Representation and simulation of molecular pathways in the stochastic π-calculus

Aviv Regev

Department of Cell Research and Immunology, Tel Aviv University
Tel Aviv 69978, Israel

Department of Computer Science and Applied Mathematics
Weizmann Institute of Science, Rehovot 76100, Israel

aviv@wisdom.weizmann.ac.il

Abstract

A formal language for describing biochemical networks is essential for rapid advancement in this field. We propose to model biomolecular processes by using the stochastic π-calculus, a process algebra originally developed for describing computer processes. Our model for biochemical processes is mathematically well-defined, while remaining biologically faithful and transparent. It is amenable to computer simulation, analysis and formal verification.

1 Introduction

A formal framework for the study of molecular pathways is one of the major goals of pathway informatics. An appropriate solution requires a formal representation language that would incorporate detailed information of the molecular, biochemical and dynamic aspects of pathways. This would facilitate the simulation, analysis and comparison of their behavior and function.

We suggest using the π-calculus for the representation and simulation of molecular pathways. The π-calculus is a formal language originally developed for specifying concurrent computational systems [5]. In such systems, multiple processes interact with each other by synchronized pair-wise communication on complementary communication channels, and modify each other by transmitting channels from one process to another. This feature, termed mobility, allows the network structure to change with interaction. Molecular pathways share many of the features of such computational systems, namely concurrency and compositionality (complex systems composed of independent parts), mobility (change in network structure as a result of interaction), and a modular and hierarchical structure.

We show how the π-calculus can be used to model biochemical [8]. We treat molecules and their individual domains as computational processes, where their complementary structural and chemical determinants correspond to communication channels. We model chemical interaction and subsequent modification as communication and channel transmission. To accurately describe the quantitative behavior of biochemical networks, we extended this qualitative view by modifying the existing stochastic variant of the π-calculus [6] according to biochemical principles [7]. Based on this modified formalism we have developed a computer system, BioPSI, for the representation and simulation of biochemical systems, and used it to study several molecular pathways.

2 Formal Representation of Molecular Pathways

Our π-calculus modeling is based on treating molecules and their constituent domains by the processes in which they may participate. We view each biochemical pathway as a process, denoted by a capitalized name, P (e.g System, figure 1A). Each pathway is defined as a collection of concurrently operating molecules, seen as processes with potential behavior. Concurrency is denoted by the PAR operator, | (e.g. Erk1 | $\cdots$ | Mek1 | $\cdots$, Figure 1A). Similarly, a protein molecule is composed of several domains, each of which is modeled as a sub-process (e.g. the three domains of Erk1, Figure 1A).

Two molecules (or domains) interact with each other based on their complementarity. Interaction is accomplished by the motifs and residues that constitute a domain. These are viewed as channels or communication ports of the molecule. Two complementary motifs are denoted by a name and co-name pair of global channels, x and $\overline{\mathrm{x}}$. An interaction event may occur only

