

CMBSlib: a Library for Comparing Formalisms and Models of Biological Systems

Sylvain Soliman and François Fages
{Sylvain.Soliman,Francois.Fages}@inria.fr

Projet Contraintes, INRIA Rocquencourt,
BP105, 78153 Le Chesnay Cedex, France.
<http://contraintes.inria.fr>

Abstract. We present CMBSlib, a library of Computational Models of Biological Systems. It is aimed at providing a list of test problems for formalisms, modeling issues and implementation issues in systems biology.

The main motivation for CMBSlib is to stimulate research on the formal modeling of biological systems, by facilitating the exchange of formal models between researchers, and by providing a forum of comparison and validation of not only models, but also modeling formalisms and implementations.

Unlike a standardization effort, CMBSlib welcomes the most exotic formalisms and models provided they attack the modeling of well documented biological systems. Models of biological systems written in any referenced formalism can be submitted to CMBSlib. No special format or standard is required.

We discuss the advantages of and problems encountered in building such a library, give an example of typical entry in the library, and most of all we invite the community to become active contributors to CMBSlib.

1 Introduction

As the first CMSB workshop proved, modeling for Systems Biology is an important new task for computer scientists, mathematicians and biologists; and to quote the Call For Papers of this second edition:

As the field matures, it is becoming increasingly obvious that there is probably no 'one-size fits all' formal language for molecular biology, but rather several meddling paradigms, each with its strengths and weaknesses relatively to specific analytical goals.

It therefore seems that the need to compare all these formal languages, and their corresponding analytical capabilities, is also increasing.

On the other hand, most of the modeling work done around systems biology is quite difficult to transfer since lots of articles mentioning new models only describe the resulting analyses (usually a simulation plot) and the biological lessons learned from it. The models themselves often lack proper publication.

We thus advocate the need for a general framework allowing to compare:

- on the one hand, different formalisms, associated tools, resulting analyses;
- on the other hand, different models of the same or related biological systems.

CMBSlib aims at becoming such a framework, and provides at the same time a global facility to store and publish models. This is in contrast to the already existing model repositories, like that of the SBML-capable tool, Cellerator [1] for instance. All model repositories are indeed oriented towards one single formalism, whether ODEs, petri-nets or process calculi [2], while CMBSlib aims at confronting various formalisms useful for Systems Biology. Furthermore, existing repositories maintain one single model of a given biological system in a given organism, there is thus no facility for comparing different models of the same biological system.

There are also attempts at unifying the current mass of languages, such as that of the BioPAX group¹ for sharing pathway information, and if such unification succeeds we would be very happy; but until then, it seems necessary to allow the use of different formalisms, description languages and tools. CMBSlib was designed for that purpose in the framework of the ARC CPBIO [3] whose more general aim is to study new languages suited to Systems Biology.

One should also remark that the creation of analogous libraries in other domains of computer science has usually resulted in a big progress in comprehension of the issues involved. These repositories are mostly benchmarking libraries, but our constraint programming origin made us aware of CSPLIB [4], a library of constraint satisfaction problems, where “representation” is also a determining factor. That existing library was an important source of inspiration for building CMBSlib.

2 Comparing Formalisms

As made clear above, there are currently many formalisms used for modeling biological systems, and it is much too early to throw away all but one.

There is however currently no existing repository trying to encompass many formalisms in order to allow the user to choose the one the most appropriate to what he wants to do with the corresponding model.

Moreover, many models were often built about one given biological system, or even similar systems, but they usually are kept separate because they are expressed using different formalisms or description languages. This separation makes it impossible to benefit in one of those models from the enrichments brought to another one; it also impeaches any meaningful comparison.

All these barriers not only impact the users of the models, since they do not know about the real reasons to choose one or another formalism, and do not benefit from cross-improvement of the existing models, they also have a negative effect on the designers of formalism who need to build test-cases and showcases from scratch and have no facility for confronting their design choices to features found useful in other formalisms from different sources.

¹ <http://www.biopax.org>

3 Submission Guidelines

To help users submit models to the library, we provide some simple guidelines, implemented as an HTML form on CMBSlib’s home page²:

The aim of the library is to become a useful resource for all researchers on formal methods in systems biology. We thus welcome the submission of *any formal model of biological system*: to submit a new model, one has only to fill-in the submission form on the CMBSlib web site.

The models are classified in CMBSlib by the biological system they refer to. We therefore specify all biological systems in CMBSlib using natural language and reference to survey papers in Biology.

CMBSlib may contain different models of the same biological system, either in different formalisms or even in the same formalism. *No standard format or language is required.*

As we want to help people compare their work in CMBSlib with minimum effort, we encourage users of the library to send us the URL of any tools that might be useful to others (e.g. parsers for data files, simulators, analyzers, translators, ...). All such code is placed in the library through the generosity of the authors, and comes with all the usual disclaimers.

To make comparison with previous work easier, links to articles that use these models are provided. References to articles using models of CMBSlib are thus solicited, in order to be added to the references section of CMBSlib.

Finally, to make it easy to compare new models with others, a record of results (simulation plots, query results, etc.) will be provided. To help us keep these records up-to-date, users are encouraged to send in their latest results.

