

## Creating a Bridge between Modelica and the Systems Biology Community

Jan Brugård<sup>1</sup>, Daniel Hedberg<sup>1</sup>, Marta Cascante<sup>2</sup>, Gunnar Cedersund<sup>3,4</sup>, Àlex Gómez-Garrido<sup>5</sup>, Dieter Maier<sup>6</sup>, Elin Nyman<sup>3</sup>, Vitaly Selivanov<sup>2</sup>, Peter Strålfors<sup>3</sup>

<sup>1</sup> MathCore Engineering AB, Teknikringen 1F, 583 30 Linköping, Sweden  
{jan.brugard, daniel.hedberg}@mathcore.com

<sup>2</sup> Departamento de Bioquímica y Biología Molecular, Universitat de Barcelona, Spain

<sup>3</sup> Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden

<sup>4</sup> Freiburg Institute for Advanced Studies, School of Life Sciences, Germany

<sup>5</sup> Grup de Recerca en Informàtica Biomèdica, Universitat Pompeu Fabra, Barcelona, Spain

<sup>6</sup> Biomax Informatics AG, Martinsried, Germany

### Abstract

The Systems Biology Markup Language (SBML) is the leading modelling language within systems biology. It is a computer-readable format for representing models of biochemical reaction networks in software. SBML has been evolving since 2000 thanks to an international community of software developers and users. At the same time the Modelica language has evolved as the leading object-oriented modelling language for convenient, component-oriented modelling of complex physical systems.

As a part of the EC-funded BioBridge project MathCore has developed the Modelica library BioChem and the *MathModelica Systems Biology* toolbox. With the release of version 1.0 of the BioChem library and the *MathModelica Systems Biology* toolbox it will be possible to import and export from Modelica to SBML. The toolbox also allows to publish the Modelica models as interactive HTML pages, with biochemical and experimental enzyme characterizations derived from the BioBridge portal and BioXM knowledge management environment.

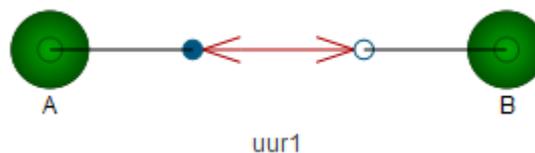
*Keywords:* Modelica tools; SBML; library; biological systems; cellular pathways; translator

## 1 The BioChem Library

The first beta version of the BioChem library was developed by Larsdotter Nilsson at the Linköping University [1].

The design idea behind the BioChem library is to create a general purpose Modelica library for modeling, simulation and visualization of biological and biochemical systems. The models implemented in

the BioChem library describe substances and reactions that can take place in-between these substances in a diverse number of biochemical pathways. An example pathway is shown in Figure 1.



**Figure 1** Simple pathway model using components from the BioChem library.

Version 1.0 of the library has now been completed and is publicly available [2].

## 2 Modelica to SBML Translator

A translator that converts Modelica to SBML [3], and vice versa, has been developed. The translator is able to convert SBML models that are compliant with SBML Level 2, version 3. The Modelica models created are Modelica 2.2 and 3 compliant. Modelica models are translated to SBML Level 2, version 1, 2 or 3.

### 2.1 SBML

SBML is based on XML, making it more suitable to read and write by computers than by humans. SBML is a model representation format for systems biology, created with an objective to become a common intermediate format for software tools. The idea is that a large support for SBML enables the use of a wide range of different tools without having to rewrite models. In principle, a SBML model consists of a number of components, describing biochemical enti-

ties (species) and transformations (reactions and rules) forming a biochemical network. In addition it is also possible to describe instantaneous discontinuous state changes using events, and operating assumptions for the model using constraints.

## 2.2 Design principles

Whereas SBML is a very specific language for describing models within the systems biology field only, Modelica is a much more general and expressive language. One of the challenges in writing a translator lies in finding a way to map different language constructs and features between the two languages. There is no obvious one-to-one mapping as there are usually many different ways of describing things in Modelica. One way to solve this is to try and catch all possible constructs in Modelica and map them to SBML. Another way is to restrict how Modelica models may be constructed in order to reduce the number of possibilities.

