Modeling Cellular Signaling Systems: An Abstraction-Refinement Approach

Diana Hermith, Carlos Olarte, Camilo Rueda, Frank Valencia

▶ To cite this version:

HAL Id: hal-00641433
https://hal.archives-ouvertes.fr/hal-00641433

Submitted on 16 Nov 2011

Modeling Cellular Signaling Systems: An Abstraction-Refinement Approach

Diana Hermith, Carlos Olarte, Camilo Rueda, and Frank D. Valencia

Abstract The molecular mechanisms of cell communication with the environment involve many concurrent processes governing dynamically the cell function. This concurrent behavior makes traditional methods, such as differential equations, unsatisfactory as a modeling strategy since they do not scale well when a more detailed view of the system is required. Concurrent Constraint Programming (CCP) is a declarative model of concurrency closely related to logic for specifying reactive systems, i.e., systems that continuously react with the environment. Agents in CCP interact by *telling* and *asking* information represented as constraints (e.g., $x > 42$). In this paper we describe a modeling strategy for cellular signaling systems based on a temporal and probabilistic extension of CCP. Starting from an abstract model, we build refinements adding further details coming from experimentation or abstract assumptions. The advantages of our approach are: due to the notion of partial information as constraints in CCP, the model can be straightforwardly extended when more information is available; qualitative and quantitative information can be represented by means of probabilistic constructs of the language; finally, the model is a runnable specification and can be executed, thus allowing for the simulation of the system. We outline the use of this methodology to model the interaction of G-protein-coupled receptors with their respective G-proteins that activates signaling pathways inside the cell. We also present simulation results obtained from an implementation of the framework.

Diana Hermith

Dept. of Natural Science and Mathematics, Pontificia Universidad Javeriana Cali, Colombia. e-mail: dhermith@javerianacali.edu.co

Carlos Olarte and Camilo Rueda

Dept. of Computer Science, Pontificia Universidad Javeriana Cali, Colombia. e-mail: {carlosolarte,crueda}@javerianacali.edu.co

Frank D. Valencia

CNRS LIX, Ecole Polytechnique, France. e-mail: fvalenci@lix.polytechnique.fr

1 Introduction

Molecular biologists use information and computer technology to process, analyze, understand, compare and share scientific knowledge. The major effort is to scale up to system biology, taking under consideration spatio-temporal interactions of molecules. Complex biological processes are orchestrated by means of a precise dynamic regulation of cell behavior, achieved through an active dialog between cells and their environment controlled by cell-surface receptors. In response to specific ligands, these translate the environmental cues into specific intracellular signaling reactions to achieve an appropriate response [1].

Formal computational methods can be useful in this setting to develop reasoning skills and to establish conceptual frameworks to handily represent biological behavior. This contributes not only to theoretical biology, but also to experimental biologists by offering a fertile substrate to think and redesign experiments.

This paper contributes in the modeling of biological phenomena by using a compositional and scalable representation of them. For this end, we shall use Probabilistic Temporal Concurrent Constraint Programming, a simple but powerful model for concurrency that allows for the specification of reactive systems where: 1) the environment reacts continuously with the system; 2) the system evolves in discrete time units; 3) some components may not be fully specified (partial information); and 4) the components react accordingly to stochastic laws.

Our approach allows for building abstract models of the system that are incrementally refined by adding new information. Furthermore, the model can be directly executed in a simulation tool. This is a salient feature for biologists since they can observe the reaction of the system when parameters are adjusted.

We report some results in the use of this method to model Guanine proteins (G-proteins) and Guanine nucleotide-binding protein-coupled receptors (GPCRs). These components are a crucial family of signal transduction molecules that govern a variety of physiological functions. They have been (and continue to be) a major exploitable drug target giving rise to a plethora of clinically relevant molecules. Compositional and extensible modeling tools as the one proposed here may help to understand the fundamental properties of these systems, thus contributing to the future of drug discovery.

2 The modeling language

Nowadays concurrent systems are ubiquitous in several domains and applications. They pervade different areas in science (e.g. biological and chemical systems), engineering (e.g., security protocols and mobile and service oriented computing) and even the arts (e.g. tools for multimedia interaction).

Process calculi such as CCS and the π-calculus [10, 13] among several others have arisen as mathematical formalisms to model and reason about concurrent systems. They treat concurrent processes much like the λ-calculus treats computable

functions. They then provide a language in which the structure of terms represents the structure of processes together with an operational semantics to represent computational steps.