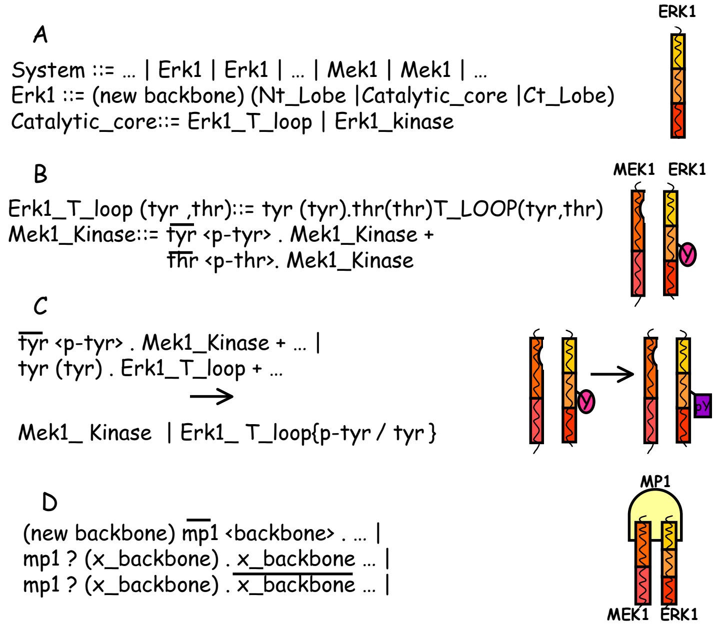


Figure 1: Principles for modeling molecular pathways in the π-calculus: (A) Domains, molecules and entire pathways are processes; (B) Molecular complementarity as channel names and co-names; (C) Interaction and modification as communication and channel mobility; (D) Complex formation as extrusion of local channel's scope

when such a complementary pair is shared by the interacting molecules (tyr and $\overline{\text{tyr}}$ in Erk1 and Mek1 respectively, Figure 1B). A process can also be defined in a parametric way where a parameter is a motif (Erk1_T_loop(tyr, thr), Figure 1B).

A sequence of interactions in which a molecule may participate is denoted by a prefix operator, . (tyr(tyr) . thr(thr) . $\cdots$, Figure 1B). In addition, a molecule may often participate in one of several mutually exclusive interactions. These are summed together, by +, denoting competitive choice (e.g. $\overline{\text{tyr}}\langle \text{p} - \text{tyr}\rangle \, . \, \cdots + \overline{\text{thr}}\langle \text{p} - \text{thr}\rangle \, . \, \cdots$ indicates that the kinase may either interact with a tyrosine or a threonine). All communications are offered, but once one is chosen the others are excluded. In our stochastic variant, this choice is probabilistic, and governed by biochemical principles (below), based on reaction rates [8]. The language also allows us to follow a molecule's fate through interaction by using recursive parametric definitions (e.g. $\mathrm{T\_loop(tyr, thr)}$ in Figure 1B).

Biochemical interaction affects subsequent events in the pathway. This is modeled using *mobility* [5]. Mobility is achieved by allowing processes to send and receive global channel names which may be used for subsequent interaction. Thus, upon interaction on the tyr channel, the Erk1_T_loop process receives a $\mathrm{p-tyr}$ channel from the Mek1_kinase. This $\mathrm{p-tyr}$ replaces tyr in the T_loop, enabling interactions with complementary motifs in phosphatases and disabling ones with kinases (Figure 1C).

A pathway is composed of defined compartments, from the covalently linked parts of a molecule, through multi-molecular complexes, to sub-cellular compartments. We represent compartments by restricted communication scopes. Channel scope is restricted using the operator new (backbone in Figure 1A). Mobility is also used to model compartment changes, such as complex formation. Private channels, representing scopes, are extruded and changed, thereby changing scopes dynamically. For example, when an MP1 scaffold interacts with both Erk1 and Mek1, the same backbone channel is sent to the two Nt_Lobe*s*. These three molecules may now continue to exclusively communicate on the backbone channel, and are thus linked (Figure 1D).

According to these general principles, detailed information on complex pathways, molecules, and biochemical events, can be represented formally.

3 A simulation system based on the stochastic π calculus

In order to allow accurate quantitative modeling, we have adapted and modified the semantics of a stochastic variant of the calculus [6], to comply with biochemical principles (as detailed in [7]). Briefly, every channel x is now attached with a base rate r, corresponding to the appropriate reaction rate, and a global (external) clock is added to the system. At each reduction step, this clock is advanced and a communication is selected from all the enabled ones, according to an appropriate "race condition". In devising and implementing this revised semantics, we relied on an existing algorithm [3] ensuring the faithfulness of our semantics.