In *MathModelica* a combination is used. The BioChem library provides the user with the building blocks that must be used in order for a model to be exportable to SBML. *MathModelica* also provides wizards to help the user build custom components, such as reactions and compartments. These wizards are interactive step-by-step guides, serving two purposes, guiding the user through complex operations and helping the user create models that will be exportable to SBML.

## 2.3 Verification of the translator

The translator has been tested by importing all curated models from the BioModels database [4] release 14. Table 1 shows the percentage of models that (a) are possible to import to *MathModelica* without failure, (b) give the same simulation result as the reference simulation, and (c) give the same simulation result after exporting to SBML.

**Table 1** Verification of import and export of SBML models. Export has only been tested for models that do not contain the piecewise function.

	Tested models	Succeeded
(a) Import	216	212 (98 %)
(b) Simulation	212	208 (98 %)
(c) Export	18	18 (100 %)

Three of the models that fail to import because they use comparison between reals, which is not allowed in Modelica, while the fourth model fails because the

translator does not yet support delay in SBML events. One of the failing simulations is due to outdated SBML code (to many equations which was allowed previously, but is not allowed in Level 2 version 3), while the three others fail due to numerical reasons.

## 2.4 Comparison with other tools

The models in the BioModels database have been developed in different SBML tools, such as Systems Biology Workbench (SBW) [5], CellDesigner [6], and Copasi [7].

The libSBML [8] library contains consistency checks to ensure that a model is consistent with the SBML specification. However, this does not ensure that a model works to simulate in a specific software.

Therefore we have performed a comparison between *MathModelica*, Systems Biology Workbench, and CellDesigner to check if the results match the published results from the original tool used for respective model on the BioModels database. Due to time constraints we could not test the full set of available models, instead a subset was chosen, by testing every fifth model (every model with model name ending with either 0 or 5). The result is presented in Table 2.

**Table 2** Comparison between different modelling environments.

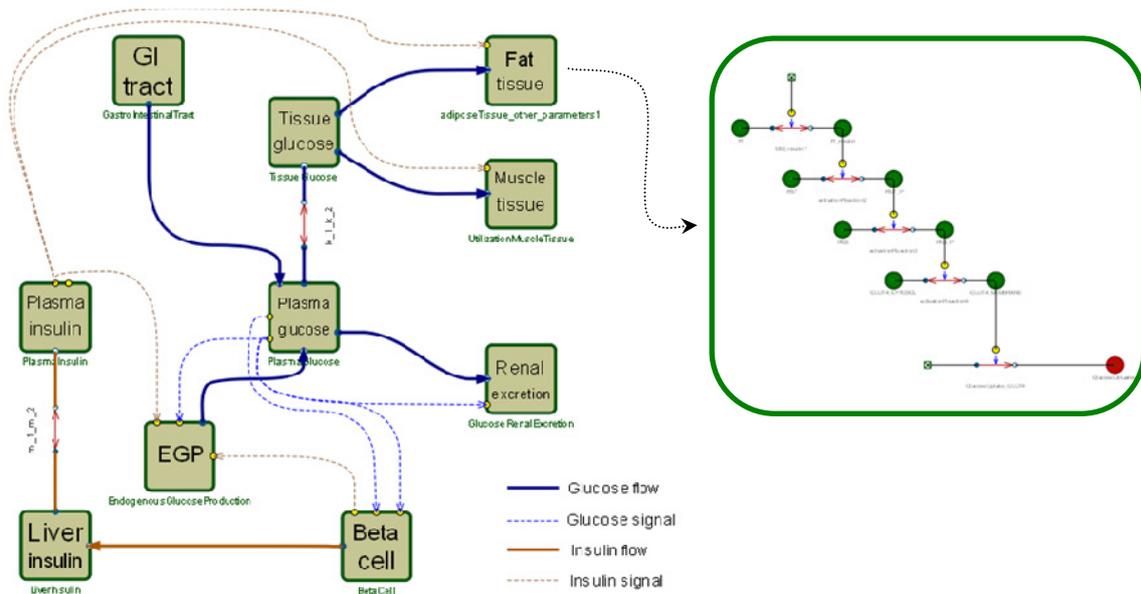
	Successful simulations
<i>MathModelica</i> 2.1 b4	98 % (42 of 43)
SBW 2.7.9	56 % (24 of 43)
CellDesigner 4.01	40 % (17 of 43)