In this paper we shall use as modeling language Temporal Concurrent Constraint Programming (CCP) [15], a model for concurrency that combines the traditional operational view of process calculi with a *declarative* one based upon logic. This combination allows CCP to benefit from the large body of reasoning techniques of both process calculi and logic (see e.g., [15, 11]) .

Agents in CCP *interact* with each other by telling and asking constraints in a global store. Constraints, that can be thought as logic formulae (e.g, $x > 42$), represent (partial) information about the variables of the system. The basic constructs in CCP are **tell**(c) adding the constraint c to the store, thus making it available to the other processes; and the *ask* **when** c **do** P querying if the current store can entail the guard c; if so, it behaves like P. Otherwise it remains blocked until more information is added. This way, ask processes define a synchronization mechanism based on entailment of constraints.

CCP features also constructs for declaring local variables as in $(\textbf{local}\,x)\,P$ and for executing processes in parallel as in $P \parallel Q$. Furthermore, temporal and probabilistic extensions of CCP have been proposed to deal with the notion of discrete time [14] and probabilistic behavior [6]. For instance, it is possible to delay one time unit the execution of P as in **next** P and to choose with a probability ρ (resp. $1 - \rho$) the execution of P (resp. Q) with the construct $P +_\rho Q$.

3 The Modeling Strategy and Preliminary Results

In this section we describe the modeling methodology and some findings in the use of it to model signaling systems of Guanine proteins (G-proteins) and Guanine nucleotide-binding protein-coupled receptors (GPCRs). We tame the complexity of the modeling task through different abstraction levels to focus on particular principles that helps to understand the behavior of the whole system.

We apply a modeling approach for processes at the cellular level describing the biochemical interactions (i.e., a signaling pathways) that operate to convert an external stimuli into an intracellular response. The study of these processes shows how a transmembrane signaling system is regulated at its most basic level. This may help us to better understand how the system will react to the presence of a certain molecule or how the control is distributed in a complex network of biochemical and enzymatic reactions.

The interactions between the components of biological systems, at any level of organization and complexity, can be explained in terms of the level below, giving rise to the function and behavior in the level above. The novelty of the modeling design is the subdivision of the macroscopic view of the signaling system by partial descriptions conditioned on a minimal set of suppositions (constraints) widely applicable as possible. This abstraction-refinement approach is certainly more difficult
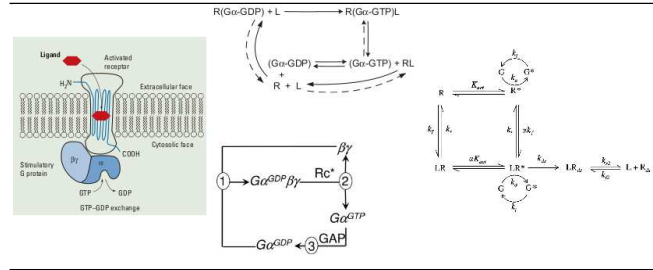
Fig. 1: Interaction Domains of the G-proteins and GPCRs: extracellular (top), transmembrane (middle) and intracellular (bottom). Taken from [4, 8, 12, 5]

to achieve if one uses models based on ordinary differential equations that contain a large number of parameters and that, in general, are not compositional (i.e., it is not trivial to compound subsystems). On the contrary, the idea of partial information as constraints and the underlying rudiments to specify concurrent systems in CCP, make this language appropriate for this aim.

As it was shown in [16], when the fine detail is not completely known, a compositional and scalable modeling strategy serves as a tool for testing hypotheses and generate predictions. These two characteristics are also natural in CCP models: more information can be added to the system by posting constraints and subsystem can be easily composed by sharing variables. In the model of the G-proteins described below, we integrate three different perspectives or points of view of the system. This leads to simple and well defined subsystems that are later integrated to enhance the understanding of the system as a whole.

The G-protein Model. The heterotrimeric G-protein molecule consists of the subunits α, β, and γ (Figure 1). Upon activation, when a free ligand binds to the transmembrane receptor (GPCR), a GDP (Guanosine diphosphate) bound to the α-subunit is exchanged with a GTP (Guanosine triphosphate), and the G-protein dissociates into different subunits which transmit the signal to downstream processes. Once the G_{α}-GTP has dissociated from the $G_{\beta\gamma}$-dimer, it can directly interact with effector proteins to continue the signaling cascade such as adenylyl cyclase (Figure 2). Despite their diversity in function and design, many signaling pathways use the same essential components, which are often highly conserved through evolution and between species [9].

