UNIVERSITY
OF TRENTO

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGY

38050 Povo – Trento (Italy), Via Sommarive 14
http://www.dit.unitn.it

CAUSAL P-CALCULUS FOR BIOCHEMICAL MODELLING

M. Curti, P. Degano, C. Priami and C.T. Baldari

2002

Technical Report # DIT-02-0022

Causal π-Calculus for Biochemical Modelling *

M. Curti, P. Degano
Dipartimento di Informatica, Università di Pisa
Corso Italia 40 I-56125 Pisa, Italy
{curtim,degano}@di.unipi.it

C. Priami
Dipartimento di Informatica e Telecomunicazioni, Università di Trento
Via Sommarive, 14 I-38050 Povo, TN, Italy
priami@dit.unitn.it

C.T. Baldari
Dipartimento di Biologia Evolutiva, Università di Siena
Via Mattioli, 4 I-53100 Siena, Italy
baldari@unisi.it

1 Introduction

An increasing number of researchers is trying to define models of biochemical pathways via theoretical and technological tools, allowing biologists to simulate reactions before doing them *in vitro*. The advantages are obvious: a computation normally requires less time then a real experiment, simulation of reactions is cheaper than doing them effectively, and so on.

The problem is to find a formal modelling language for systems biology that represents:

- the "actors" of the system (molecules, proteins, genes...) at different level of abstraction;
- the qualitative evolution of the system in terms of their reactions possibly expressing causality between reactans or reactions;
- all the quantitative aspects of the pathways (quantity of reagents, temperature, reaction rates, ...)

Recently Regev, Silvermann & Shapiro [8] proposed the π-calculus [3] as a qualitative model of biochemical pathways, seen as network of proteins. Then [6] use a stochastic variant of the π-calculus [5], yielding a language that permits to model also quantitative aspects of biochemical pathways.

*The first three authors have been partially supported by UE project IST-32072-DEGAS within the FET initiative on Global Computing.

Since causality often allows more accurate and more concise representations of system behavior, we extend the π-calculus to mechanically derive a causality relation between reactions. As a matter of fact, the qualitative and the quantitative aspects are orthogonal to each other. So, the stochastic and the causal semantics can be combined together to yield even more detailed and accurate models of biochemical processes. Our version of the calculus is flexible and supports both qualitative and quantitative modelling: the stochastic [4] and the causal descriptions of biological evolution can be specified within a uniform single model.

We briefly show how the π-calculus is used to model network of processes. Entities can interact by exchanging messages. Essentially, molecules are a set of *domains* denoted by *motifs*. In the π-calculus we view molecules and domains as concurrent processes composed via parallel operators (written as $A_1 \mid \ldots \mid A_n$), while motifs are represented by global channels $(a, b, \ldots, x, y, \ldots)$.

Two molecules can interact if they have two *complementary* motifs (the same channel on which a molecule performs a send operation and the other a receive operation). Reactions are then modelled as communications, in which the two reactants communicate between a global channel (where $a(w)$ denotes an input action on the channel a, and $\overline{a}\langle y\rangle$ an output one):

$$\overline{a}\langle y\rangle.A \mid a(w).B \xrightarrow[R]{\langle \overline{a}\langle y\rangle, a(w)\rangle} A \mid B \tag{1}$$

The interaction produces a result called the *residual* of the reaction $(A|B)$.

Any molecule may have some private "information", the *backbone*, that determines its identity. The interaction between two molecules can be seen as sharing a backbone. The same happens with protein complexes or cellular compartments [6].

In the model, backbones are private names $((\nu\ n)A)$ of processes declared by a restriction. So, interaction between molecules results also in a scope enlargement.

2 Causality

We say that two reactions are in a causal relation if the first reaction is a necessary condition for the other to fire. To express causality we enrich motifs with informations about logical placement in the network of the molecules to which they belong. This is represented by a logical address ϑ automatically computed from the description of the biological system in our modelling language.

