Application of a stochastic name-passing calculus to representation and simulation of molecular processes

Corrado Priami* Aviv Regev†‡ Ehud Shapiro†§ William Silverman†

Abstract

We describe a novel application of a stochastic name passing calculus for the study of biomolecular systems. We specify the structure and dynamics of biochemical networks in a variant of the stochastic π-calculus, yielding a model which is mathematically well-defined and biologically faithful. We adapt the operational semantics of the calculus to account for both the time and probability of biochemical reactions, and present a computer implementation of the calculus for biochemical simulations.

Keywords: stochastic π-calculus, bioinformatics, stochastic simulation.

1 Introduction

Biomolecular processes, carried out by large complex networks of interacting proteins, are responsible for most of the information processing within the living cell. Previous attempts at modeling such processes have used continuous mass-action differential equations, discrete Monte-Carlo simulations, or Petri nets (e.g. [9]). While each of these approaches captures some of the information regarding pathways and their components, none fully integrates dynamics, molecular, and biochemical detail. As an alternative we have proposed [15] to represent molecular systems as mobile communicating systems in the π-calculus [12],[1] yielding a model of the molecular realm which is both highly detailed and visible. For this purpose we have implemented a qualitative simulation

*Dipartimento Scientifico Tecnologico, Università di Verona Ca' Vignal 2, Strada Le Grazie 15, I-37134 Verona, Italy – priami@sci.univr.it

†Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, 76100, {aviv,bill,udi}@wisdom.weizmann.ac.il

‡Department of Cell Research and Immunology, Life Sciences Faculty, Tel Aviv University, Tel Aviv 69978, Israel.

§Department of Biological Chemistry, Weizmann Institute of Science, Rehovot, 76100.

[1] We assume familiarity with the π-calculus.

system based on the π-calculus in Flat Concurrent Prolog [17]. Importantly, unlike previous implementations of the calculus [13], our system, BioPSI, supports full synchronized[2] communications and full choice (including mixed choice between input and output actions).

The π-calculus's non-determinism is well-suited for qualitatively modeling biomolecular systems. However, for more accurate quantitative modeling, we propose using the stochastic π-calculus. Stochastic process algebras were introduced by [8, 10, 3, 4] to compositionally model quantitative aspects of distributed systems, and have been mostly used for performance evaluation. Their application to biomolecular processes is novel.

The paper is structured as follows. In Section 2 we describe biomolecular processes and their qualitative representation in the π-calculus, and briefly review quantitative information. In Section 3 we present a stochastic variant of the π-calculus suitable for biochemical dynamics. In Section 4 we describe the extension of our implementation to the stochastic calculus.

2 Biomolecular processes in the π-calculus

Biomolecular processes are carried out by networks of interacting protein molecules, each composed of several distinct independent structural parts, called *domains*. Pair-wise interaction between domains depends on structural and chemical complementarity of particular portions, called *molecular determinants* or *motifs*.[3] Interaction between proteins causes biochemical modification of motifs (e.g. covalent changes). These modifications affect the potential of the modified protein to interact with other proteins. Since protein interactions directly affect cell function, these modifications are the main mechanism underlying many cellular functions, making the π-calculus particularly suited for their modeling as mobile communicating systems.

We view molecules and domains as processes, represent complementary motifs by global channel names and co-names, and identify complexes and cellular compartments by newly declared private channels. We model chemical interaction and subsequent modification as communication and channel transmission. The operational semantics of the calculus thereby defines the dynamic

[2] Molecular interaction is invariably a synchronous event.

[3] There is currently no standard definition of a motif. The term encompasses a variety of factors, including electrostatics, hydrogen bonding and three dimensional structure.

behaviour of the modeled system.

To accurately describe the quantitative behavior of biochemical networks, this qualitative view must be extended. The actual rate of a reaction between two proteins is determined according to a *basal rate*[4] and the concentrations or quantities[5] of the reactants [11]. We distinguish between two types of reactions that are common in biomolecular networks. In the usual reaction, two different reactant molecules, P and Q, are involved, and the reaction rate is given by $Brate \times |P| \times |Q|$, where $Brate$ is the reaction's basal rate, and $|P|$ and $|Q|$ are the concentrations of P and Q in the chemical solution. In another prevalent kind of reaction, homodimerization, two identical proteins, Q, bind together. The interacting processes are represented in this case as mixed choice constructs offering both an input and an output communication on the same channel. The rate in this case is $1/2 \times Brate \times |Q| \times (|Q|-1)$. In the next section, we present the biochemical stochastic π-calculus, which provides a formal semantics for these reaction rates.