As we said before, our strategy is to integrate different points of view of the system. In the model of the G-protein we study three domains of interactions: *(i)* the extracellular (ED), that is the model of the signaling of G-protein; *(ii)* the transmembrane (TD), that is the model of signaling of the GPCRs including G-protein activation and receptor desensitization, and *(iii)* the intracellular (ID), that is the model for the cycle of the heterotrimeric G-protein.

Following [2], each environmental domain is modeled by a set of stoichiometric equations of the form $a_1 X_1 + ... a_n X_n \dashrightarrow b_1 Y_1 ... + b_m Y_m$ where $X_1 ... X_n$ are reac-
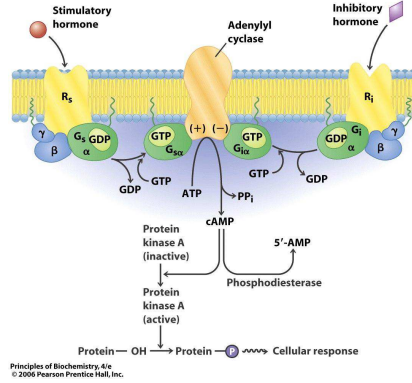
Fig. 2: A typical scenario for control of intracellular metabolic processes. Taken from [7]

tants that interact (and are consumed) yielding to the products $Y_1...Y_m$. Each type of molecule is represented as a variable (e.g., X_i) and equations as CCP processes of the form:

$$\texttt{eq-proc} = \textbf{when } X_1 \geq a_1 \wedge ... \wedge X_n \geq a_n \textbf{ do} \quad \textbf{next}^{(t)}\textbf{tell}(Y_1 = Y_1' + b_1 \wedge ... \wedge Y_m = Y_m' + b_m)$$
$$\| \textbf{ next tell}(X_1 = X_1' - a_1 \wedge ... \wedge X_n = X_n' - a_n)$$

Roughly speaking, when the reactants are available, they are consumed and the right hand components are produced t time units later. The parameter t allows to model kinetic parameters of the system where the speed of reactions may vary.

The set of equations leads to a simple description of the system by means of stoichiometric analysis. We consider equations to describe binding, dissociation, complex formation, and transfer of molecule groups [4, 8, 12, 18].

For example, to represent the biochemical behavior of the intracellular domain, we have the following set of chemical equations that captures the main features of the G-protein signaling cycle:

Reaction for molecular complex formation.	$G\alpha GDP + \beta\gamma \xrightarrow{k_{ass}} G\alpha GDP\beta\gamma$
The transmembrane receptor (actived) promotes GDP/GTP exchange.	$G\alpha GDP\beta\gamma \xrightarrow{Rc^*} G\alpha GTP + \beta\gamma$ $G\alpha GDP\beta\gamma + Rc^* \leftrightharpoons [G\alpha GDP\beta\gamma]Rc^*$ $[G\alpha GDP\beta\gamma]Rc^* \xrightarrow{k_{diss}} G\alpha GTP + \beta\gamma + Rc^*$
Reaction of hydrolysis.	$G\alpha GTP \xrightarrow{GAP} G\alpha GDP$ $G\alpha GTP + GAP \leftrightharpoons [G\alpha GTP]GAP$ $[G\alpha GTP]GAP \xrightarrow{k_{hydr}} G\alpha GDP + GAP + P_i$

Given the set of processes representing the reactions, a probabilistic choice ($+_\rho$) is used to determine the next reaction (whose left hand side components are available) to occur. At present we assume the same probability for each reaction. Nevertheless,

if the propensity of the reactions to occur are known, this information can be easily added to the model by fixing ρ above.

The reader may refer [3] for the complete model of the the intercellular domain as well as for the model of the other two interaction domains.

Preliminary Results. CCP processes can be seen as runnable specifications of a system: the model can be directly simulated in tools as BioWays [3]. We implemented the model above and simulated a million of time units under different kinetic parameters and concentrations of system molecules. We obtained some results similar to those reported in [19] and we describe them in the following. Refer [3] for the complete data and plots obtained with the tool BioWays.

