Database update

MEMOSys 2.0: an update of the bioinformatics database for genome-scale models and genomic data

Stephan Pabinger[1,2,*], Rene Snajder[1,3], Timo Hardiman[4], Michaela Willi[1,3], Andreas Dander[1,3] and Zlatko Trajanoski[1]

[1]Division for Bioinformatics, Innsbruck Medical University, 6020 Innsbruck, Austria, [2]Health & Environment Department, AIT–Austrian Institute of Technology, Molecular Diagnostics, 1190 Vienna, Austria, [3]Oncotyrol, Center for Personalized Cancer Medicine, 6020 Innsbruck, Austria and [4]Development Anti-Infectives Microbiology, Sandoz GmbH, 6250 Kundl, Austria

*Corresponding author: Tel: +43 512 9003 71400; Fax: +43 512 9003 73401; Email: stephan.pabinger@gmail.com

The MEtabolic MOdel research and development System (MEMOSys) is a versatile database for the management, storage and development of genome-scale models (GEMs). Since its initial release, the database has undergone major improvements, and the new version introduces several new features. First, the novel concept of derived models allows users to create model hierarchies that automatically propagate modifications along their order. Second, all stored components can now be easily enhanced with additional annotations that can be directly extracted from a supplied Systems Biology Markup Language (SBML) file. Third, the web application has been substantially revised and now features new query mechanisms, an easy search system for reactions and new link-out services to publicly available databases. Fourth, the updated database now contains 20 publicly available models, which can be easily exported into standardized formats for further analysis. Fifth, MEMOSys 2.0 is now also available as a fully configured virtual image and can be found online at http://www.icbi.at/memosys and http://memoys.i-med.ac.at.

Database URL: http://memosys.i-med.ac.at

Introduction

The extensive use of next-generation sequencing (NGS) technology in biological research has spurred the creation of genome-scale metabolic models for a multitude of organisms (1). Recently, new versions of metabolic reconstruction for the model organism yeast (2) and the human organism (3) were published and widely discussed in the research community.

Genome-scale models (GEMs) can be basically described as a network of metabolites that are connected by reactions occurring in the living organism. GEMs integrate various omics data types (bibliomics, metabolomics, proteomics) in a structured format, and can be used to perform computational and quantitative queries to answer various questions about the capabilities of the investigated organisms. Applications for GEMs are manifold, including biological discovery (4), flux analysis (5), gene deletion studies (6) and metabolic engineering (7). Furthermore, they can be used to provide an alternative context from which experimental data can be interpreted (4).

Over the past years, bacteria, fungi, as well as other organisms have been engineered to increase the yield of industrial enzymes and vitamins, as they are ideal hosts for the production of recombinant proteins. To further explore the characteristics of these important industrial micro-organisms, GEMs have become more and more relevant for biotechnology to get a comprehensive understanding of its metabolism (8). The rapid development of methods

for obtaining high-throughput data has now made it possible to study human metabolism on a genome-wide scale. GEMs provide a useful way to investigate numerous omics data types and address various research questions. Recently, human metabolic models have been used to study cancer metabolism (9) and predict potential drug targets and biomarkers (10).

The generation of new models is a well-defined iterative process comprising a multitude of different steps (11) where usually several intermediate revisions are generated (12). Owing to the vast amount of existing models, it is now reasonable to assume that new GEMs might be based on existing models of closely related organisms. During the construction process, the newly created GEM is compared with existing networks, and simulated results are constantly verified with experimental data. Therefore, it is of great importance to (i) have easy access to software for creating and manipulating models, (ii) be able to review all previous modifications, (iii) extract previous versions and (iv) export GEMs into a standardized format. Furthermore, the use of existing models for creating new GEMs might lead to bias where components of the novel organisms are underrepresented. Therefore, the ability to unequivocally compare models is of great help when designing new models.

To deal with these challenges, we previously developed the MEtabolic MOdel research and development System (MEMOSys) to support the construction, modification and management of GEMs (13). The web-based bioinformatics database stores all properties of a metabolic model and offers a powerful search system, a feature-rich comparison mechanism and standardized references to external databases. It uses a custom-version control system to automatically store the complete developmental history of all model components, which enables researchers to query and download all intermediate versions of a GEM. Furthermore, the database supports exchanging of models in the Systems Biology Markup Language (SBML) format.

Here, we report the second major release of MEMOSys, which introduces several updates to the database. The system has undergone a substantial redesign of the interface and business functionality and has been enhanced with a new feature to create derived models. Furthermore, new publicly available models were integrated into the database, and the complete MEMOSys system is now available as a fully configured virtual machine image. We believe that this major update to MEMOSys results in a sophisticated database that will facilitate the creation and investigation of GEMs by researchers worldwide.