Our idea is to enrich processes and labels of transitions with addresses ϑ. For instance, in (1) we can rewrite the processes inserting addresses of their sequential components yielding

$$||_0\overline{a}\langle y\rangle.P \mid ||_1 a(w).Q \tag{2}$$

and the enriched label on the transition

$$||_0\overline{a}\langle y\rangle.P \mid ||_1 a(w).Q \xrightarrow[R]{\langle ||_0\overline{a}\langle y\rangle, ||_1 a(w)\rangle} ||_0 P \mid ||_1 Q \tag{3}$$

Note that we essentially record the position of processes with respect to the parallel structure of the overall system using the tag $||_0$ ($||_1$) to denote the left (right) side of a parallel operator. Furthermore, we can also have parallel tags outside pairs on labels of transitions as in the case

$$R\,|(P \mid Q) \xrightarrow[R]{||_1\langle\vartheta_0\mu_0,\vartheta_1\mu_1\rangle} R\,|(P' \mid Q') \tag{4}$$

where reactants P and Q are in parallel (in the same solution) with a third reactant R. Now we sketch how the relation of causality is mechanically derived. Roughly speaking, if we have the sequence of reactions

$$P \xrightarrow[R]{\vartheta\langle||_0\vartheta_0\mu_0,||_1\vartheta_1\mu_1\rangle} Q \xrightarrow[R]{\vartheta'\langle||_0\vartheta'_0\mu'_0,||_1\vartheta'_1\mu'_1\rangle} R \tag{5}$$

we can say that the first reaction causes the second one if $\vartheta||_0\vartheta_0$ is a prefix of $\vartheta'||_0\vartheta'_0$ or of $\vartheta'||_1\vartheta'_1$ (and similarly for $\vartheta||_1\vartheta_1$). This conditions reflects the nesting of actions in the specification (for a full account onn the definition see [1]). If there is no causal relation the two reactions can occour in any temporal order: they are independent.

Although it offers a qualitative view of processes, causality seems to play a relevant role in understanding complex biochemical reactions. As a side effect, causal-based representation of a biochemical pathway is more readable and more expressive then other proposals, because (i) the model identifies reactants involved in the actions, (ii) determines the state of the system during the evolution, and (iii) represents causal relations between reactions.

3 An Example

In this section we briefly show an application of our reduction semantics to a well-characterized biochemical process: the activation of the transcription factor $NF-AT$, which plays a crucial role in the process of T-cell activations.

$NF-AT$ is composed of two subunits, a cytosolic subunit, which belongs to the family of $NF-AT$ proteins, and a nuclear subunit, identified as $AP-1$. Activation of $NF-AT$ requires translocation of the cytosolic subunit to the nucleus, where it can assemble with $AP-1$. The latter in turn requires to be phosphorylated to become transcription-competent. Regulation of the subcellular localization of $NF-AT$ is achieved by phosphorylation on specific serine residues. When phosphorylated, the nuclear localization signal of NF–AT is masked and $NF-AT$ is therefore segregated to the cytosol. T-cell antigen receptor (TCR) engagement triggers an increase in intracellular calcium ions, which induces the activation of the phosphatase calcineurin. Dephosphorylation of NF–AT by calcineurin results in exposure of the nuclear localization sequence and translocation of $NF-AT$ to the nucleus. $AP-1$ on the other hand is heterodimer of Jun and Fos and is activated as the result of phosphorylation of Jun on serine residues. Activation of the MAP kinase pathway, which is also triggered by the TCR, is required for this process to occur [7]. Hence TCR

engagement results in the activation of both pathways, the Ras/MAP kinase and the calcium/calcineurin pathway, required for $NF-AT$ activation [2].

The model of the system follows:

$$Sys = ((((||_0||_0||_0||_0 TCR \mid ||_0||_0||_0||_1 ZAP\text{–}70) \mid ||_0||_0||_1 Grb2) \mid ||_0||_1 SOS) \mid (||_1||_0 RAS \mid (||_1||_1||_0 RAF \mid (||_1||_1||_1||_0 cNF-AT \mid ||_1||_1||_1||_1 PLC)))$$

$$TCR = (\nu\ tcr)\ \overline{bind\_z}\langle tcr\rangle$$

$$ZAP\text{–}70 = bind\_z(tcr).LAT$$

$$LAT = (\nu\ lat1)||_0\overline{bind\_grb2}\langle lat1\rangle \mid (\nu\ lat1)||_1\overline{bind\_PLC}\langle lat1\rangle$$

