On Representing Biological Systems through Multiset Rewriting

F. Martinelli[1], S. Bistarelli[1], I. Cervesato[2⋆], G. Lenzini[3], and R. Marangoni[4,5]

[1] Istituto di Informatica e Telematica — C.N.R.
Via G. Moruzzi 1, I-56100 Pisa — Italy
{fabio.martinelli,stefano.bistarelli}@iit.cnr.it

[2] Advanced Engineering and Sciences Division, ITT Industries, Inc.
Alexandria, VA 22303 — USA
iliano@itd.nrl.navy.mil

[3] Istituto di Scienze e Tecnologie Informatiche — C.N.R.
Via G. Moruzzi 1, I-56100 Pisa — Italy
lenzini@iei.pi.cnr.it

[4] Istituto di Biofisica — C.N.R.
Via G. Moruzzi 1, I-56100 Pisa — Italy
{roberto.marangoni}@ib.pi.cnr.it

[5] Dipartimento di Informatica, Università di Pisa,
Via F. Buonarroti 2, 56127 Pisa — Italy.

1 Overview

In the post-genomic era, the most prominent biological problems are detecting, describing and analyzing the informational flows that make a set of molecules a living organism [5]. Genomic and proteomic techniques, in fact, are producing the largest set of biological data available ever, but the problem of detecting and describing how these entities (genes and proteins) interact with each other in the complex molecular machinery of the cell has just begun being addressed. It is necessary to find easy, comprehensive, and biological-friendly *models* to describe molecules and their interactions.

It is common opinion [5] that an ideal model for biological enquiring has to satisfy three requirements:

- It must be suitable for describing metabolic networks, in order to create metabolic databases allowing the user to search for and compare biochemical pathways in living organisms (like the genomic and proteomic database are already doing).
- It must be implementable into a simulation machine, in order to realize dynamic models of metabolic pathway that allow studying possible critical situations and steady states, and generally predicting that certain conditions will happen.

⋆ Cervesato was partially supported by NRL under contract N00173-00-C-2086. This work was completed while this author was visiting Princeton University.

– It must be possible to run dynamic simulations in which to evaluate how external agents interfere with molecules and processes, in order to infer the consequences on the metabolic network stability. This kind of applications is a useful in silico test of possible side effects of a drug.

For these reasons, proper theories and instruments of Formal Methods research community may help in defining formal models and tools (e.g., see [6]), since they have been used so far to represent different kinds of relationships and dynamic interactions among objects and processes in distributes systems. In this paper, we use Multiset Rewriting (MSR) [3, 2, 4, 7], a logic-based formalisms based on term rewriting. *MSR* offers both a formal language for a precise description of molecular interaction maps, and an execution model allowing simulation of dynamics of molecular networks with the theoretical possibility of predicting optimal values for certain parameters used in the system description. Basic mechanisms in *MSR* include: (a) a multiset of items or predicates (possibly with an internal structure), used to describe a system state, which can represent objects or resources or generic entities; (b) a set of rewriting rules which act on a state by consuming and producing items. We use stochastic MSR: The rewriting step is not immediate but has a duration with a certain probability distribution, usually the exponential one. This duration depends on the state where the rule is applied, allowing the modeling of reaction inhibitors. Those simple and abstract mechanisms are expressive enough in describing both the static and the dynamic aspects of metabolic pathways.

In the field of security protocol analysis, many description languages have been proposed. Recently, there is a successful research trend that exploits MSR as a low level language for describing such protocols, and the other formalisms are compiled into it. Indeed, due to the simplicity and uniformity of the *MSR* rewrite mechanism, many researchers feel easier to develop analysis algorithms for *MSR*.

In this paper, we show that a well known model of analysis may be encoded into the *MSR*, i.e. Stochastic Petri Nets (SPSs) and we argue that it will be possible to faithfully encode also other specification languages as the stochastic process algebra [6] into stochastic *MSR* (similarly to the encoding we proposed in [1] for security analysis). A benefit is that the intermediate language is itself a significant one for biologists.

Stochastic Petri Nets (SPNs) are a family of distributed calculi, based on the notions of *places*, *tokens*, *markings* and *transitions*. The idea is that places stores resources, i.e. tokens. A marking is an instantaneous picture of the tokens present in the places. A transition is a relation among set of places to set of places and describes how markings change during the computation. A link from a place to the transition itself is equipped with a number which expresses how many token are necessary from that place to enable the transition. A link from a transition to a place is equipped with a number that expresses how many token are produced when the transition is fired. Stochastic Petri Nets have a weight in the transition that represents the duration of the transition.

An analogy between *SPN*'s and *MSR* may be the following: Places correspond to Predicate names; a token in a place is a predicate of a certain kind in the multiset and thus a marking is essentially a multiset of atomic predicates. Transitions are encoded as rewriting rules: A transition that consumes n tokens from a certain place is is encoded through a rule which requires n instances of the predicate corresponding to the place in the left hand side; a transition that produces m tokens in a place corresponds to a rule which requires m predicates on the right end side. The weight function of the rewriting rule is the same of the transition, provided the analogy among the marking and the multiset. Thus, computation graph produced by the *MSR* from an initial multiset precisely corresponds to the computation graph of the initial marking of the corresponding Petri net.

Future work. We have mainly three goals:

- We plan to encode other formal computation frameworks, e.g. the calculus of [6], within *MSR*, in order to strengthen our claim about the MSR as a suitable low level language for modeling biological systems;
- We plan also to adapt the rich theory already developed for *MSR*, which comprises very useful formal notions like composition, abstraction, equivalence and congruences, to obtain suitable definitions for biological systems;
- We plan to produce an optimized simulation environment based on the input syntax of *SMSR* (possibly refined with built-in predicates or constraints) and map into other specification formalisms.

References

1. S. Bistarelli, I. Cervesato, G. Lenzini, and F. Martinelli. Relating process algebras and multiset rewriting (for example for security protocol analysis). Technical report, Istituto di Scienza e Tecnologie dell'Informazione (ISTI-CNR), 2002. To appear.
2. I. Cervesato. A Specification Language for Crypto-Protocols based on Multiset Rewriting, Dependent Types and Subsorting. In G. Delzanno, S. Etalle, and M. Gabbrielli, editors, *Workshop on Specification, Analysis and Validation for Emerging Technologies — SAVE'01*, pages 1–22, Paphos, Cyprus, 2001.
3. Iliano Cervesato, Nancy Durgin, Patrick D. Lincoln, John C. Mitchell, and Andre Scedrov. A Meta-Notation for Protocol Analysis. In *12th Computer Security Foundations Workshop — CSFW-12*, pages 55–69, Mordano, Italy, 28–30 June 1999. IEEE Computer Society Press.
4. S. Eker, M. Knapp, K. Laderoute, P. Lincoln, J. Meseguer, and K. Sonmez. Pathway logic: Symbolic analysis of biological signaling. In *Proc. Pacific Symposium on Biocomputing*, volume 7, pages 400–412, 2002.
5. A. Pevzner. *Computational Molecular Biology.* MIT press, 2000.
6. C. Priami, A. Regev, E. Shapiro, and W. Silverman. Application of a stochastic name-passing calculus to representation and simulation of molecular processes. *Information Processing Letters*, 80, 2001.
7. M. Lange und U. Scholz. R. Hofestdt. Molecular information fusion for metabolic networks. In *In G.X. Xue, Y.B. Xue, Z.H. Xu, R. Holmes, G. Hammond und H.A. Lim, Herausgeber, Gene Families: Studies of DNA, RNA, Enzymes and Proteins*, pages 221–232, 2001.