We implemented the biochemical stochastic π-calculus as part of the BioPSI application, based on the FCP platform Logix [9]. We devised an appropriate insulated surface syntax, and built a compiler to FCP. The interactions between processes on channels are mediated by a monitor, implementing the race condition. Several tracing and debugging tools are available for following a simulation [1]. These include a full ordered and timed trace of all events, which is post-processed to produce a quantitative time-evolution for each kind of process.

4 Stochastic π-calculus models for the RTK-MAPK pathway and the circadian clock

We have so far developed π-calculus models for several signaling pathways, including the RTK-MAPK pathway [8]. The RTK-MAPK model is composed of 15 molecular processes, with 24 different domains and 15 sub-domains. Four compartments (extracellular, membrane, cytoplasm and nucleus) were defined. A major portion of current knowledge has been incorporated into this concise (250 lines) formal representation [1].

The compositional nature of the calculus allows us to easily perturb our model not only at the quantitative level but also at the molecular one: we introduced "mutant" molecules by "hacking" their code. Such manipulation is similar to mutational analysis in laboratory experiments: domains and residues are deleted, inserted or modified. Specifically, we introduced dominant negative "monomer" ligands and "truncated" receptors by removing sub-processes, as well as constitutively active "membrane-bound" Raf and "mutated " Erk1 by changing channel parameters.

While most perturbations have yielded the expected effects, we have observed that many perturbations can be buffered by the system, albeit not at extreme doses, indicating potentially robust architecture. These results, which require detailed examination, are discussed in detail in [8].

A second, simpler model which we are actively studying is that of the circadian clock. We based our model on an existing generalized one [2]. In this system (Figure 2A), an oscillator is created from interleaving a positive and a negative feedback loop in the same pathway. Previous work [2] has shown that this architecture is robust to perturbations, as long as differential rates are maintained. Our stochastic π-calculus program includes six molecular process types (for the gene, RNA and protein of each molecule), with a straightforward code (e.g. Figure 2B). This molecular model gives the expected oscillatory behavior (e.g. Figure 2C). Using the compositionality of the calculus we are currently comparing this molecular model to a functional module, and examine the interaction between the circadian clock and other regulatory pathways in the cell (unpublished data).

5 Discussion

Representing biochemical networks in the π-calculus offers a synthesis of both declarative and procedural representations. On the one hand, molecular detail is clearly defined in a well-structured, biologically faithful fashion. This is based on a strong correspondence between the syntax of the calculus and biochemical networks. The compositionality of both biochemical systems and the calculus allows incremental modeling of complex systems. On the other hand, the formal stochastic semantics which we have associated with this syntax can serve as a basis for simulation (as part of the BioPSI system) and analysis.

Our approach enjoys many of the benefits of other CS approaches which are currently being adapted for studying biochemical systems. Similar to Petri nets [4], it treats biochemical pathways as concurrent processes. Petri nets have been successful in treating many, mostly metabolic, biochemical systems. The advantage of the π-calculus is in the explicit, detailed description of each node or process in complex signaling pathways.

In order to become a full bioinformatics approach further development is required. In particular, the detailed level offered by the formalism can become deterring at times. Furthermore, some complex biochemical events require elaborate encoding. These problems may be alleviated by taking advantage of the modular nature of the calculus. Thus, different components of the same system can be specified in different levels of detail. This approach may be highly suitable for molecular modules, and we employ it in studying the circadian clock.

A specification-based approach also allows to formally analyze and compare the behavioral properties of modeled systems, using methods and tools for formal verification [5]. This opens up completely novel possibilities in the study of biochemical systems. Comparison of similar pathways is the first step in studying the evolution of entire processes, establishing a *homology* *of processes*. Comparative methods can also be employed for modular analysis, in order to prove that the molecular information indeed supports the behavior we expect and specify.

Acknowledgments

The development of the stochastic π-calculus and BioPSI was done jointly with Udi Shapiro and Bill Silverman from the Weizmann Institute, and Corrado Priami from the University of Verona, Italy. The work on the circadian clock is done in collaboration with Naama Barkai, from the Weizmann Institute. A.R. is a Colton Fellow.