4 An Example

To get an idea of what information gets stored in the library, let us take an example, that of the “Mammalian cell cycle control”.

4.1 Specification

To define what this is about we first need a *specification* of the system: this usually takes the form of a few natural language sentences describing the system. For instance:

“A model of the known interactions of the mammalian cell cycle regulatory network at the molecular level.”

4.2 References

The informal specification should always for clarity be accompanied by some references to the relevant literature, in the present case the survey paper:

² <http://contraintes.inria.fr/CMBSlib/>

“Kurt K. Kohn, Molecular Interaction Map of the Mammalian Cell Cycle Control and DNA Repair Systems. Molecular Biology of the Cell, Vol. 10, pp.2703-2734, August 1999.”

Whenever possible links to databases such as PubMed³ or PNAS Online⁴ should be included, or if copyrights allow it, links to an electronic version of the article hosted (or mirrored) in the library.

4.3 Models and Analyses

Next, one can try and provide some models of that system in different *formalisms*, like κ 0 [5] and BIOCHAM [6, 7], two transition based formalisms. Then one can show what type of analyses are possible with each formalism and tool, what part of the model is correctly captured and what should still be improved. BIOCHAM for instance allows for some model-checking of CTL properties of the defined model.

It is interesting however to note that there are other models coming from the same survey of K. Kohn [8], like one using Pathway Logic [9] which uses the formal tools of Maude⁵ for analysis purposes.

Those different models each improved on the original Kohn map by elucidating unclear zones while formalizing the model, some also benefited from later improvements using the literature, it would thus be very fruitful for all of them to compare the results and correct/combine what needs so.

The same can be said of most KEGG [10] maps, which have been completed and corrected by different people a number of times, but since each of those corrections resulted in a model in a different language, they were usually never compared, nor reused.

The comparisons might be in some cases limited to some informal reasoning about the analyses obtained on different models, and in the first phase of CMBSlib’s life this type of exchange will probably be the most natural, and we believe already very fruitful. However in most cases some concrete guidelines can be elaborated for comparison, and the CMBSlib team will strive for such a result, with the help of the contributing modelers.

5 Conclusion

The existence of the CMSB conference and of its community already shows the need for a platform for sharing and comparing models and formalisms.

CMBSlib aims at becoming such a forum of exchange about models and formalisms, by providing tools like a model repository, a mailing-list, translators, etc. It should also stimulate research on the comparison of different formalisms

³ <http://www.ncbi.nlm.nih.gov/entrez/>

⁴ <http://www.pnas.org/>

⁵ <http://maude.csl.sri.com>

for expressing and analyzing a single model, and of different models for representing a single system. We believe that the resulting cross-fertilization of models and formalisms is an important step towards the difficult task of modeling issues in Systems Biology.

It is hoped that the community will contribute to the library since that is the only way it will develop, and since such a contribution will be beneficial to all contributors, whether mathematicians, computer-scientists or biologists, to improve their formalisms, tools and models.

References

1. Shapiro, B.E., Levchenko, A., Meyerowitz, E.M., Wold, B.J., Mjolsness, E.D.: Cellerator: extending a computer algebra system to include biochemical arrows for signal transduction simulations. *Bioinformatics* **19** (2003) 677–678 <http://www-aig.jpl.nasa.gov/public/mls/cellerator/>.
2. Regev, A., Silverman, W., Shapiro, E.Y.: Representation and simulation of biochemical processes using the pi-calculus process algebra. In: *Proceedings of the sixth Pacific Symposium of Biocomputing*. (2001) 459–470
3. ARC CPBIO: Process calculi and biology of molecular networks (2002–2003) <http://contraintes.inria.fr/cpbio/>.
4. Gent, I.P., Walsh, T.: CSPLIB: A benchmark library for constraints. In Jaffar, J., ed.: *Proceedings of the Fifth International Conference on Principles and Practice of Constraint Programming (CP'99)*. Volume 1713 of *Lecture Notes in Computer Science*, Springer-Verlag (1999) 480–481 A longer version appeared as technical Report TR-APES-09-1999.
5. Chabrier, N., Chiaverini, M., Danos, V., Fages, F., Schächter, V.: Modeling and querying biochemical networks. *Theoretical Computer Science* **To appear** (2004)
6. Chabrier-Rivier, N., Fages, F., Soliman, S.: The biochemical abstract machine BIOCHAM. In Danos, V., Schächter, V., eds.: *CMSB'04: Proceedings of the second Workshop on Computational Methods in Systems Biology*. *Lecture Notes in Computer Science*, Springer-Verlag (2004)
7. Chabrier, N., Fages, F., Soliman, S.: BIOCHAM's user manual. INRIA. (2003–2004) <http://contraintes.inria.fr/BIOCHAM/>.
8. Kohn, K.W.: Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Molecular Biology of Cell* **10** (1999) 703–734
9. Eker, S., Knapp, M., Laderoute, K., Lincoln, P., Meseguer, J., Sönmez, M.K.: Pathway logic: Symbolic analysis of biological signaling. In: *Proceedings of the seventh Pacific Symposium on Biocomputing*. (2002) 400–412
10. Kanehisa, M., Goto, S.: KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **28** (2000) 27–30