The results indicate that *MathModelica* has a good support for the SBML language, especially in comparison with SBW and CellDesigner, however improvements are still needed to reach a complete support.

## 2.5 Related work

The BioDyn [8][9][10] software can make external calls to a few other software, including OpenModelica [11][12]. It also includes import functionality from Modelica to SBML. However, SBML models are imported as flat Modelica i.e. the pathway diagram is lost in the translation.

There is also some other related work done, but with limited or no public information available.



**Figure 2** The top-level of the whole-body model (left) shows the organs, the flows, and the information exchange between the organs, in an underlying level the detailed fat tissue module (right) is found.

### 3 Modelling Example

The Dalla Man model [4] is one of the best available models for the whole-body glucose-insulin system, and the model has even been formally accepted as a viable tool when certifying new drugs for diabetes. This model is, however, a non object-oriented model developed in MATLAB.

Based on this model a multi-level object-oriented whole-body model of the glucose-insulin system has been developed using *MathModelica*, MATLAB, and – for the communication between them – SBML.

#### 3.1 Creating an object-oriented model

The original MATLAB model was exported to SBML and inserted in *MathModelica*, and then reformulated into an object-oriented format, in which each object typically correspond to an organ, thus making it possible to successively replace respective organ with more detailed models. The top level of the model can be seen in Figure 2. As long as the input-output profiles of the replaced organ is preserved this can be done with negligible effects on the whole-body level.

#### 3.2 Detailed adipose tissue model

A detailed model for the adipose tissue (fat tissue) was developed (in MATLAB) using mechanistic *a*

*priori* information and existing and novel data from experiments on human fat cells. The parameters of the model were then optimized so that the simulations fitted both the input-output profile of the old model, and the cell-level data.

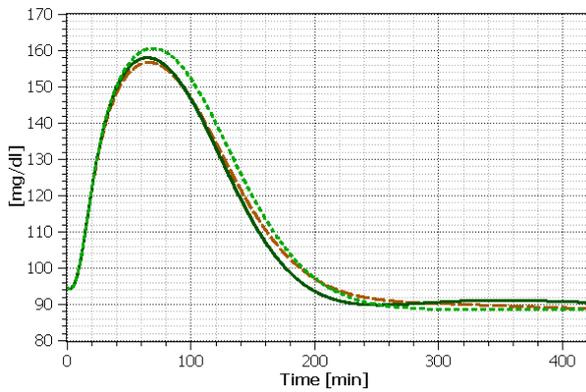
The model was exported to SBML, and then imported to *MathModelica*, and inserted in the whole-body model. As expected, this new detailed module did not cause any major changes in the whole-body behaviour (Figure 3), even though the intercellular glucose utilization rate (i.e. within the altered adipose tissue) had a somewhat different profile (Figure 4). Hence, the resulting object-oriented model can be said to have a "zoomable" adipose tissue.

To further show the strength of the object-oriented model, a parameter in the insulin signalling cascade was changed, so that the adipose tissue became 5 times less sensitive to insulin. This corresponds to a diabetic fat tissue in an otherwise healthy body. The simulations (Figure 2 and 3) predict that these rather big intracellular changes only leads to slightly higher glucose in the plasma, since a healthy body easily can adapt to the new situation, by slightly increasing the insulin level.

