Novel Developments of the MetaCrop Information System for Facilitating Systems Biological Approaches

Klaus Hippe[1], Christian Colmsee[1], Tobias Czauderna[1], Eva Grafahrend-Belau[1], Björn H. Junker[1], Christian Klukas[1], Uwe Scholz[1], Falk Schreiber[1,2] and Stephan Weise[1*]

[1]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Corrensstr. 3, 06466 Gatersleben, Germany

[2]Institute of Computer Science, Martin Luther University Halle-Wittenberg, Von-Seckendorff-Platz 1, 06120 Halle, Germany

Summary

Crop plants play a major role in human and animal nutrition and increasingly contribute to chemical or pharmaceutical industry and renewable resources. In order to achieve important goals, such as the improvement of growth or yield, it is indispensable to understand biological processes on a detailed level. Therefore, the well-structured management of fine-grained information about metabolic pathways is of high interest. Thus, we developed the MetaCrop information system, a manually curated repository of high quality information concerning the metabolism of crop plants. However, the data access to and flexible export of information of MetaCrop in standard exchange formats had to be improved. To automate and accelerate the data access we designed a set of web services to be integrated into external software. These web services have already been used by an add-on for the visualisation toolkit VANTED. Furthermore, we developed an export feature for the MetaCrop web interface, thus enabling the user to compose individual metabolic models using SBML.

1 Introduction

Crop plants are a main source for human and animal nutrition. Furthermore, they increasingly contribute to chemical or pharmaceutical industry and renewable resources [1, 2, 3, 4]. In order to advance the comprehension of complex biological processes occurring in crop plants on a detailed level (e.g. for improvement of growth or yield), it is of high interest to reconstruct detailed metabolic models. Therefore, detailed metabolic information up to cell compartment level needs to be collected and managed in a well-structured way. On this account we developed the MetaCrop information system [5].

MetaCrop is a manually curated repository of high quality information concerning the metabolism of seven major crop plants with high agronomical importance and two model plants widely used in plant research. This comprises pathway diagrams, reactions, locations, transport processes, reaction kinetics, taxonomy and literature (Tab. 1). The web interface, available at `http://metacrop.ipk-gatersleben.de`, supports an easy exploration of the information from overview pathways to single reactions and therefore helps users to understand the metabolism of crop plants.

*To whom correspondence should be addressed. Email: weise@ipk-gatersleben.de

Table 1: Overview of information contained in MetaCrop

Organism	Pathways	Reactions	Transporters	Compartments	References
Hordeum vulgare	36	293	8	5	391
Triticum aestivum	34	273	6	5	365
Oryza sativa	35	281	9	4	355
Zea mays	35	276	27	5	392
Solanum tuberosum	34	211	14	3	272
Brassica napus	32	170	7	4	206
Beta vulgaris	35	228	—	4	366
Arabidopsis thaliana	35	289	15	4	525
Medicago truncatula	34	239	—	4	337
total	38	394	60	5	1100

Together with every conversion element (reaction or translocation) managed in MetaCrop, fine-grained information can be stored. This includes reaction or translocation type, formula, synonyms, EC numbers, literature references and also kinetic data, such as V_{max} values, affinity or inhibitor constants. Moreover, all this information is assigned to the compartment the respective element is located in, thus considering that conversions take place at different locations inside an organism depending on the developmental state and environmental effects. Due to different quality levels, data stored in MetaCrop can be complemented with quality tags. Furthermore, it is possible to store pathway information in different parallel versions. We reckon that the time-consuming process of manual data curation is indispensable since there are only a few models for crop plants existing at all.

The main application of MetaCrop is to support mathematical modelling of metabolism, thus allowing to analyse the structure, dynamics and behaviour of metabolic networks. This enables to verify and to extend the understanding of complex processes as well as the generation of new hypotheses and the identification of suitable targets for metabolic engineering by performing *in silico* experiments. MetaCrop contains the necessary data in a consolidated manner and supports the process of model reconstruction by different tools. Subject to the available data, different model analysis techniques can be applied to the metabolic models, such as purely stoichiometric or kinetic approaches, respectively. Since there are only a few models for crop plants existing at all MetaCrop is of high importance for this area of research.