$$Grb2 = bind\_grb2(lat1).\overline{bind\_SOS}\langle lat1\rangle$$

$$SOS = bind\_SOS(lat1).\overline{bind\_RAS}\langle lat1\rangle$$

$$RAF = bind\_RAF(lat1).\tau_{MEK}.\tau_{ERK}.AP\text{–}1$$

$$PLC = bind\_PLC(lat2).CN$$

$$AP\text{–}1 = bind\_ap1(cn)$$

$$CN = (\nu\ cn)\overline{bind\_sub}\langle cn\rangle$$

$$cNF-AT = bind\_sub(cn).nNF-AT$$

$$nNF-AT = \overline{bind\_ap1}\langle cn\rangle$$

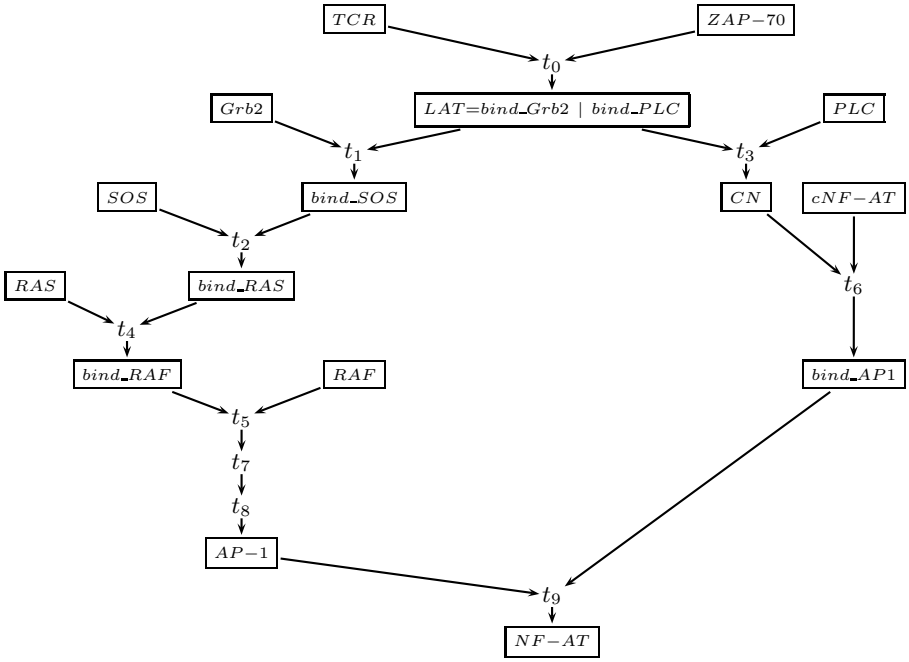

Figure 1: A computation of our model.

Figure 1 is built according to the causal relation between transitions $t_1 \ldots t_9$. For readability, the processes, enclosed in boxes, have no address. Causality

(both on transitions and processes) is represented by the (Hasse diagram resulting from the) arrows; their absence makes it explicit concurrent activities. For example, the initial interaction t_0 causes both t_1 and t_3, that instead are unrelated. The transitive closure of the arrows give raise newly to the causal relation. It is immediate to see which transition are causal related and which aren't.

References

[1] P. Degano and C. Priami. Non-interleaving semantics for mobile processes. *Theoretical Computer Science*, 216, 1-2:237–270, 1999.

[2] L.P. Kane, J. Lin, and A. Weiss. Signal transduction by the tcr for antigen. *Curr. Opin. Immunol.*, 12:242–249, 2000.

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

[4] C. Nottegar, C. Priami, and P. Degano. Performance evaluation of mobile processes via abstract machines. *IEEE Transactions on Software Engineering*, 27, 10:867–889, 2001.

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

[6] C. Priami, A. Regev, W. Silverman, and E. Shapiro. Application of a stochastic passing-name calculus to representation and simulation of molecular processes. *Information Processing Letters*, 80:25–31, 2001.

[7] A. Rao, C. Luo, and P.G. Hogan. Transcription factors of the nfat family: regulation and function. *Annu. Rev. Immunol.*, 15:707–747, 1997.

[8] A. Regev, W. Silverman, and E. Shapiro. Representation and simulation of biochemical processes using the π-calculus process algebra. In *Pacific Symposium of Biocomputing (PSB2001)*, pages 459–470, 2001.