References

[1] The BioPSI project web site. `http://www.wisdom.weizmann.ac.il/~aviv`.

[2] N. Barkai and S. Leibler. Circadian clocks limited by noise. *Nature*, 403(6767):267–268, 2000.

[3] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[4] P. J. E. Goss and J. Peccoud. Quantitative modeling of stochastic systems in molecular biology by using stochastic petri nets. *Proceedings of the National Academy of Sciences USA*, 95(12):6750–6755, 1998.

[5] R. Milner. *Communicating and Mobile Systems: the π-calculus*. Cambridge University Press, 1999.

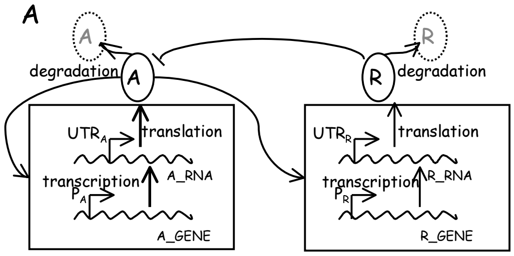
[6] C. Priami. Stochastic π-calculus. *The Computer Journal*, 38(6):578–589, 1995.

[7] 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*, 2001. in press.

[8] A. Regev, W. Silverman, and E. Shapiro. Representation and simulation of biochemical processes using the pi- calculus process algebra. In R. B. Altman, A. K. Dunker, L. Hunter, and T. E. Klein, editors,

Pacific Symposium on Biocomputing, volume 6, pages 459–470, Singapore, 2001. World Scientific Press.

[9] W. Silverman, M. Hirsch, A. Houri, and E. Shapiro. The Logix system user manual, version 1.21. In E. Shapiro, editor, *Concurrent Prolog (vol. II)*, pages 46–78. MIT Press, Cambridge, Massachusetts, 1987.



B

```
A_GENE::= PROMOTED_A + BASAL_A
PROMOTED_A::= pA ? {e}.ACTIVATED_TRANSCRIPTION_A(e)
BASAL_A::= bA ? [].( A_GENE | A_RNA)
ACTIVATED_TRANSCRIPTION_A::=
          τ1 . (ACTIVATED_TRANSCRIPTION_A | A_RNA) +
          e ? [] . A_GENE

RNA_A::= TRANSLATION_A + DEGRADATION_mA
TRANSLATION_A::= utrA ? [] . (A_RNA | A_PROTEIN)
DEGRADATION_mA::= degmA ? [] . 0

A_PROTEIN::= (new e1,e2,e3) .
             PROMOTION_A-R + BINDING_R + DEGRADATION_A
PROMOTION_A-R ::=   pA!{e2}.e2![]. A_PROTEIN +
                    pR!{e3}.e3![]. A_PRTOEIN
BINDING_R ::= rbs ! {e1} . BOUND_A_PRTOEIN
BOUND_A_PROTEIN::= e1 ? [].A_PROTEIN  + degpA ? [].e1 ![].0
DEGRADATION_A::= degpA ? [].0
```

C

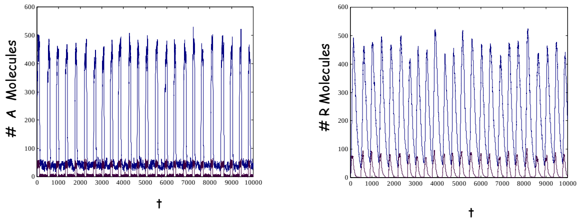


Figure 2: The circadian clock: (A) A generalized model of the circadian clock [2]. Interleaved positive (A-driven) and negative (R driven) feedback loops result in oscillation in A and R quantities. (B) π-calculus code for the "A" molecules of the circadian clock. The full model is available at [1]. (C) BioPSI simulation results for the circadian clock: Oscillating protein (higher peaks) and RNA (low peaks) quantities for the A and R species.