Results

The architecture and development platform of the database MEMOSys have been retained in this update as described in the original publication (13). However, the system has undergone significant changes to implement new functionality and features (the new database schema is depicted in Supplementary Figure S1). During the development of the database update, substantial effort has been devoted to testing the system and setting up a continuous integration test environment. Thereby, all functionality tests are executed for each added feature, which ensures that new functionality is compatible with the previous state.

An important cornerstone of MEMOSys is the compliance with the SBML. Import and export of models are now based on the native JAVA library JSBML (14), improving the overall performance of the system and considerably facilitating the installation of MEMOSys 2.0.

Extension as a research database

As MEMOSys has proven to be a valuable research platform, we included several published GEMs into the database. The system now hosts 20 publicly available models (Table 1), which can be easily exported for further analysis. Original SBML models were modified and extended to be compatible with previously included model components. Among others, we have included the recently released models for *Saccharomyces cerevisiae* and *Homo sapiens*, as well as the model organisms *Mus musculus*, *Arabidopsis thaliana* and *Escherichia coli.*

Derived model creation

During the construction and refinement of GEMs, it is often necessary to analyze models that deviate only in a few particular features. Therefore, MEMOSys 2.0 offers now the possibility to create derived models out of existing models without specifying a limit on the number of derived models. Each derived model is in essence an exact copy of the parent model, where the model itself and every reaction contain a reference to the parent. As soon as a property in the parent model is modified, MEMOSys implements a mechanism to propagate changes to all derived models. If a child reaction is modified, it loses its reference to the parent reaction and is now treated as an independent reaction. Figure 1 depicts the described scenarios.

Derived models are shown together with their parent model in the model list section of MEMOSys 2.0 (Figure 2). The system graphically depicts the hierarchical level of each derived model and lists the name of the respective parent model. To account for the changed data structure, the derived model history functionality was extended to include the modification history of the parent model. MEMOSys 2.0 is capable of listing all changes of a model, including its hierarchical dependencies. The model comparison mechanism, as introduced in the previous release, has been enhanced to allow the comparison of different derived models. Comparison result for reactions, metabolites and genes is either displayed in lists or graphically presented.

Table 1. Listed are publicly available GEMs in the MEMOSys database

Model	Organism	Author	Reference
iLF583	*Amycolatopsis balhimycina*	Vongsangnak *et al.,* 2012	(15)
Arabidopsis	*Arabidopsis thaliana*	Mintz-Oron *et al.,* 2012	(16)
iHD666	*Aspergillus nidulans*	David *et al.,* 2008	(17)
iMA871	*Aspergillus niger*	Andersen *et al.,* 2008	(18)
iWV1314	*Aspergillus oryzae*	Vongsangnak *et al.,* 2008	(19)
iMZ1055	*Bacillus megaterium*	Zou *et al.,* 2013	(8)
iJO1366	*Escherichia coli*	Orth *et al.,* 2011	(20)
Recon1	*Homo sapiens*	Duarte *et al.,* 2006	(21)
Recon2	*Homo sapiens*	Thiele *et al.,* 2013	(3)
Llactis	*Lactococcus lactis*	Oliveira *et al.,* 2005	(22)
mmu_08	*Mus musculus*	Queck *et al.,* 2008	(23)
iAL1006	*Penicillium chrysogenum*	Agren *et al.,* 2013	(24)
iLC915	*Pichia pastoris*	Caspeta *et al.,* 2012	(25)
iSS884	*Pichia stipitis*	Caspeta *et al.,* 2012	(25)
iSB1139	*Pseudomonas fluorescens*	Borgos *et al.,* 2013	(26)
iFF708	*Saccharomyces cerevisiae*	Förster *et al.,* 2003	(27)
iIN800	*Saccharomyces cerevisiae*	Nookaew *et al.,* 2008	(28)
ymn2_0	*Saccharomyces cerevisiae*	Herrgard *et al.,* 2008	(29)
ymn7_0	*Saccharomyces cerevisiae*	Aung *et al.,* 2013	(2)
iIB711	*Streptomyces coelicolor*	Borodina *et al.,* 2005	(30)

Component annotation

GEMs rely heavily on annotations to unambiguously identify model components. Therefore, MEMOSys supports annotating reactions, metabolites, genes and compartments with references to external databases using the 'minimum information requested in the annotation of biochemical models' (MIRIAM) (31) notation. To avoid redundancy and improve comparability and re-usability, MEMOSys stores only unique metabolites and genes in the database. As these components can be used by multiple reactions, updates or annotation changes are automatically visible for all attached reactions. To enhance these components with additional annotations, the newest version of MEMOSys allows users to load SBML files into the system to extract and store these annotations. Files are uploaded to the database using the file upload functionality. The parsing process can be started by dedicated users and after successfully loading the file, the system displays the annotations in a user-friendly way. Next, the user can select which annotations should be kept and stored in the database (Figure 3).