3 The biochemical stochastic π-calculus

The original semantics of the stochastic π-calculus [14] required some modification in order to accurately describe chemical reactions. In this section we describe the reduction semantics of the biochemical variant that we have developed.

The prefix $\pi.P$ of the π-calculus [12] is replaced in the stochastic variant by $(\pi, r).P$ where r is the single parameter of an exponential distribution that characterizes the stochastic behaviour of the activity corresponding to the prefix π. Thus, r corresponds to the basal rate of a biochemical reaction.[6] Otherwise, original π-calculus syntax [12] remains intact. The structural congruence $\equiv$ [12] is extended with $A(\tilde{y}) \equiv P\{\tilde{y}/\tilde{x}\}$ (if $A(\tilde{x}) ::= P$ is the unique defining equation of constant A). Similarly to [12] we assume all processes in head normal form. In particular, a process P is in *head normal form* if either it is the null process or $P \equiv \sum_i (\pi_i, r_i).P_i$ and $\forall i \neq j \, . \, sbj(\pi_i) \neq sbj(\pi_j)$.[7]

[4] The basal rate of a reaction is an empirically-determined constant, which depends on the specific reaction, the temperature, etc.

[5] Under certain assumptions [7], molecule concentration and quantity can be interchanged.

[6] In the original stochastic π-calculus [14] the rate is associated with the prefix. However, in a chemical reaction both reactants share a single basal rate. This is resolved by associating the basal rate with the channel name. For clarity purposes, we continue to specify the rate r in the prefixes throughout the paper, implicitly assuming that two prefixes have the same rate when using the same channel name.

[7] $sbj(\pi)$ denotes the subject of π, i.e. its output or input link.

Note, that this condition is justified since we assume at most one occurrence of a given motif in a domain.

As for semantics, the dynamic evolution of systems is driven by a *race condition*, yielding a probabilistic model of computation. All the activities that are enabled in a state compete and the fastest one succeeds. The continuity of exponential distributions ensures that the probability that two activities end simultaneously is zero.

Since reaction rates depend on the number of interacting processes, we define two auxiliary functions, $In, Out : 2^{\mathcal{P}} \times \mathcal{N} \to \mathbb{N}$ that inductively count the number of receive and send operations on a channel x enabled in a process:

$$
\begin{aligned}
& In_x(\mathbf{0}) = 0 \\
& In_x(\sum_{i \in I} (\pi_i, r_i).P_i) = |\{(\pi_i, r_i) | i \in I \wedge sbj(\pi_i) = x\}| \\
& In_x(P_1 | P_2) = In_x(P_1) + In_x(P_2) \\
& In_x((\nu\, z)P) = \begin{cases} In_x(P) & \text{if } z \neq x \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

Out_x is similarly defined, by replacing any occurrence of In with Out and the condition $sbj(\pi_i) = x$ with $sbj(\pi_i) = \overline{x}$.

Table 1 shows the reduction semantics of the biochemical stochastic π-calculus. We distinguish the two types of chemical reactions according to the channel name. A subset of all the names, $\mathcal{H} \subseteq \mathcal{N}$, is used to identify channels used in homodimerization reactions. A usual reaction is implemented by the three parameters r_b, r_0 and r_1, where r_b represents the basal rate, and r_0 and r_1 denote the quantities of interacting molecules, and are computed compositionally via In_x and Out_x while deducing transitions. The first axiom in Table 1 corresponds to usual reactions, with two different molecules. The second one corresponds to homo-dimerization reactions.