As it is shown in Figure 3 for the trimeric G-protein cycle, the concentrations of signaling components in the environment should be high in order to achieve a steady state. This can be observed in the intracellular domain trough different modes for kinetic parameters. Another issue related to the activation kinetics is that it seems to depend on the expression levels of the activated receptor (green curve in Figure 4).
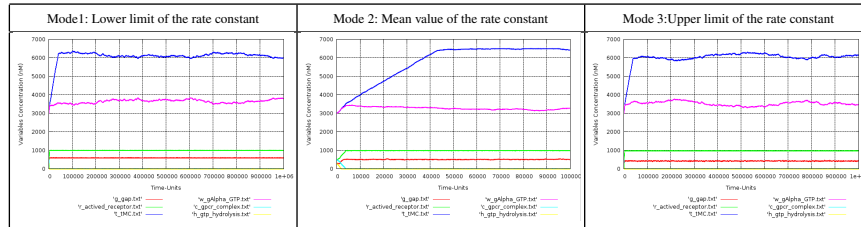

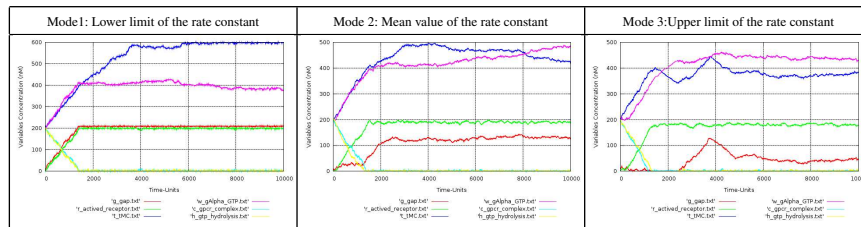Fig. 3: Simulation results for the trimeric G-protein Cycle taking the maximum value of the concentrations

Fig. 4: Simulation results for the trimeric G-protein Cycle taking the lower limit of the concentrations

When we observe the extracellular domain for the reaction scheme of G-protein signaling (see Figure 4 in [3]), it is interesting to observe the high levels of expression of the ligand-receptor complex. This suggests that at least one step in the signal

transduction cascade is mediated via collision coupling between the receptor and the ligand. In this line, the work in [17] proposes a hypothesis for a collision coupling model: a receptor-agonist complex acts as a mobile catalyst for the activation of the signal in the plasma membrane. Our simulation shows that indeed there is no interaction between the receptor and the G-protein in the absence of the ligand.

Simulation results for GPCR signaling, including G-protein activation and receptor desensitization, were obtained under different values of the parameters (see Figures 5-10 in [3]). This provided insights about which parameters are ligand dependent. The ability of a receptor to transduce a response may depend on its active or desensitized state, as well as on the identity of the ligand that is bound. In this sense, the transmembrane domain seems to be more sensible to these changes: any receptor in the R^* or R_{ds} form, being ligand-bound or not, is capable of activating and inactivating G-protein.

A better understanding of the signaling efficacy will provide not only quantitative information but also a qualitative comprehension of cellular responses, to address e.g., the development of improved drug therapies for diseases that involve GPCRs. For this end, the cross viewpoint of our method allows some interpretations of the system to the same stimuli that can be explored systematically. For example, the simulation results shown that the target molecule $G\alpha GTP$ remains relatively constant after activation in the intracellular domain, while the levels of the receptor and the ligand change when we observe the other domains.

The recognition and interpretation of the above mentioned behaviors provide insights about the relationship between G-proteins, GPCRs and ligands. These results are certainly more difficult to obtain with other modeling approaches as differential equations since they cannot be easily composed and they consider the estimation of parameters as a key role more than the interactions of molecules on the scale of cells.

Ongoing Work. Our model focuses on qualitative patterns (supported by quantitative information) of the time evolution of the key components. If the molecules concentrations and rate constants were widely available (fully experimental sources), the model parameters could be re-estimated in an easy and modular way to fit experimental data, thus obtaining a predictive model.

We are currently working on models of the control system of intracellular metabolic processes for the signaling pathway of glycogen breakdown (Figure 2). This signal transduction system consists of three protein components, a receptor, a transducer, and an effector.

