Systems Biology Markup Language Standard

Research coordinator
Hiroaki Kitano  The Systems Biology Institute: JAPAN  kitano@sbi.jp

Research team members
John Doyle  California Institute of Technology: USA
Andrew Finney  University Hertfordshire: UK
Poul Nielsen  University of Auckland : New Zealand

Duration : April 2002 – March 2005

Abstract
Building and analyzing models is the approach in systems biology. However, no ideal software platform is available today for research, which is a big issue to be solved. To solve the problem, we need (a) development and standardization of modeling language that can represent biological phenomena, and (b) development of software tools which is compliant to the standardized language.

Therefore, we have proposed Systems Biology Markup Language (SBML) as a modeling language, and have developed Systems Biology Workbench (SBW). The developed software has been publicly released under open source license (LGPL).

With the NEDO funding, we conducted the following activities:
(1) Definition of SBML Level-2 specification
(2) Proposal of SBML Level-3 specification
(3) Promotion of SBML

More than 90 groups commit developing SBML-compliant software. More are expected as Level-3 becomes available.

From international standardization viewpoint, we plan to create SBML to be “de facto”, then “de jure”. For de facto standardization, major journals such as Nature have decided to endorse SBML as standard representation for model data. Several commercial and public pathway databases have employed SBML.

For de jure standard promotion, IETF has been accepted. We have approached OMG, and continue to pursue more established approval bodies such as IEEE and ISO.

Keywords:  Systems Biology, SBML, SBW, Standardization

1. Introduction
One of the big challenges in post-genomics biology is to build and understand entire process of biological phenomena from cellular behavior to gene regulatory networks to metabolic pathways. Building of models and analyzing from various perspectives is the approach in computational systems biology that may bring a major impact to science and drug discovery. Currently a large number of developers and researchers are developing softwares to enable the modeling and its analysis. Despite those efforts, the softwares are scarcely compatible with
each other, thus as is often the case that a model built upon software cannot be easily transferred to other software. Additionally, wide varieties of methodologies for analysis or even for model building exist, and it is not realistic to cover all functions comprised by a single software. Therefore, no ideal software platform is available today for systems biology research, which is a big issue to be solved.

We believe that to solve the problem is (a) development and standardization of modeling language by which various phenomena could be represented, and (b) development of software tools which is compliant to the standardized language, which can be exchanged with each other.

2. Strategy

Based on this idea, we have proposed Systems Biology Markup Language (SBML) as a modeling language, and have developed Systems Biology Workbench (SBW). The developed software has been publicly released under open source license (LGPL).

With the help of the NEDO funding, we were able to conduct primarily following activities regarding standardization of SBML.

(1) Definition of SBML Level-2 specification
(2) Proposal of SBML Level-3 specification
(3) Publicity activities of SBML

3. Research

3.1 Definition of SBML specifications

SBML, the first version of Level-1 specification released March, 2001, is now available its Level-2 specification (released June 2003), and have adopted by many software.

With the latest version of Level-3 (although not yet released), various features will be added, which could not be represented in previous version, such as (1) spatial information of components, (2) modulation of proteins, and (3) multi-cellular model. Those features will serve modelers or researchers to simulate/analyze models in much practical manner.

Up to now, 80 groups commit developing software with SBML compliant features included. More groups are expected to join our community in proportion to version up of SBML Level, which will be continued next fiscal year.

<table>
<thead>
<tr>
<th>A) Modeling</th>
<th>B) Simulation</th>
<th>D) Utilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>CellDesigner</td>
<td>BALSA</td>
<td>Kinsolver</td>
</tr>
<tr>
<td>NetBuilder</td>
<td>BIOCHAM</td>
<td>MesoRD</td>
</tr>
<tr>
<td>CADLIVE</td>
<td>BioCharon</td>
<td>MMT2</td>
</tr>
<tr>
<td>Bio Sketch Pad</td>
<td>BioNetGen</td>
<td>Moleculizer</td>
</tr>
<tr>
<td>BioTapestry</td>
<td>BioSPICE/Â Dashboard</td>
<td>NetBuilder</td>
</tr>
<tr>
<td>BioUML</td>
<td>BioSpreadsheet</td>
<td>PathwayLab</td>
</tr>
<tr>
<td>JDesigner</td>
<td>BioUML</td>
<td>PathwayBuilder</td>
</tr>
<tr>
<td>PaVESy</td>
<td>CADLIVE</td>
<td>PNK</td>
</tr>
<tr>
<td>SRS Pathway Editor</td>
<td>CellDesigner</td>
<td>PROTON</td>
</tr>
<tr>
<td></td>
<td>Cellerator</td>
<td>PySCeS</td>
</tr>
</tbody>
</table>
Table 1: List of SBML-compliant software