Interface improvements

The user interface of MEMOSys 2.0 underwent a major overhaul. In general, unnecessary information was removed from the main pages, and frequently used features were put in more prominent places. The new version of MEMOSys provides a simple interface that allows users to directly browse to specific reactions using either its defined identifier or its KEGG ID. The improved model import mechanism displays several key figures (such as number of imported reactions, number of balanced/unbalanced reactions) after the process has finished providing the user with a quick overview of model characteristics. Another new feature allows enhancing reactions with citation information, where corresponding fields are automatically populated using valid PubMed IDs. The query mechanism for reactions was enhanced to enable searching for KEGG IDs, reactants and genes. To improve readability, the versioning for models has been redesigned, which is now consistently implemented throughout the whole database. Moreover, MEMOSys 2.0 offers a new BioCyc linking service using gene and organism IDs to create references to corresponding BioCyc Web sites. Based on the organism name, the system provides suggestions of matching IDs facilitating the assignment of correct organism IDs.

Distribution

Two critical factors that contribute, among others, to the success of a bioinformatics software suite are easy installation (32) and the availability of a detailed user guide. Since its first release, the source code of MEMOSys has been open-source, and we provided step-by-step installation and usage instructions. The database can be installed on
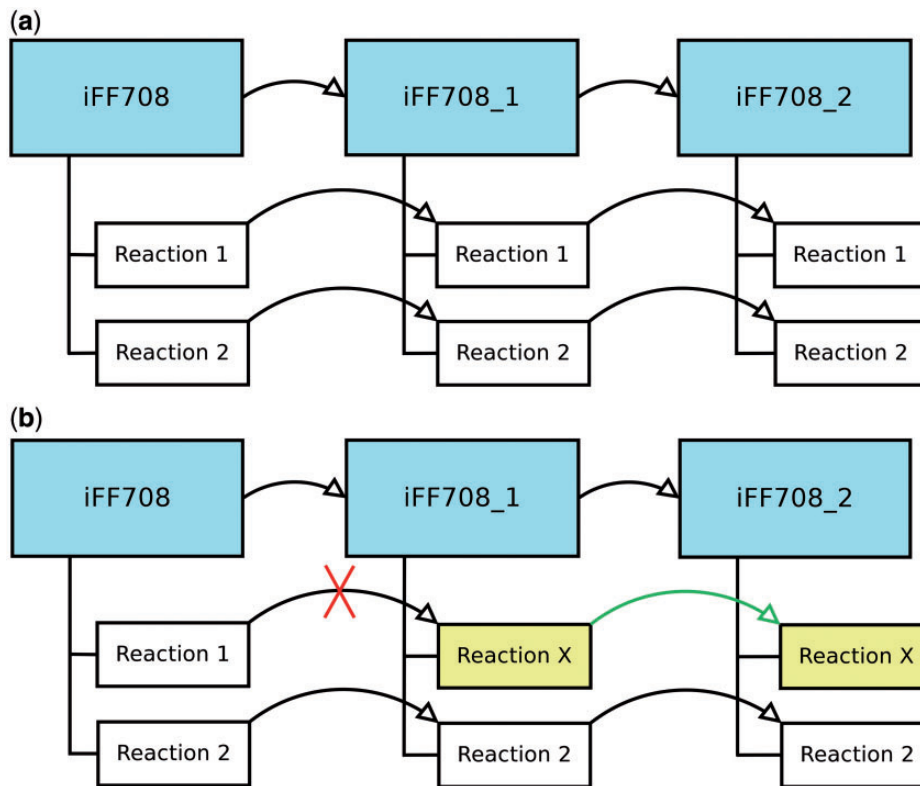
Figure 1. (a) Displays three GEMs where model iFF708_1 is a derived model of iFF708, and iFF708_2 is a derived model of iFF708_1. Model iFF708_1 and iFF708_2 are copies of model iFF708 and contain exact copies of all reactions present in model iFF708. (b) Shows a use case where Reaction 1 was modified in model iFF708_1 (its name was changed to *reaction X*) resulting in the loss of the reference to Reaction 1, and in the update of Reaction 1 in model iFF708_2. If Reaction 1 in model iFF708 is updated, reaction X will not be changed in model iFF708_1 and model iFF708_2.