We illustrate our model with the following biomolecular system, regulating gene expression by positive feedback. The system Sys specified in Table 2 (and illustrated in Fig. 1) includes two genes ($Gene\_A$ and $Gene\_TF$), their transcribed mRNAs (RNA_A and RNA_TF), the corresponding translated proteins ($Protein\_A$ and $Protein\_TF$) and the degradation of both RNA

$$(\ldots + (\overline{x}\langle z\rangle, r).Q)|((x(y), r).P + \ldots) \xrightarrow{x, r_b \cdot 1 \cdot 1} Q|P\{z/y\},\ x \notin \mathcal{H}$$

$$(\ldots + (\overline{x}\langle z\rangle, r).Q + (x(y), r).P)| \\ ((\overline{x}\langle z\rangle, r).Q + (x(y), r).P + \ldots) \xrightarrow{x, 1/2 \cdot r_b \cdot 2 \cdot (2-1)} Q|P\{z/y\},\ x \in \mathcal{H}$$

$$\frac{P \xrightarrow{x, r_b \cdot r_0 \cdot r_1} P'}{P|Q \xrightarrow{x, r_b \cdot r_0' \cdot r_1'} P'|Q}, \begin{cases} r_0' = r_0 + In_x(Q) \\ r_1' = r_1 + Out_x(Q) \end{cases}$$

$$\frac{P \xrightarrow{x, r_b \cdot r_0 \cdot r_1} P'}{(\nu x)P \xrightarrow{x, r_b \cdot r_0 \cdot r_1} (\nu x)P'} \qquad \frac{Q \equiv P, P \xrightarrow{x, r_b \cdot r_0 \cdot r_1} P', P' \equiv Q'}{Q \xrightarrow{x, r_b \cdot r_0 \cdot r_1} Q'}$$

Table 1: Reduction semantics of the biochemical stochastic π-calculus.

and protein molecules.[8] The events are mediated by interaction with cellular machineries for DNA transcription ($Transcr$), RNA translation ($Transl$) and RNA and protein degradation (RNA_Deg and $Protein\_Deg$). Each of these interactions involves different molecular motifs (channels $basal$, utr, $degm$, and $degp$) and occurs at a different rate. In addition, $Protein\_A$ activates $Protein\_TF$ in a two-step mechanism. First, A binds TF, through A's $Binding\_Site$ domain ($bind$ channel), to form a complex by extrusion of A's **private** $backbone$ channels (bbi) to TF. Second, A's $Kinase$ domain modifies the bound TF protein, by sending the **global** channel $ptail$ (on the private $bb2$). All interactions between the parts of the complex are mediated on the private backbone channels: unbinding is mediated on $bb1$, while degradation, initiated on $degP$ in protein A, is propagated throughout the complex on $bb3$. Following modification and unbinding, TF can rapidly bind the transcription machinery using its newly acquired $ptail$ channel, causing faster promotion of transcription, closing a positive feedback loop. A computation leading to this situation is shown in Figure 2.[9] Note, that in this example, only few of the interactions we show are mobile. However, in a typical biological specification many such mobile communications take place (see examples in [15],[1]).

[8] A polyadic version of the calculus is used and the trailing 0 is omitted as usual.

[9] where we let $S_A = Gene\_A|RNA\_A|Protein\_A$, $S_{TF} = Gene\_TF|RNA\_TF|Protein\_TF$ and $S = Transcr|Transl|RNA\_Deg|Protein\_Deg$.

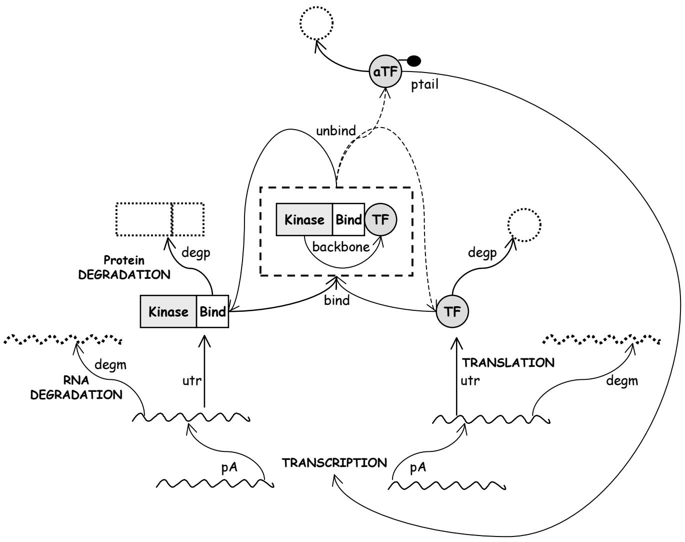

Figure 1: A simple biomolecular process: Transcriptional regulation by positive feedback.