In the response of liver cells to glucagon, the interaction of these components stimulate adenylyl cyclase, a transmembrane protein which catalyzes the conversion of ATP (Adenosine triphosphate) into cyclic AMP (Cyclic adenosine monophosphate), an intracellular second messenger. The binding of the ligand at the cell surface stimulates synthesis of a second messenger inside the cell, which is a desirable metabolic response.

Glycogen is the most immediately available large-scale source of metabolic energy in living beings. Moreover, glycogen breakdown is a hormone controlled process well studied in the literature. Our aim is to find some principles of the G-protein

cycle activation in a complete context: the refinement taking into account the model of the conversion of the ATP into cyclic AMP will allow us to observe the dynamic of the glycogen breakdown under the effect of the interaction between G-protein and GPCR.

References

1. Anand R. Asthagiri and Douglas A. Lauffenburger. Bioengineering models of cell signaling. *Annual Review of Biomedical Engineering*, 2:31–53, 2000.
2. D. Chiarugi, M. Falaschi, C. Olarte, and C. Palamidessi. Compositional modelling of signalling pathways in timed concurrent constraint programming. In *Proc.of ACM BCB'10*, pages 414–417, New York, NY, USA, 2010. ACM.
3. Davide Chiarugi, Diana Hermith, Moreno Falaschi, and Carlos Olarte. Bioways: A ccp-based tool for simulating biological systems. Available at http://www.dianahermith.com/bioways/.
4. David Csercsika, Katalin M. Hangosa, and Gyrgy M. Nagy. A simple reaction kinetic model of rapid (g protein dependent) and slow (betha-arrestin dependent) transmission. *Journal of Theoretical Biology*, 255(1):119–128, 2008.
5. David Filmore. Cell-based screening assays and structural studies are fueling g-protein coupled receptors as one of the most popular classes of investigational drug targets. *Modern Drug Discovery*, 7(11), 2004.
6. Vineet Gupta, Radha Jagadeesan, and Vijay A. Saraswat. Probabilistic concurrent constraint programming. In *Proc. of CONCUR 97*, London, UK, 1997. Springer-Verlag.
7. Robert Horton, Laurence A. Moran, Gray Scrimgeour Marc Perry, and David Rawn. *Principles of Biochemistry*. Prentice Hall, 4th edition, 2006.
8. Vladimir L. Katanaev and Matey Chornomorets. Kinetic diversity in g-protein-coupled receptor signalling. *Biochem J.*, 401(2):485–495, 2006.
9. Edda Klipp and Wolfram Liebermeister. Mathematical modeling of intracellular signaling pathways. *BMC Neurosciences*, 7(1):1–16, 2006.
10. Robin Milner. *Communicating and Mobile Systems: the Pi-Calculus*. Cambridge University Press, 1999.
11. M. Nielsen, C. Palamidessi, and F.D. Valencia. Temporal concurrent constraint programming: Denotation, logic and applications. *Nordic Journal of Computing*, 9(1), 2002.
12. Todd A. Riccobene, Geneva M. Omann, and Jennifer J. Linderman. Modeling activation and desensitization of g-protein coupled receptors provides insight into ligand efficacy. *Journal of Theoretical Biology*, 200(2):207–222, 1999.
13. D. Sangiorgi and D. Walker. *The π-calculus: a Theory of Mobile Processes*. Cambridge University Press, 2001.
14. Vijay Saraswat, Radha Jagadeesan, and Vineet Gupta. Foundations of timed concurrent constraint programming. In *Proc. of LICS'94*. IEEE CS, 1994.
15. Vijay A. Saraswat. *Concurrent Constraint Programming*. MIT Press, 1993.
16. S. Schnell, R. Grima, and P. K. Maini. Multiscale modeling in biology. *American Scientist*, 95(1):134–142, 2007.
17. Lonnie Shea and Linderman Jennifer J. Mechanistic model of g-protein signal transduction, determinants of efficacy and effect of precoupled receptors. *Biochem Pharmacol.*, 53(4):519–30, 1997.
18. Eduardo D. Sontag. Molecular systems biology and control. In *Control 11:396-435. of Boolean networks 15*, 2005.
19. Tau-Mu Yi, Hiroaki Kitano, and Melvin I. Simon. A quantitative characterization of the yeast heterotrimeric g protein cycle. *PNAS*, 100(19):10764–10769, 2003.