For automating and accelerating the access to the MetaCrop data we developed a set of web services. The availability of web services ensures a widespread accessibility of the data and they are also platform independent. This approach offers many advantages for bioinformatics, which have already been demonstrated in the open source project BioMOBY [6]. These web services can be used by any application. We developed a new add-on for the pathway visualisation toolkit VANTED (Visualization and Analysis of Networks containing Experimental Data) [7] which uses these web services. To enable simulation and analysis of metabolic pathways, e. g. with software toolkits such as COPASI [8] or FBASimVis [9], it is necessary to compose individual network models. Therefore, we developed an export functionality applying the Systems Biology Markup Language (SBML) [10].

This paper is structured as follows. First, we describe the design and implementation of the web services intended to provide an automated access to the MetaCrop system. Second, we focus on the development of a VANTED add-on for pathway visualisation and third, we describe the

composition of individual metabolic network models using SBML.

2 Web Services

All relevant information of MetaCrop is available via web services. For each of the five categories of MetaCrop data (pathway, conversion, substance, taxonomy and publication), an own web service was developed comprising several methods. Thus, it is easy to integrate new methods for a specific category and the administration of each web service remains simple. The various methods of every web service conform one of the following categories:

- all information about a single data set,
- all data sets of a MetaCrop category and
- search by a keyword on a specific attribute.

Each of the five web services possesses a *getAll*-method returning all data sets of the certain category. With the parameter *limit* the result set can be reduced. The method *getPathway* of the pathway web service returns data in Systems Biology Graphical Notation (SBGN) [11] or SBML format, respectively. This data can be used with appropriate applications such as VANTED. Table 2 illustrates the MetaCrop web services and their methods.

Table 2: Overview of the methods of the MetaCrop web services

Web service	Method	Input	Output
pathway	getPathway	p. ID	p. ID, p. name, SBML, SBGN, conversion
	getAllPathways	limit	p. ID, p. name
	searchPathway	p. name	p. ID, p. name
conversion	getConversion	c. ID	c. ID, c. name, reversible, catalysed, formula, pathway, substance, location, values
	getAllConversions	limit	c. ID, c. name, reversible, catalysed
	searchEcNumber	EC number	c. ID, c. name, reversible, catalysed
	searchConversion	c. name	c. ID, c. name, reversible, catalysed
substance	getSubstance	s. ID	s. ID, s. name, type, molecular mass, formula, CAS number, EC number, conversion, pathway, location
	getAllSubstances	limit	s. ID, s. name, type
	searchCasNumber	CAS number	s. ID, s. name, type
	searchSubstance	s. name	s. ID, s. name, type

To be continued on next page

Continuation of Tab. 2

Web service	Method	Input	Output
taxonomy	getTaxonomy	t. ID	t. ID, NCBI ID, species, cultivar, synonyms, class, order, family, subfamily, tribe, genus, conversion, pathway, publication
	getAllTaxonomy	limit	t. ID, NCBI ID, species, cultivar
	searchCultivar	cultivar name	t. ID, NCBI ID, species, cultivar
	searchNcbiId	NCBI ID	t. ID, NCBI ID, species, cultivar
	searchSpecies	species name	t. ID, NCBI ID, species, cultivar
publication	getPublication	p. ID	p. ID, PubMed ID, p. name, type, substance, conversion
	getAllPublications	limit	p. ID, PubMed ID, p. name, type
	searchPublicationType	p. type	p. ID, PubMed ID, p. name, type
	searchPubmedId	PubMed ID	p. ID, PubMed ID, p. name, type
	searchPublication	p. name	p. ID, PubMed ID, p. name, type

The web services were developed by means of the open source framework Apache Axis2 v1.4.1 and Java v1.4 using the Application Server OracleAS v10gR2. Apache Axis2 supports the web service standards SOAP (Simple Object Access Protocol), WSDL (Web Service Description Language), REST (Representational State Transfer) and MTOM (SOAP Message Transmission Optimisation Mechanism). Furthermore, it is possible to extend this framework with additional modules, so that more functionalities are available. One of these modules is Apache Rampart. By this extension the Web Services Security standard is available in Apache Axis2. Particularly, the usability, performance and flexibility are key factors of Apache Axis2. Thus, the MetaCrop web services were developed with this framework. The communication protocol SOAP is the main engine of these web services. Figure 1 illustrates the system architecture and the functionality of the MetaCrop web services.