4 Implementation

We implemented the biochemical stochastic π-calculus as part of the BioPSI application, based on the FCP platform Logix [17, 16]. We devised an appropriate insulated surface syntax, and built a compiler to FCP. Two unique features of FCP made it suitable for our purposes. First, the ability to pass logical variables in messages is used to implement name passing. Second, FCP's support of guarded atomic unification allows synchronized interaction between input and output guards.

In BioPSI, each channel is an object (a persistent procedure) and is associated with a basal rate. BioPSI processes send requests to the channel, via an FCP vector. There are four kinds of requests: send, receive, send & receive (for homodimerization), and withdraw. Requests to a channel which has an infinite rate are satisfied as soon as possible. Requests to a channel which has a finite rate (> 0) are queued.

Each time that a new event is required the central BioPSI monitor and all channel objects with a finite, non-zero rate, jointly determine a communication event. Each channel object determines a weighted rate, according to its basal rate and the numbers of send and receive offers. Based on an existing algorithm (detailed in [7]), the monitor selects randomly among the weighted rates, and

$$
\begin{array}{l}
\mathrm{Sys} = \mathrm{Gene\_A}|\mathrm{Gene\_TF}|\mathrm{Transcr}|\mathrm{Transl}|\mathrm{RNA\_Deg}|\mathrm{Protein\_Deg} \\
\mathrm{Gene\_A} = (\mathrm{basal}(),4).(\mathrm{Gene\_A}|\mathrm{RNA\_A}) \ + \ (\mathrm{pA}(),40).(\mathrm{Gene\_A}|\mathrm{RNA\_A}) \\
\mathrm{RNA\_A} = (\mathrm{utr}(),1).(\mathrm{RNA\_A}|\mathrm{Protein\_A}) \ + \ (\mathrm{degm}(),1) \\
\mathrm{Protein\_A} = (\nu \mathrm{bb1},\mathrm{bb2},\mathrm{bb3})(\mathrm{Binding\_Site}|\mathrm{Kinase}) \\
\mathrm{Binding\_Site} = (\overline{\mathrm{bind}}\langle \mathrm{bb1},\mathrm{bb2},\mathrm{bb3}\rangle,0.1).\mathrm{Bound\_Site} \ + \ (\mathrm{degp}(),0.1).(\overline{\mathrm{bb3}},\infty) \\
\mathrm{Bound\_Site} = (\overline{\mathrm{bb1}},10).\mathrm{Binding\_Site} \ + \ (\mathrm{degp}(),0.1).(\overline{\mathrm{bb3}},\infty).(\overline{\mathrm{bb3}},\infty) \\
\mathrm{Kinase} = (\overline{\mathrm{bb2}}\langle \mathrm{ptail}\rangle,10).\mathrm{Kinase} \ + \ (\mathrm{bb3}(),\infty) \\
\mathrm{Gene\_TF} = (\mathrm{basal}(),4).(\mathrm{Gene\_TF}|\mathrm{RNA\_TF}) \ + \ (\mathrm{pA}(),40).(\mathrm{Gene\_TF}|\mathrm{RNA\_TF}) \\
\mathrm{RNA\_TF} = (\mathrm{utr}(),1).(\mathrm{RNA\_TF}|\mathrm{Protein\_TF}) \ + \ (\mathrm{degm}(),1) \\
\mathrm{Protein\_TF} = (\mathrm{bind}(\mathrm{c\_bb1},\mathrm{c\_bb2},\mathrm{c\_bb3}),0.1).\mathrm{Bound\_TF} \ + \ (\mathrm{degp}(),0.1) \\
\mathrm{Bound\_TF} = (\mathrm{c\_bb1}(),10).\mathrm{Protein\_TF} \ + \ (\mathrm{c\_bb3}(),\infty) \ + \\
\qquad (\mathrm{c\_bb2}(\mathrm{tail}),10).((\mathrm{c\_bb1}(),10).\mathrm{Active\_TF}(\mathrm{tail}) \ + \ (\mathrm{c\_bb3}(),\infty)) \\
\mathrm{Active\_TF}(\mathrm{tail}) = (\overline{\mathrm{tail}},100).\mathrm{Active\_TF}(\mathrm{tail}) \ + \ (\mathrm{degp}(),0.1) \\
\mathrm{Transcr} = (\overline{\mathrm{basal}},4).\mathrm{Transcr} \ + \ (\mathrm{ptail}(),100).(\overline{\mathrm{pA}},40).\mathrm{Transcr} \\
\mathrm{Transl} = (\overline{\mathrm{utr}},1).\mathrm{Transl} \\
\mathrm{RNA\_Deg} = (\overline{\mathrm{degm}},1).\mathrm{RNA\_Deg} \\
\mathrm{Protein\_Deg} = (\overline{\mathrm{degp}},0.1).\mathrm{Protein\_Deg}
\end{array}
$$

