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. We describe a modeling strategy based on a probabilistic temporal concurrent constraint (CCP) calculus. 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 CCP; finally, the model is a runnable specification and can be executed. 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.

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 a probabilistic temporal concurrent constraint language that allows the modeling of reactive systems where: 1) the environment reacts continuously with the system; 2) the system evolves in discrete time units; 3) there is no a complete description of the some components (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 of our 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 as biological systems, security protocols, mobile systems, etc. In general, concurrent systems exhibit complex forms of interaction, not only among their internal components, but also with the surrounding environment. In computer science, process calculi have arisen as mathematical formalisms to model and reason about concurrent systems.

We shall use as modeling language Concurrent Constraint Programming (CCP) [9], a model for concurrency that combines the traditional operational view of process calculi with a declarative one based upon logic. CCP has successfully been used in the modelling and verification of several concurrent scenarios, e.g., timed, reactive and stochastic systems [9, 6].

Agents in CCP interact with each other by telling and asking constraints in a global store. Constraints (e.g., \( x > 42 \)) represent (partial) information about the variables of the system. The basic constructs in CCP are \textbf{tell}(c)\(^1\) adding the constraint \( c \) to the store, thus making it available to the other processes; and the \textbf{ask when} \( c \) \textbf{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 (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 [8] and probabilistic behavior [4]. For instance, it is possible to delay one time unit the execution of $P$ as in next$P$ and to choose with a probability $\rho$ (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 G-proteins and GPCRs. We tame the complexity of the modeling task through different abstraction levels. This allows us to focus on particular principles that helps to understand the behavior of the system.

The most simple picture of the system is the cell-surface receptor, the ligand, the G-proteins components, and other supporting molecules interacting in three environmental domains (Figure 1a). The extracellular domain (ED) is the model of the signaling of G Protein. The transmembrane domain (TD) is the model of signaling of the GPCRs including G Protein activation and receptor desensitization. The intracellular domain (ID) is the model for the cycle of the heterotrimeric G Protein.

Each environmental domain is modeled by a set of stoichiometric equations of the form $a_1X_1 + \ldots + a_nX_n \rightarrow b_1Y_1 + \ldots + b_mY_m$ where $X_1 \ldots X_n$ are reactants that interact (and are consumed) yielding to the products $Y_1 \ldots Y_m$. Each type of molecule is represented as a variable (e.g., $X_i$) and equations as CCP processes of the form:

$$eq-proc = when X_1 \geq a_1 \wedge \ldots \wedge X_n \geq a_n \ do \ next(tell(Y_1 = Y_1' + b_1 \wedge \ldots \wedge Y_m = Y_m' + b_m) \parallel next(tell(X_1 = X_1' - a_1 \wedge \ldots \wedge X_n = X_n' - a_n)$$
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 us to model systems where the speed of reactions may vary (see [2] for the complete CCP model).

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 [3, 5, 7, 10]. Besides, when rates for a reaction to occur are known, they can be added to the model by choosing probabilistically (+\( p \)) the reaction to be applied.

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 to achieve if we one uses models based on ordinary differential equations that contain a large number of parameters.

**Simulating the Model.** CCP processes can be seen as runnable specifications of a system: the model can be directly simulated in tools as BioWays [2]. We implemented the model above and simulated a million of time units under different kinetic parameters and concentrations of system molecules (see [2] for details).

The Figure 1b shows some results similar to those reported in [11]. For this case, there are six species in the graph. \( G Protein \) (turquoise) represents that the \( \alpha \) unit and the \( \beta\gamma \) complex are bound and the \( \alpha \) unit is bound to \( GDP \) in the ID. \( RL \) (in green for TD and purple for ED) represents that the receptor (GPCRs) and ligand are bound. \( G Protein \) (red for TD and yellow for ID) represents that the \( \alpha \) unit is separated from the \( \beta\gamma \) complex and is bound to \( GTP \) (G Protein activation). The curves of \( G\alpha GTP \) production considering the TD and ED, decrease oppositely to the increase of G Protein complex in ID. \( RL \) in TD decreases below G-protein in the ID-viewpoint. \( G\alpha GTP \) is keeping relatively constant in the ID.

This cross viewpoint allows some interpretations of the system to the same stimuli, that can be explored systematically. Although our target molecule \( G\alpha GTP \) remains relatively constant after activation, the interaction between receptor and ligand changes when we look inside each domain. The recognition and interpretation of these behaviors could provide some insights into the relationship between G-Proteins, GPCRs and ligands.

**Ongoing Work.** Our model focuses on the characteristic qualitative patterns (supported by quantitative information) of the time evolution of the key components. If the molecules concentrations and rate constants were 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. Here CCP may be helpful since each sub-system can be altered locally without modifying other components.

We are currently working on models of the control system of intracellular metabolic processes for the signaling pathway of glycogen breakdown. This aims at finding some principles of the G-Protein cycle activation in a complete context.
References