Connections to the MetaCrop database are realised by a connection pool, thus enabling many clients to use the web services simultaneously. Also, this approach consumes little system resources only. Interactions between applications and web services are realised by service clients, which can be generated automatically by the Apache Axis2 framework. All methods essential for the interaction are available in these service clients. Thus, it gets very convenient for every software to get information about any MetaCrop category or any data set of the MetaCrop database.

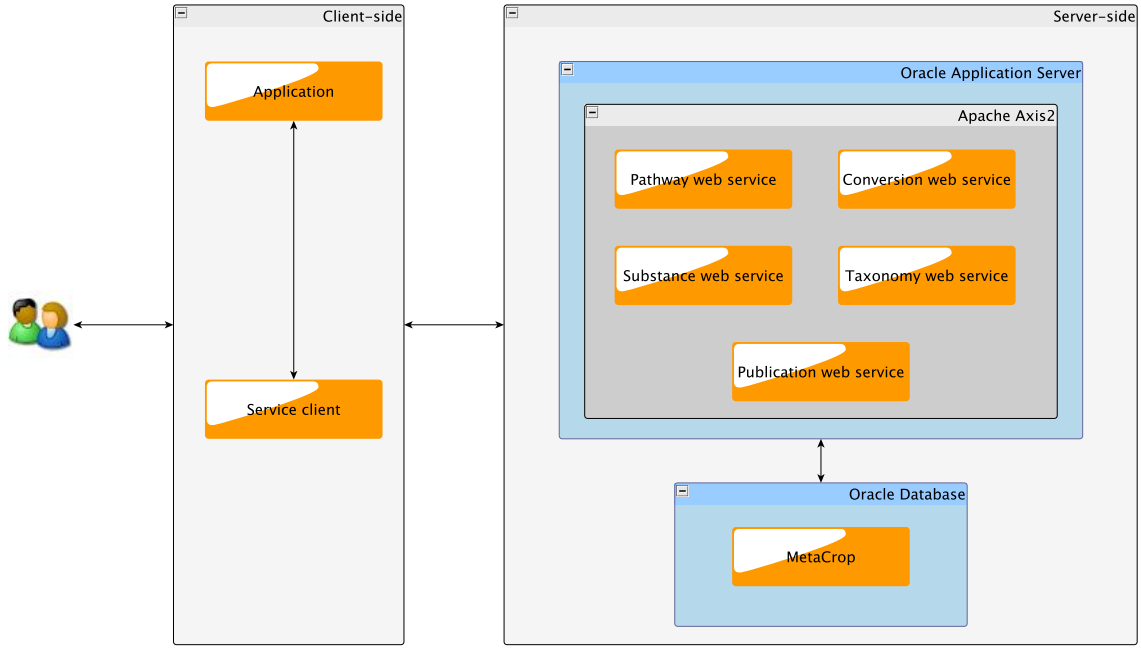

Figure 1: Architecture overview of the MetaCrop web services, which is based on a client-server architecture and provides the corresponding functionalities. On server-side it comprises an Oracle Application Server and the MetaCrop database. This application server is the runtime environment for Apache Axis2, which is essential for the MetaCrop web services. The interaction between the users and the MetaCrop web services is enabled by the client-side consisting of an application and a generated service client.

The MetaCrop web services can be accessed at `http://metacrop.ipk-gatersleben.de/metacrop_ws`.

3 VANTED Add-on

In order to use the MetaCrop data in conjunction with the visualisation toolkit VANTED an add-on was developed applying the web services described above. VANTED is an application for the visualisation and analysis of networks with related experimental data. Also, data from large-scale biochemical experiments can be mapped onto a network that is drawn with VANTED itself, downloaded from the KEGG [12] pathway database or imported using standard network exchange formats, respectively. VANTED is available at `http://vanted.ipk-gatersleben.de`. Figure 2 shows a screenshot of the MetaCrop add-on.

Any category of MetaCrop has got an own area within this add-on. Thus, it is very convenient for the user to change between the different categories. The single area has the following four components:

- search component,
- control component,
- search results and
- detailed information about a single entry.