Table 2: Specification of a biomolecular system.

stochstically selects according to the sum of weighted rates an appropriate reaction time interval to advance a "clock" counter. The chosen channel completes one transmission (send/receive pair), relaying the sent message to the receiver.

The completion of the send and receive requests is synchronized by the channel. In addition, other messages offered on this and other channels by the same two processes whose requests were completed, are withdrawn (mutually exclusive choice). The withdrawals are not synchronized, but they do precede continuation of their respective processes.

Each BioPSI process is transformed to an FCP procedure, and its channel set (global channels, arguments, newly declared channels and channels, bound by input, to be instantiated only following communication) is identified, thus allowing full use of channels as in the original calculus. Note, that the BioPSI process retains a segment of a short circuit, which is extended when the channel is passed to more than one process (including itself, recursively) and closed when the channel reference is no longer required. When all segments of the short circuit have been closed, the channel object terminates.

Several tracing and debugging tools are available for following a simulation [1]. These include

$$
\begin{tikzcd}
Sys \arrow[r, "{basal,4\cdot 2\cdot 1}"] & S_0 \arrow[r, "{utr,1\cdot 1\cdot 1}"] & S_1 \arrow[r, "{basal,4\cdot 2\cdot 1}"] & S_2 \arrow[r, "{utr,1\cdot 2\cdot 1}"] & S_3 \arrow[d, "{bind,0.1\cdot 1\cdot 1}"] \\
 & S_7 & S_6 \arrow[l, "{ptail,100\cdot 1\cdot 1}"'] & S_5 \arrow[l, "{bb1,10\cdot 1\cdot 1}"'] & S_4 \arrow[l, "{bb2,10\cdot 1\cdot 1}"']
\end{tikzcd}
$$

$$S_0 = Gene\_A|RNA\_A|Gene\_TF|S \qquad S_1 = S_A|Gene\_TF|S$$

$$S_2 = S_A|Gene\_TF|RNA\_TF|S \qquad S_3 = S_A|S_{TF}|S$$

$$
\begin{aligned}
S_4 &= Gene\_A|RNA\_A|((\nu bbi)(Bound\_Site|Kinase))|Gene\_TF|RNA\_TF| \\
&\quad Bound\_TF\{bbi/c\_bbi\}|S \\
S_5 &= (\nu bbi)(Gene\_A|RNA\_A|Bound\_Site|Kinase|Gene\_TF|RNA\_TF| \\
&\quad ((bb1(),10).Active\_TF(ptail)\ +\ (bb3(),\infty)))|S \\
S_6 &= (\nu bbi)(S_A|Gene\_TF|RNA\_TF|Active\_TF(ptail))|S \\
S_7 &= (\nu bbi)(S_A|Gene\_TF|RNA\_TF|Active\_TF(ptail))| \\
&\quad (\overline{pA},40).Transcr|Transl|RNA\_Deg|Protein\_Deg
\end{aligned}
$$

Figure 2: A computation of Sys.

a full ordered and timed trace of all events, which is post-processed to produce a quantitative time-evolution for each kind of process. For example, the behavior of the simple biochemical system described in Table 2 is presented in Figure 3. Higher levels of Protein A in the presence of positive feedback due to the TF activator gene are clearly observed.

The use of Gillespie's well-established [7] algorithm for the implementation of the race condition ensures the biochemical faithfullness of BioPSI stochastic simulations. The implementation was tested with several simple and realistic biomolecular models (including the circadian, the RTK-MAPK, and the wnt pathways) yielding results which are in agreement both with published simulation and analysis data (e.g. for the circadian clock [2]) and with experimental observations (e.g. for the RTK-MAPK pathway).