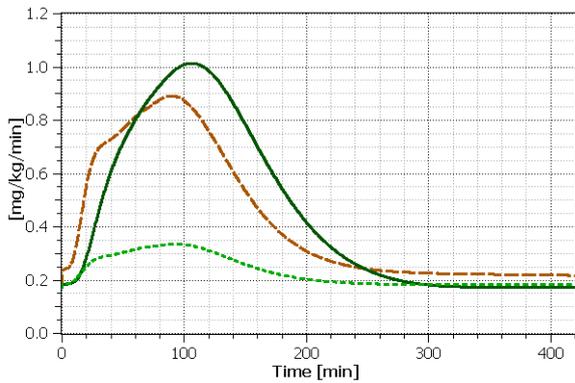
A main strength of the developed model is that it has the ability to translate mechanistically oriented simulations on the biochemical/cellular level, which is the level where drugs act, to the whole-body level, which is the level of clinical interest.

Another important benefit of object-oriented modelling is that the modeller is forced to adopt the cell-

oriented experiments to physiologically realistic conditions, compatible with the whole-body dynamics.



**Figure 3** Simulation results of glucose concentration in plasma of the original model (—), with the mechanistic fat tissue module included (---), and with changed parameters on fat tissue module-level (···).



**Figure 4** Simulation results of glucose utilization rate by adipose tissue of the original model (—), with the mechanistic fat tissue module included (---), and with changed parameters on fat tissue module-level (···).

## 4 Publishing on the Web

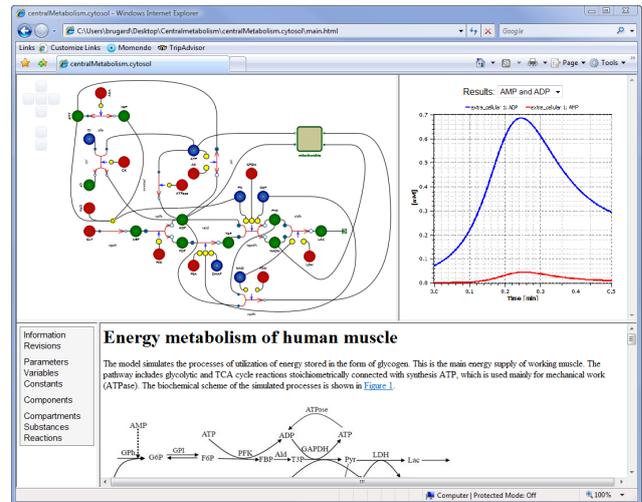
In order to communicate results with others a publish feature has been developed for *MathModelica*. Once a Modelica model has been developed it can be exported to an interactive HTML page.

It is possible to publish not only the selected model, but also its components. The exported model contains interactive model diagram based on the Silverlight technology [14], model information (documentation), and selected results.

Figure 5 shows an example of a published central metabolism model, originally developed by Selivanov et al [3].

### 4.1 Publishing on the BioBridge portal

Models can also be published as an integrated part of the BioBridge portal. The BioBridge portal allows modellers (and other users) to access an integrated resource including clinical, genomic, proteomic, and metabolomic information for the metabolic pathways affected by disorders such as chronic obstructive pulmonary disease (COPD), cardiac disease and diabetes. This information is then applied to model the underlying metabolic network. This, in turn, can then be utilized by the modeller for simultaneous analysis of multilevel data in order to improve existing knowledge on complex disorders.



**Figure 5** Screenshot of the published central metabolism model. Upper left corner shows the (clickable) model diagram of the cytosol pathway, upper right the concentration of AMP and ADP substances, lower left browser menu, and finally the lower right shows the model documentation (at the cytosol level).

### 4.2 Connecting to BioXM

The *MathModelica Systems Biology* toolbox also includes the possibility to tag substances in order to associate a specific substance with information available in the knowledge management environment BioXM [17] [18]. When a user selects a substance with a tag, an http request is made to the BioBridge portal and information is retrieved using BioXM, as illustrated in Figure 6.