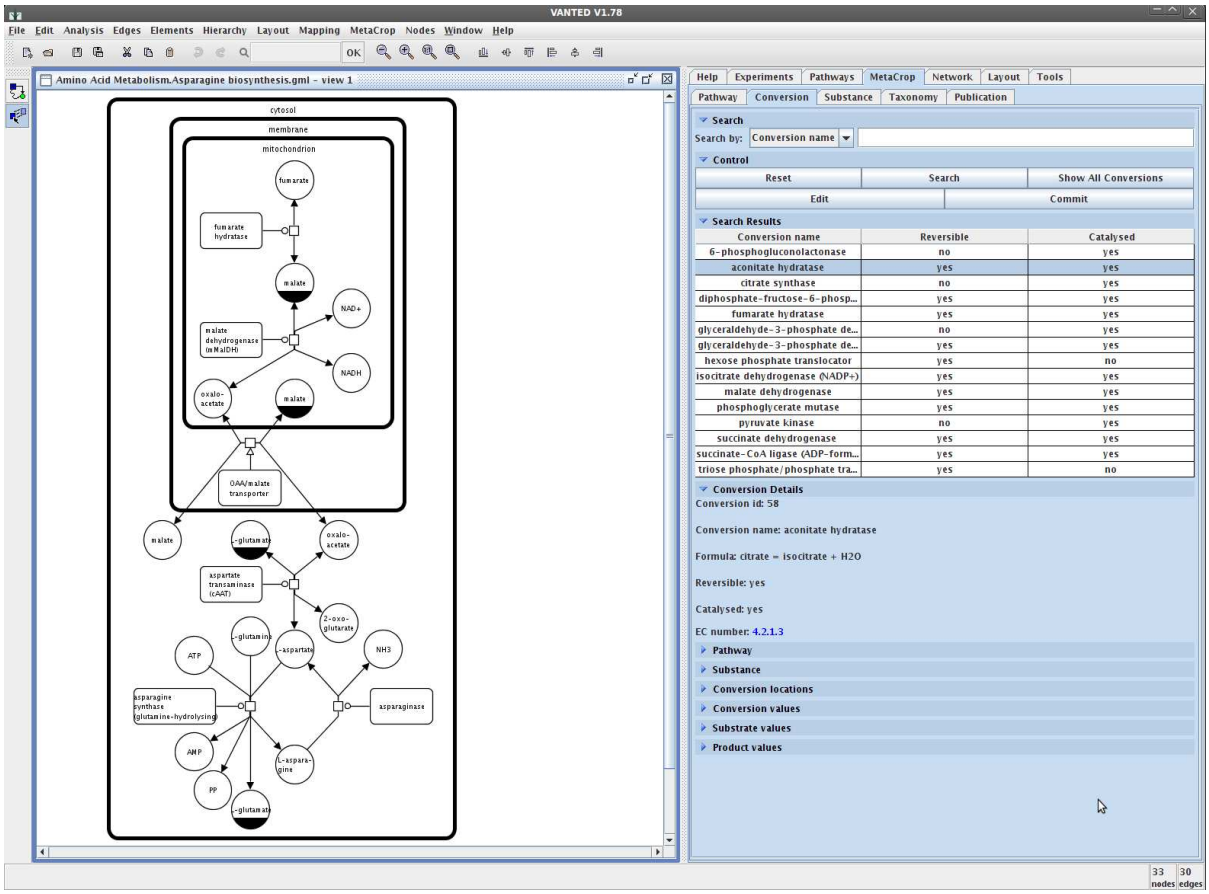

Figure 2: MetaCrop add-on for VANTED. Through this expansion all MetaCrop functions and data are available in VANTED. The right side shows the main area of the add-on with their options and on the left side the pathway is shown as a graph using SBGN notation.

The respective components use one of the web services of MetaCrop. Thus, for a user it is quite simple to interact with the system. The search components of the relevant areas have got several search options. Table 3 lists each category and their search options currently available via the MetaCrop add-on.

Table 3: Overview of information contained in MetaCrop

Category	Search option
pathway	name
conversion	name, EC number
substance	name, CAS number
taxonomy	cultivar, NCBI ID, species
publication	publication, publication type, PubMed ID

Furthermore, this add-on identifies and displays the relationships between the different cate-

gories and their biological elements or processes. The example in Fig. 2 shows the data set for the metabolic pathway *Asparagine biosynthesis*.