Figure 2. Depicted is the model list of MEMOSys. The first column graphically displays the hierarchical level of the model followed by various model properties. New derived models can be created by clicking on the icon displayed in the rightmost column.

Linux and Mac, and since its version 2.0, also on Windows. Additionally, we now offer MEMOSys 2.0 as a ready-to-use virtual machine image including a fully configured system. The virtual machine image contains a PostgreSQL database, a JBoss application server, the MEMOSys system, as well as a web-based user-management application. Furthermore, the complete *Penicillium chrysogenum* model is already stored in the database, which contains numerous annotated components that can be used as a starting point for future GEM constructions.

Discussion

Version 2.0 is a significant update to the MEMOSys database designed for the management, development and storage of metabolic models. The new derived model functionality allows researchers to build hierarchical relationships of models and quickly compare different versions of one GEM. Moreover, the provided propagation feature ensures that component updates are pushed to all derived models, which avoids inconsistency and cumbersome repetitive work.

Figure 3. Displayed is the user interface for enhancing the annotation of metabolites and genes. After loading an SBML file, the system identifies new or different annotations and offers the possibility to select the correct annotation for each component. Furthermore, the user can choose to replace all current component annotations with the newly loaded ones, or store only new component annotations while keeping the currently existing annotations.

It has been shown that the development of new models is facilitated by using previously inserted well-annotated components (11). Consequently, MEMOSys 2.0 includes a new annotation improvement functionality offering a simple way to add new annotations to already stored components. Owing to the flexible data design of MEMOSys, additional references to the existing databases can be easily added to all stored components. The inclusion of new models further increases the value of MEMOSys as a research system, where researchers can easily export and query one of the 20 existing GEMs. The authors of the database will continually improve the annotation of publicly available models, and new models will be included upon request.

In contrast to existing systems such as the Biochemical Genetic and Genomic (BiGG) database (33) or the BioModels database (34), MEMOSys support researchers during the reconstruction of novel metabolic models by providing an automatic auditing system. Furthermore, the database offers a rich web-based editing functionality for all components and features a comparison mechanism to quickly get an overview of model differences.

The distribution of MEMOSys 2.0 as a virtual machine will make it accessible to more users because it removes the burden of manual installation. Furthermore, the ability of virtual machines to reset states, save snapshots or export appliances presents tremendous opportunities for collaborative GEM development and model analysis. Additional analysis software can be installed directly into the virtual machine, and researchers can easily share the complete appliance including current and previous states of models by simply exporting it.

Since its initial release, the interface of MEMOSys has undergone several changes that addressed usability issues and design deficiencies. The improvements were based on long-term user feedback and include new query mechanisms, link-out services and page refinements. We believe that MEMOSys 2.0 will continue to be a useful tool for the research community, and the new update will facilitate the creation of new models and allow users to effectively explore and analyze existing GEMs.

Supplementary Data

Supplementary data are available at *Database* Online.

Acknowledgements

Funding

References

1. Henry,C.S., DeJongh,M., Best,A.A. *et al*. (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotechnol.*, **28**, 977–982.

2. Aung,H.W., Henry,S.A. and Walker,L.P. (2013) Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism. *Ind. Biotechnol.*, **9**, 215–228.

3. Thiele,I., Swainston,N., Fleming,R.M.T. *et al*. (2013) A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.*, **31**, 419–425.

4. McCloskey,D., Palsson,B.Ø. and Feist,A.M. (2013) Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. *Mol. Syst. Biol.*, **9**, 661.

5. Orth,J.D., Thiele,I. and Palsson,B.Ø. (2010) What is flux balance analysis? *Nat. Biotechnol.*, **28**, 245–248.

6. Choi,H.S., Lee,S.Y., Kim,T.Y. *et al*. (2010) In silico identification of gene amplification targets for improvement of lycopene production. *Appl. Environ. Microbiol.*, **76**, 3097–3105.

7. Yen,J.Y., Nazem-Bokaee,H., Freedman,B.G. *et al*. (2013) Deriving metabolic engineering strategies from genome-scale modeling with flux ratio constraints. *Biotechnol. J.*, **8**, 581–594.

8. Zou,W., Zhou,M., Liu,L. *et al*. (2013) Reconstruction and analysis of the industrial strain *Bacillus megaterium* WSH002 genome-scale in silico metabolic model. *J. Biotechnol.*, **164**, 503–509.