BioXM directly, or by access to their Webservice query interfaces, integrates more than 20 different public databases and ontologies (see

Table 3 for integrated databases) representing a total of 80 793 genes (30 246 human, 27 237 mouse, 23 310 rat), 1 307 pathways, 78 528 compounds, 1 525 474 protein interactions a total of 3 666 313 connections within the knowledge network and the entire Gene Expression Omnibus database. Clinical and

experimental data from the two clinical BioBridge studies of chronic obstructive pulmonary disease (COPD) has been integrated with these public resources.

This information is combined with BioBridge literature-mining derived molecular networks for COPD, cardiac disease, chronic systemic inflammation, diabetes and lung and muscle specific signalling sub-networks. Experimental molecular sets encompass gene expression, metabolomics and proteomics data. Exercise and COPD specific kinetic and metabolic data relevant to constrain muscle metabolic models has been extracted from 30 published clinical trials. Finally the mathematical models and probabilistic networks generated within the BioBridge project are fed back into BioXM.

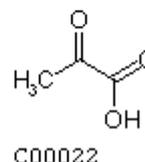
▼ **Metabolite Structural Information**

▼ **C00022**

**Object:**  [C00022](#)

**Formula:** C<sub>3</sub>H<sub>4</sub>O<sub>3</sub>  
C<sub>3</sub>H<sub>3</sub>O<sub>3</sub>-  
C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>

**Structure:**



**Figure 6** Metabolite structural information for Pyruvate given by BioXM.

**Table 3:** Public molecular data resources integrated into BioXM

Source Database	Information Type	Current Statistics	Level of curation	Updates/Version
<b>BIND</b>	Protein Interaction Molecular complexes Pathways	6256 Interactions	High throughput data submission and hand curated from the literature	last public version 20.3.07
<b>BioGrid</b>	Protein interaction	19 707 interactions	Manually curated from literature Different evidence codes	updated monthly
<b>Biomodels</b>	SBML models	224 SBML models	Partial verification of model simulation	updated monthly
<b>ChEBI</b>	Compound information	15 367	Curated from different data sources	updated weekly
<b>Comparative Toxicogenomics Database (CTD)</b>	Compound-gene, Compound-disease and Gene-disease relationships	259 898 relations	Manually curated from the published literature	updated monthly
<b>GEO</b>	Expression data	12 543 studies, 257 312 expression sets	Partial manual curation of meta-information into study sets	updated monthly
<b>EntrezGene</b>	Gene functional information	80 793 human, mouse and rat genes	Curated information integrated from different databases, based on RefSeq genomes	updated weekly
<b>Enzyme</b>	Enzyme related functional information	4 833	Manually curated from the published literature	updated weekly
<b>IntAct</b>	Protein interaction	21 584 binary interactions	Literature curation User submission	updated weekly
<b>KEGG</b>	Pathways	418 pathways	Manually curated from the published literature	updated monthly
<b>LIGAND</b>	Compound information	15 185	Manually curated from the published literature	updated monthly
<b>MIPS Mammalian</b>	Protein Interaction	410 interactions	Manually curated from the published literature	current Release 31.10.07
<b>OMIM</b>	Gene - disease relations	19 769	Curated from the published literature	updated weekly
<b>Pfam</b>	Protein family information	10 340 families	Manually curated from sequence alignments	23.0

## 5 Limitations and Current Work

The translator is currently in a beta stage and lacks support for unit conversions, and some, as it appears, less commonly used elements of SBML Level 2 version 3, such as initial assignments and constraints. Warnings will be issued if unsupported elements are detected during translation. One element (the piecewise function of SBML) is supported by the SBML to Modelica translator but not yet by the Modelica to SBML translator. Most of these shortcomings will be addressed in the final release of the translator.

The implementation where BioXM is used to access and retrieve information has served as a test implementation and will be extended to make use of the SBML RDF annotations in order to associate entities with biochemical or biological semantics.

During the development of the translator, SBML Level 2 version 4 was released. This version has not yet been looked at and hence is not supported by the translator.

Due to the lack of an accepted standard for storing graphical diagrams of models in SBML, the translator does not support conversion of graphical representations of models. SBML Level 3 is expected to resolve this issue.