All available information and relations of this pathway will be illustrated on the right side. The example pathway entry in Fig. 2 has got several conversions. When a user selects a certain conversion, all information about the selected entry will be requested and shown in their category. Hence, it is simple for the user to get more information about a conversion which is involved in the pathway. On the left side, pathways are shown in SBGN notation. Therefore, the method *getPathway* from the pathway web service returns the necessary information and the add-on then generates the graph using SBGN notation. Via the add-on the user gets access to all available data of MetaCrop and can use them for other operations with VANTED. Also it is possible to query MetaCrop for specific information. Thereby, the work is considerably simplified and the user has got more possibilities using VANTED.

4 SBML Exporter

In order to perform analysis and simulation with external software toolkits it is often necessary to compose individual metabolic models. Such models are helpful in interpreting the dynamic behaviour of specific metabolic pathways. Therefore, we developed a new export feature using the standardised SBML format, which is integrated into the MetaCrop web interface. This feature uses the principle of a shopping cart, thus enabling the user to put elements such as certain reactions, translocations or even whole pathways into the cart.

Applying a wizard, in the next step the user is able to modify several default parameter settings, such as multiple value handling, compartment, restrictions to data from certain species only or export file name. The following steps of the wizard comprise lists of reactions and translocations contained in the selected pathways (Fig. 3). Now, the user can modify the kinetic model or the compartment of the respective elements.

The generated SBML file contains function definitions, all involved species, annotations and the reactions including their kinetic data. Additionally, it is possible to export structural models only.

The MetaCrop SBML exporter is available via the web interface at `http://metacrop.ipk-gatersleben.de`.

5 Conclusion

In this paper we presented recent developments in order to improve the MetaCrop information system for facilitating systems biology. For using the MetaCrop data in conjunction with other software environments, e. g. for simulation or visualisation, there exist two alternatives. First, we developed a set of web services intended to provide automated access to MetaCrop. As a proof of concept we described the MetaCrop add-on for the visualisation toolkit VANTED. Second, to compose individual metabolic models interactively, we enhanced the MetaCrop web interface by a feature using the standardised SBML format to export user-defined pathways and models.