<table>
<thead>
<tr>
<th>Software</th>
<th>Tool / Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>BioSens (UCSB)</td>
<td>DBsolve</td>
</tr>
<tr>
<td>Cytoscape</td>
<td>Dizzy</td>
</tr>
<tr>
<td>XPPAUT</td>
<td>E-Cell</td>
</tr>
<tr>
<td>FluxAnalyzer</td>
<td>ESS</td>
</tr>
<tr>
<td>MetaboLogica</td>
<td>Fluxor</td>
</tr>
<tr>
<td>MetaFluxNet</td>
<td>Gepasi</td>
</tr>
<tr>
<td>MetaboLogica</td>
<td>INSILICO discovery</td>
</tr>
<tr>
<td>SCIpath</td>
<td>Jarnac</td>
</tr>
<tr>
<td>SimWiz</td>
<td>JigCell</td>
</tr>
<tr>
<td>JigCell</td>
<td>Virtual Cell</td>
</tr>
<tr>
<td>JSIM</td>
<td>WinSCAMP</td>
</tr>
</tbody>
</table>

3.2 International Standardization

From international standardization viewpoint, we created SBML to be de facto standard, by involving large numbers of researchers and organizations in our community — this is the first step. As second step, we will commit on de jure standard of SBML under carefully planned scheme. During this period, we focused on creating de facto standard for SBML, but at the same time, we wrote documents for de jure standardization. As a matter of fact, our SBML to IETF (=The Internet Engineering Task Force, http://www.ietf.org/) as “application/sbml+xml MIME media type standards” has been accepted.

As far as de facto standard formation, it is noteworthy that major journals such as Nature have decided to endorse SBML as standard representation for model data that should be submitted with the paper. In addition, several commercial and public pathway databases have started to use SBML as underlying representation standard.

For de jure standard promotion, we are already contacting with OMG (=Object Management Group, http://www.omg.org), and continue to pursue more established approval bodies such as IEEE and ISO.

Submission of SBML files

The Systems Biology Markup Language (SBML) is a computer-readable format for representing models of biochemical reaction networks. SBML is applicable to metabolic networks, cell-signaling pathways, regulatory networks, and many others. Where relevant and possible, authors are encouraged to submit datasets in SBML format. Authors should select 'SBML' from the available list of data file formats when uploading the data set file.

http://mts-msb.nature.com/cgi-bin/main.plex?form_type=display_auth_instructions

Fig 2: From Nature Molecular Biology Website
4. Discussions

Today, SBML is well known to be the de facto standard for computational modeling in systems biology. The success of SBML has led to requests from the community for new features and continued evolution of the language. We’ve role as organizers and editors in the development of SBML, and respected many feedbacks from the open community. We’ve invited individual researchers who are interested in SBML to our SBML forum, and had a discussion about the evolution of SBML. Simultaneously, we have developed software infrastructure, including programming libraries, conversion utilities, interface packages for commonly-used software environments, and web based online tools. Most of these software are distributed as an open-source software to maximize the accessibility and utility of the products.

We think that such two-pronged approach (open community based discussion, and support of software infrastructure) is essential for the standardization.

5. Conclusions

Overall, our project went well and has achieved its goal in establishing de facto standard in the field of biological modeling. This is clearly illustrated by the fact over 80 software adapted SBML as its model representation language. In addition, in May 5 issue of Nature, Nature Publishing Group made it clear that they endorse SBML as a format for submitting biological model. The next step is to make this as de jury standard, and we are working on this goal too. However, it is important that SBML has established de facto status with strong back up from major publishers such as Nature made it almost certain that we can also achieve this goal.

Acknowledgement

SBML.ORG より転記

The development of SBML was originally funded by the Japan Science and Technology Agency under the ERATO Kitano Symbiotic Systems Project. Support for the continued development of SBML and associated software and activities today comes from the following sources:

• National Human Genome Research Institute (USA)
• National Institute of General Medical Sciences (USA)
• International Joint Research Program of NEDO (Japan)
• ERATO-SORST Program of the Japan Science and Technology Agency (Japan)
• Ministry of Agriculture (Japan)
• Ministry of Education, Culture, Sports, Science and Technology (Japan)
• BBSRC e-Science Initiative (UK)
• DARPA IPTO Bio-Computation Program (USA)
• Army Research Office's Institute for Collaborative Biotechnologies (USA)
• Air Force Office of Scientific Research (USA)

Additional support is provided by the California Institute of Technology (USA), the University of Hertfordshire (UK), the Molecular Sciences Institute (USA), and the Systems Biology Institute (Japan).

References

The list of the most important papers and patents from the project.

Papers (Total 34)

Presentations (Total 151)

Patents (Total 0)