While the wizards in *MathModelica* do a good job assisting the user in creating SBML exportable compartments, substances, and reactions in Modelica, they do not let the user edit these components once created. We intend to address this limitation in future releases.

The current version includes Metabolic Control Analysis (MCA) and dynamic sensitivity analysis; however there are various other common tasks like parameter estimation, model reduction, and stochastic simulation that could be added in the future.

## 6 Conclusions

The ambition is to create a bridge between the systems biology community and the Modelica community. The main benefits of the project is that it makes it possible to 1) convert an existing SBML model and benefit from Modelica's multi domain technology when extending the model, and 2) get direct access to biological databases and tools when developing a biochemical pathway in Modelica.

A public available Modelica library, BioChem, for modelling cellular pathways in Modelica has been

developed and a toolbox, *MathModelica Systems Biology*; that allows import/export to SBML, publishing of models and results with connections to biochemical databases has also been developed.

The translation from SBML to Modelica allows people that work with SBML, to import their models to *MathModelica* and use the full strength of the Modelica language to expand their models.

All in all, we believe that hierarchical modelling has a strong future for modelling of complex biological processes, and that *MathModelica*, in communication with SBML-based software and models, has the potential to take a leading role in such efforts.

## Acknowledgements

The BioBridge project is supported by the European Commission and is part of the Sixth Framework Programme.

## References

- [1] Larsdotter Nilsson, E. and Fritzson P., (2003). BioChem - A Biological and Chemical Library for Modelica. Proceedings of the 3rd International Modelica Conference: 215-220.
- [2] BioChem library. (Accessed 2009-08-18) <http://www.mathcore.com/products/mathmodelica/libraries/biochem.php>
- [3] Systems Biology Markup Language (Accessed 2009-08-18) <http://www.sbml.org>
- [4] The BioModels Database - A Database of Annotated Published Models. (Accessed 2009-08-18) <http://www.ebi.ac.uk/biomodels-main/>
- [5] Systems Biology Workbench. (Accessed 2009-08-20) <http://sbw.sourceforge.net/>
- [6] CellDesigner. (Accessed 2009-08-20) <http://www.celldesigner.org/>
- [7] COPASI - Complex Pathway Simulator. (Accessed 2009-08-20) <http://www.copasi.org>
- [8] libSBML. (Accessed 2009-08-18) <http://sbml.org/Software/libSBML>
- [9] BioDyn. (Accessed 2009-08-18) <http://cbbl.imim.es:8080/ByoDyn>
- [10] Gómez-Garrido, À. et al. (2008) ByoDyn: integrating computational methods for the analysis of biochemical models. P oster pre-

sented at the 4th Meeting of the Spanish Systems Biology Network (REBS).

- [11] OpenModelica. (Accessed 2009-08-18) <http://www.ida.liu.se/~pelab/modelica/OpenModelica.html>
- [12] Fritzson, P. et al. (2006) OpenModelica - A Free Open-Source Environment for System Modeling, Simulation, and Teaching. IEEE International Symposium on Computer-Aided Control Systems Design.
- [13] Dalla Man C, Rizza R A and Cobelli C (2007). Meal simulation model of the glucose-insulin system, IEEE transactions on bio-medical engineering, Vol. 54, No. 10, pp. 1740-1749.
- [14] Microsoft Silverlight. (Accessed 2009-08-18) <http://silverlight.net>
- [15] Selivanov VA, et al., (2008). The changes in the energy metabolism of human muscle induced by training. Journal of Theoretical Biology 252, 402-410.
- [16] The BioBridge portal. (Accessed 2009-08-18) <http://www.biobridge.eu>
- [17] Sameith, K. et al., (2008). Functional Modules integrating essential cellular functions are predictive of the response of leukaemia cells to DNA damage. Bioinformatics 24: 2602-2607.
- [18] Losko, S. et al. (2006) Knowledge Networks of Biological and Medical Data: An Exhaustive and Flexible Solution to Model Life Science Domains. Lecture Notes in Computer Science 4075: 232-239.