5 Conclusions

We described a biochemical variant of the stochastic π-calculus that is suitable to specify biomolecular processes. We have modified the original stochastic π-calculus to account for the rates of chemical reactions. We believe that the complexity and importance of biomolecular processes justify the introduction of such a variant. Even a simplified biomolecular system (Fig. 1 and 2) is relatively complex, and requires automated tools for quantitative analysis. We present such

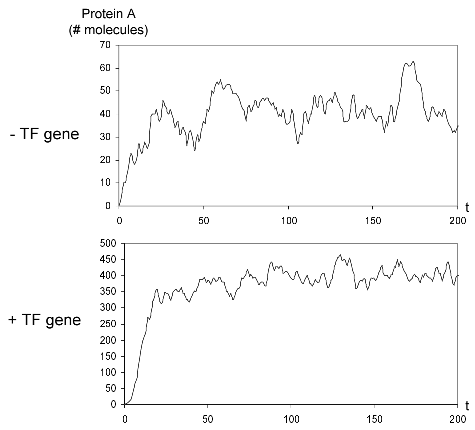

Figure 3: BioPSI simulation output for a simple biochemical system: Time evolution (sec.) of the absolute number of protein A molecules in the absence (top) and presence (bottom) of the TF activator gene.

a tool, BioPSI, for stochastic discrete simulation of biochemical systems, implemented in FCP. While discrete-event simulators have been implemented in the past (e.g. [6, 5]), BioPSI is the first to support mobility, a key feature for simulating biochemical systems. The BioPSI application is being used to study a variety of biomolecular systems [1].

In addition to simulation studies, formal methods could also bring benefits to medicine and biology by providing rigorous means for exploring and comparing biomolecular processes. Similar, but not identical biomolecular processes operate under different physiological, pathological, and evolutionary conditions. Revealing such similarities and differences is key to understanding the function of these processes in physiology and disease.

Acknowledgements. The second and third author thank Gideon Schechtman from the Weizmann Institute for his help.

References

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

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

[3] M. Bernardo, L. Donatiello, and R. Gorrieri. A formal approach to the integration of performance aspects in the modelling and analysis of concurrent systems. *Information and Computation*, 144:83–154, 1998.

[4] P. Buchholz. On a Markovian process algebra. Technical Report 500, Informatik IV, University of Dortmund, 1994.

[5] P.R. D'Argenio, J.-P. Katoen, and E. Brinksma. An algebraic approach to the specification of stochastic systems. In *Proceedings of PROCOMET'98, IFIP Series*, pages 126–147. Chapman-Hall, 1998.

[6] T. Field, P. Harrison, and K. Kanani. Performance modelling and verification of cache coherency protocols. In M. Ribaudo, editor, *Proceedings of PAPM96*, pages 199–214. CLUT Torino, 1996.

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

[8] N. Goetz, U. Herzog, and M. Rettelbach. Multiprocessor and distributed system design: The integration of functional specification and performance analysis using stochastic process algebras. In *Proceedings of Performance Evaluation of Computer and Communication Systems, LNCS 729*, pages 121–146, 1993.

[9] P.J.E. Goss and J. Peccoud. Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. In *In Proceedings of the National Academy of Sciences USA*, volume 95, pages 6750–6754, 1998.

[10] J. Hillston. *A Compositional Approach to Performance Modelling*. Cambridge University Press, 1996.

[11] B. Mahan and Myers R.J. *University Chemistry*. Addison-Wesley, 1987.

[12] R. Milner, J. Parrow, and D. Walker. A calculus of mobile processes (I and II). *Information and Computation*, 100(1):1–77, 1992.

[13] B.C. Pierce and D.N. Turner. PICT: A programming language based on the pi-calculus. In *Proof, Language and Interaction: Essays in honour of Robin Milner*. MIT Press, 1999.

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

[15] A. Regev, W. Silverman, and E. Shapiro. Representing biomolecular processes with computer process algebra: π-calculus programs of signal transduction pathways. In *Proceedings of the Pacific Symposium of Biocomputing 2000*, 2000. To appear.

[16] E. Shapiro. Concurrent prolog: a progress report. In E. Shapiro, editor, *Concurrent Prolog (vol. I)*, pages 157–187. MIT Press, Cambridge, Massachusetts, 1987.

[17] W. Silverman, M Hirsch, A. Houri, and E. Shapiro. *The Logix system user manual, Version 1.21. - Concurrent Prolog (vol. II)*. MIT Press, 1987.