9. Hu,J., Locasale,J.W., Bielas,J.H. *et al*. (2013) Heterogeneity of tumor-induced gene expression changes in the human metabolic network. *Nat. Biotechnol.*, **31**, 522–529.

10. Jerby,L. and Ruppin,E. (2012) Predicting drug targets and biomarkers of cancer via genome-scale metabolic modeling. *Clin. Cancer Res.*, **18**, 5572–5584.

11. Thiele,I. and Palsson,B.Ø. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.*, **5**, 93–121.

12. Pabinger,S. and Trajanoski,Z. (2013) Genome-scale model management and comparison. *Methods Mol. Biol.*, **985**, 3–16.

13. Pabinger,S., Rader,R., Agren,R. *et al*. (2011) MEMOSys: bioinformatics platform for genome-scale metabolic models. *BMC Syst. Biol.*, **5**, 20.

14. Dräger,A., Rodriguez,N., Dumousseau,M. *et al*. (2011) JSBML: a flexible Java library for working with SBML. *Bioinformatics*, **27**, 2167–2168.

15. Vongsangnak,W., Figueiredo,L.F., Förster,J. *et al*. (2012) Genome-scale metabolic representation of *Amycolatopsis balhimycina*. *Biotechnol. Bioeng.*, **109**, 1798–1807.

16. Mintz-Oron,S., Meir,S., Malitsky,S. *et al*. (2012) Reconstruction of *Arabidopsis* metabolic network models accounting for subcellular compartmentalization and tissue-specificity. *Proc. Natl. Acad. Sci. USA*, **109**, 339–344.

17. David,H., Ozçelik,I.S., Hofmann,G. *et al*. (2008) Analysis of *Aspergillus nidulans* metabolism at the genome-scale. *BMC Genomics*, **9**, 163.

18. Andersen,M.R., Nielsen,M.L. and Nielsen,J. (2008) Metabolic model integration of the bibliome, genome, metabolome and reactome of *Aspergillus niger*. *Mol. Syst. Biol.*, **4**, 178.

19. Vongsangnak,W., Olsen,P., Hansen,K. *et al*. (2008) Improved annotation through genome-scale metabolic modeling of *Aspergillus oryzae*. *BMC Genomics*, **9**, 245.

20. Orth,J.D., Conrad,T.M., Na,J. *et al*. (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism 2011. *Mol. Syst. Biol.*, **7**, 535.

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

22. Oliveira,A.P., Nielsen,J. and Förster,J. (2005) Modeling *Lactococcus lactis* using a genome-scale flux model. *BMC Microbiol.*, **5**, 39.

23. Quek,L.-E. and Nielsen,L.K. (2008) On the reconstruction of the *Mus musculus* genome-scale metabolic network model. *Genome Inform.*, **21**, 89–100.

24. Agren,R., Liu,L., Shoaie,S. *et al*. (2013) The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. *PLoS Comput. Biol.*, **9**, e1002980.

25. Caspeta,L., Shoaie,S., Agren,R. *et al*. (2012) Genome-scale metabolic reconstructions of *Pichia stipitis* and *Pichia pastoris* and in silico evaluation of their potentials. *BMC Syst. Biol.*, **6**, 24.

26. Borgos,S.E.F., Bordel,S., Sletta,H. *et al*. (2013) Mapping global effects of the anti-sigma factor MucA in *Pseudomonas fluorescens* SBW25 through genome-scale metabolic modeling. *BMC Syst. Biol.*, **7**, 19.

27. Förster,J., Famili,I., Fu,P. *et al*. (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.*, **13**, 244–253.

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

29. Herrgård,M.J., Swainston,N., Dobson,P. *et al*. (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat. Biotechnol.*, **26**, 1155–1160.

30. Borodina,I., Krabben,P. and Nielsen,J. (2005) Genome-scale analysis of *Streptomyces coelicolor* A3(2) metabolism. *Genome Res.*, **15**, 820–829.

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

32. Nocq,J., Celton,M., Gendron,P. *et al*. (2013) Harnessing virtual machines to simplify next-generation DNA sequencing analysis. *Bioinformatics*, **29**, 2075–2083.

33. Schellenberger,J., Park,J.O., Conrad,T.M. *et al*. (2010) BiGG: a biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, **11**, 213.

34. Li,C., Donizelli,M., Rodriguez,N. *et al*. (2010) BioModels database: an enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Syst. Biol.*, **4**, 92.