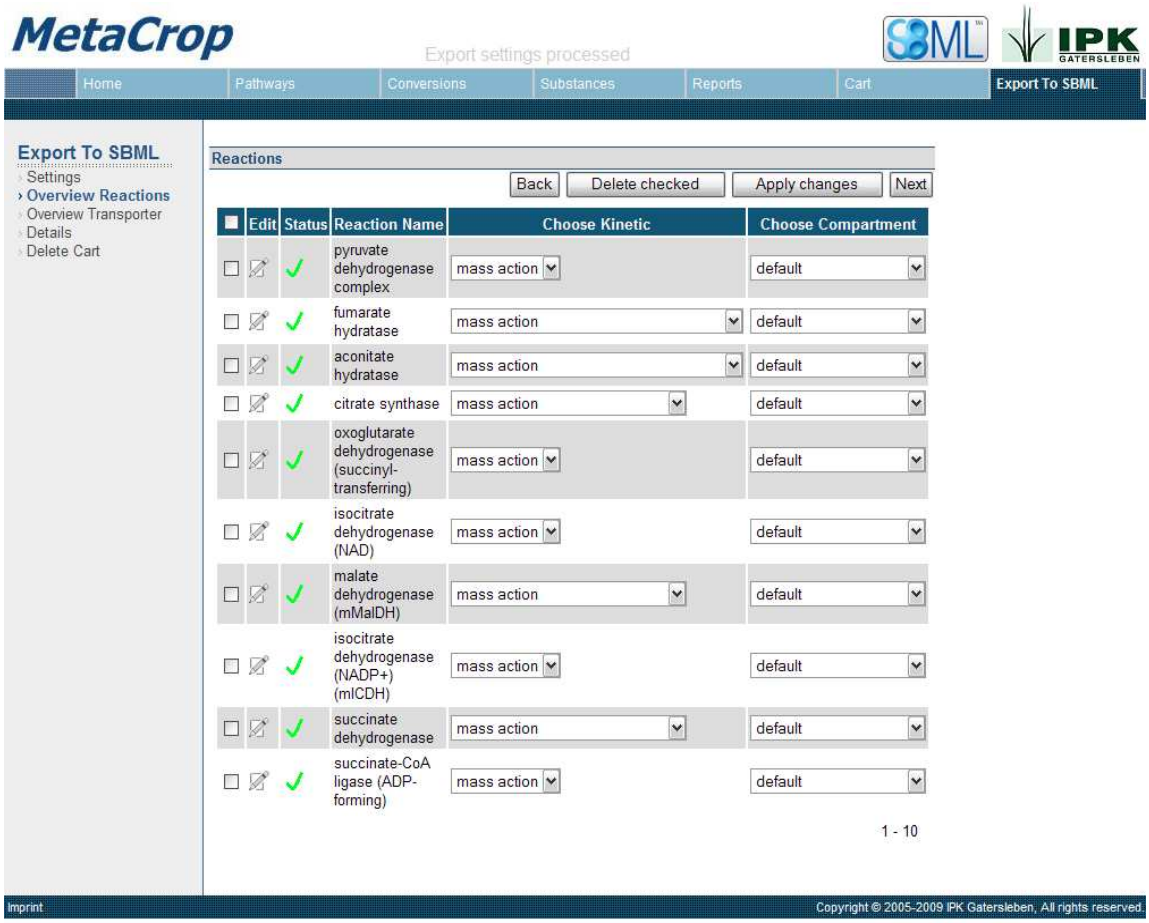

Figure 3: Screenshot from the MetaCrop SBML exporter. Here, a list of reactions chosen for export is shown enabling the user to modify parameters such as the kinetic type or the compartment a reaction takes place at.

References

[1] D. Tilman, J. Hill, and C. Lehman. Carbon-Negative Biofuels from Low-Input High-Diversity Grassland Biomass. *Science*, 314(5805):1598–1600, 2006.

[2] J.O. Metzger and U. Bornscheuer. Lipids as renewable resources: current state of chemical and biotechnological conversion and diversification. *Applied Microbiology and Biotechnology*, 71(1):13–22, 2006.

[3] J. Scheller, K. H. Guhrs, F. Grosse, and U. Conrad. Production of spider silk proteins in tobacco and potato. *Nature Biotechnology*, 19(6):573–577, 2001.

[4] D. DellaPenna. Nutritional Genomics: Manipulating Plant Micronutrients to Improve Human Health. *Science*, 285(5426):375–379, 1999.

[5] E. Grafahrend-Belau, S. Weise, D. Koschützki, U. Scholz, B.H. Junker, and F. Schreiber. MetaCrop: a detailed database of crop plant metabolism. *Nucleic Acids Research*, 36(Database issue):D954–D958, 2008.

[6] M. D. Wilkinson and M. Links. BioMOBY: An open source biological web services proposal. *Briefings In Bioinformatics*, 3(4):331–341, 2002.

[7] B. Junker, C. Klukas, and F. Schreiber. VANTED: A system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics*, 7(1):e109.1–13, 2006.

[8] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. COPASI–a COmplex PAthway SImulator. *Bioinformatics*, 22(24):3067–3074, 2006.

[9] E. Grafahrend-Belau, C. Klukas, B.H. Junker, and F. Schreiber. FBA-SimVis: interactive visualization of constraint-based metabolic models. *Bioinformatics*, 25(20):2755–2757, 2009.

[10] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, II Goryanin, W. J. Hedley, T. C. Hodgman, J. H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novère, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, and J. Wang. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003.

[11] N. Le Novère, M. Hucka, H. Mi, S. Moodie, F. Schreiber, A. Sorokin, E. Demir, K. Wegner, M.I. Aladjem, S.M. Wimalaratne, F.T. Bergman, R. Gauges, P. Ghaza, H. Kawaji, L. Li, Y. Matsuoka, A. Villéger, S.E. Boyd, L. Calzone, M. Courtot, U. Dogrusoz, T.C. Freeman, A. Funahashi, S. Ghosh, A. Jouraku, S. Kim, F. Kolpakov, A. Luna, S. Sahle, E. Schmidt, S. Watterson, G. Wu, I. Goryanin, D.B. Kell, C. Sander, H. Sauro, J.L. Snoep, K. Kohn, and H. Kitano. The Systems Biology Graphical Notation. *Nature Biotechnology*, 27(8):735–741, 2009.

[12] M. Kanehisa, S. Goto, M. Hattori, K. F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, and M. Hirakawa. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Research*, 34(Database issue):D354–D357, 2006.