacker,M., Latendresse,M., Dale,J.M., Lee,T.J., Kaipa,P., Gilham,F., Spaulding,A., Popescu,L. *et al.* (2010) Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Brief Bioinform.*, **11**, 40–79.
9. Lang,M., Stelzer,M. and Schomburg,D. (2011) BKM-react, an integrated biochemical reaction database. *BMC Biochem.*, **12**, 42.
10. Kumar,A., Suthers,P.F. and Maranas,C.D. (2012) MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases. *BMC Bioinformatics*, **13**, 6.
11. Bernard,T., Bridge,A., Morgat,A., Moretti,S., Xenarios,I. and Pagni,M. (2014) Reconciliation of metabolites and biochemical reactions for metabolic networks. *Brief Bioinform.*, **15**, 123–135.
12. Gao,J., Ellis,L.B.M. and Wackett,L.P. (2010) The University of Minnesota Biocatalysis/Biodegradation Database: improving public access. *Nucleic Acids Res.*, **38**, D488–D491.
13. Ganter,M., Bernard,T., Moretti,S., Stelling,J. and Pagni,M. (2013) MetaNetX.org: a website and repository for accessing, analysing and manipulating metabolic networks. *Bioinformatics*, **29**, 815–816.
14. Ganter,M., Kaltenbach,H.-M. and Stelling,J. (2014) Predicting network functions with nested patterns. *Nat. Commun.*, **5**, 3006.
15. UniProt Consortium. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.
16. Morgat,A., Axelsen,K.B., Lombardot,T., Alcántara,R., Aimo,L., Zerara,M., Niknejad,A., Belda,E., Hyka-Nouspikel,N., Coudert,E. *et al.* (2015) Updates in Rhea - a manually curated resource of biochemical reactions. *Nucleic Acids Res.*, **43**, D459–D464.
17. Aimo,L., Liechti,R., Hyka-Nouspikel,N., Niknejad,A., Gleizes,A., Götz,L., Kuznetsov,D., David,F.P.A., van der Goot,F.G., Riezman,H. *et al.* (2015) The SwissLipids knowledgebase for lipid biology. *Bioinformatics*, **31**, 2860–2866.
18. Overbeek,R., Olson,R., Pusch,G.D., Olsen,G.J., Davis,J.J., Disz,T., Edwards,R.A., Gerdes,S., Parrello,B., Shukla,M. *et al.* (2014) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.*, **42**, D206–D214.
19. Forster,M., Pick,A., Raitner,M., Schreiber,F. and Brandenburg,F.J. (2002) The system architecture of the BioPath system. *In Silico Biol. (Gedrukt)*, **2**, 415–426.
20. Hastings,J., de Matos,P., Dekker,A., Ennis,M., Harsha,B., Kale,N., Muthukrishnan,V., Owen,G., Turner,S., Williams,M. *et al.* (2013) The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Res.*, **41**, D456–D463.
21. Wishart,D.S., Jewison,T., Guo,A.C., Wilson,M., Knox,C., Liu,Y., Djoumbou,Y., Mandal,R., Aziat,F., Dong,E. *et al.* (2013) HMDB 3.0 - The Human Metabolome Database in 2013. *Nucleic Acids Res.*, **41**, D801–D807.
22. Kanehisa,M., Goto,S., Sato,Y., Kawashima,M., Furumichi,M. and Tanabe,M. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.*, **42**, D199–D205.
23. Sud,M., Fahy,E., Cotter,D., Dennis,E.A. and Subramaniam,S. (2012) LIPID MAPS-nature lipidomics gateway: an online resource for

students and educators interested in lipids. *J. Chem. Educ.*, **89**, 291–292.
24. Croft,D., Mundo,A.F., Haw,R., Milacic,M., Weiser,J., Wu,G., Caudy,M., Garapati,P., Gillespie,M., Kamdar,M.R. *et al.* (2014) The Reactome pathway knowledgebase. *Nucleic Acids Res.*, **42**, D472–D477.
25. Morgat,A., Coissac,E., Coudert,E., Axelsen,K.B., Keller,G., Bairoch,A., Bridge,A., Bougueleret,L., Xenarios,I. and Viari,A. (2012) UniPathway: a resource for the exploration and annotation of metabolic pathways. *Nucleic Acids Res.*, **40**, D761–D769.