ABSTRACT

Most biological functions are mediated by protein interactions. These interactions can be physical, such as when two proteins form a complex, or “logical,” such as when one or more proteins control the behavior of one or more other proteins without physical interaction. Metabolic pathways provide us with many examples of these kind of interactions. These molecules have an extracellular domain, a membrane domain, and an intracellular domain. The association of the corresponding intracellular domains allows the molecule to transmit the signal inside the cell. The number of proteins and other molecules participating in the events involving signal transduction increases as the process emanates from the initial stimulus, resulting in a "signal cascade," beginning with a relatively small stimulus that elicits a large response.

We are planning gain a comprehensive and predictive understanding of the dynamic, interconnected processes underlying living systems, in these case, the G Protein Signal Cascade, with a NTCC model that could resolve some biological queries. The first goal would be, identify and characterize the biomolecular machinery of G Protein Signal Cascade as concurrent computation, using process algebra: the NTCC calculus. The second one, develop the computational capabilities to advanced understanding of complex biological systems and predict their behavior.
I. INTRODUCTION

Most of biological functions are mediated by protein interactions. These interactions can be physical, such as when two proteins form a complex, or “logical,” such as when one or more proteins control the behavior of one or more other proteins without physical interaction. Examples of physical interactions are stable complexes, in which the functional unit is formed by more than one protein chain, as in the case of the glycogen phosphorylase enzyme, and transient associations, in which the protein chains are stable by themselves but can also interact to transmit a signal or as a response to external conditions. In logical interactions, one protein affects another protein by, for example, regulating its expression or changing the concentration of a factor that, in turn, is sensed by the target protein. The two modes of interaction are not exclusive. The same proteins can interact both physically and logically.\(^1\)

Metabolic pathways provide us with many examples of logical interactions. The concentration of a product is often “sensed” by other proteins in its synthetic cascade and modulates their activity. The presence of hormones is detected by cell surface receptors and transmitted to other proteins in the cell that can interact with the genetic material to activate or repress genes. These logical interactions can coexist with physical interactions. For example, hemoglobin senses the binding of oxygen and transmits the information from one of its subunits to the others via physical interaction. Other examples can be found in cell surface receptors. These molecules have an extracellular domain, a membrane domain, and an intracellular domain. Binding of a ligand to the extracellular domain can cause these molecules to form

dimers (i.e., to associate with another receptor chain). The association of the corresponding intracellular domains allows the molecule to transmit the signal inside the cell (Figure 1).

**Fig. 1.** The binding of a ligand to the extracellular domain of a transmembrane receptor might cause its binding to a coreceptor (which can be the same or a different protein). The subsequent interaction between the intracellular domains can trigger signaling, for example, by activating a transcription factor that, in turn, activates the required genes.

In biology, signal transduction refers to any process by which a cell converts one kind of signal or stimulus into another. Most processes of signal transduction involve ordered sequences of biochemical reactions inside the cell, which are carried out by enzymes, activated by second messengers, resulting in a signal transduction pathway. Such processes are usually rapid, lasting on the order of milliseconds in the case of ion flux, or minutes for the activation of protein- and lipid-mediated kinase cascades, but some can take hours, and even days (as is the case with gene expression), to complete. The number of proteins and other molecules participating in the events involving signal transduction increases as the process emanates from the initial stimulus, resulting in a "signal cascade," beginning with a relatively small stimulus that elicits a large response. This is referred to as amplification of the signal.²

An environmental signal, such as a hormone, is first received by interaction with a cellular component, most often a cell-surface receptor. The information that the signal has arrived is then converted into other chemical forms, or transduced. The signal is often amplified before evoking a response. Feedback pathways regulate the entire signaling process.

**Fig 2. Signal Transduction**

² http://en.wikipedia.org/wiki/Signal_transduction
II. STATE OF THE ART (Short)

Cell Signaling or Signal Transduction, is the study of the mechanisms that enable the transfer of biological information. Signaling impinges on all aspects of biology, from development to disease. Many diseases, such as cancer, involve malfunction of signal transduction pathways. Mathematical modeling and simulation in this field has the purpose to help and guide the biologist in designing experiments and generally to establish a conceptual framework in which to think (Kitano et al, 2003).

G-proteins represent a crucial family of signal transduction molecules that govern a variety of physiological functions. Moreover, GPCRs have traditionally been (and continue to be) a major exploitable drug target, giving rise to a plethora of clinically relevant molecules. Thus, a more complete understanding of the fundamental properties of GPCRs and how they interact with, and activate, their target G-proteins is of utmost importance to future drug discovery (Johnston et al, 2006).

How GPCRs operate is one of the most fundamental questions in the field of transmembrane signal transduction.

Thus, in addition to our curiosity about the fascinating mechanism that cells use to respond to signals, there is practical motivation to better understanding the processes of cellular signaling, in which protein-protein interactions play a central role.

The behavior of a signal-transduction system depends on dynamic interactions among its proteins. The combined effects of these interactions are difficult to predict from intuition alone. When intuition is insufficient, a mathematical model is often useful for acquiring a quantitative and predictive understanding of a complex dynamical system, and mathematical modeling is being increasingly used to aid in studies of cellular signaling. However, current models are still far from capturing all of the relevant mechanistic details of signal-transduction systems that must be considered to provide realistic and complete pictures of how these systems work. In particular, models often fail to account for the complexities of protein-protein interactions, such as how these interactions depend on contextual details at the level of protein sites. New modeling approaches that address this problem involve the use of rules to represent protein-protein
interactions, rules are also useful for representing other types of biomolecular interactions.

The introduction of rules greatly eases the task of specifying a model that incorporates details at the level of protein sites. A rule—such as “ligand binds receptor with rate constant k whenever ligand and receptor have free binding sites”—describes the features of reactants that are required for a particular type of chemical transformation to take place. Rules simplify the specification of a model when the reactivity of a component in a system is determined by only a subset of its possible features (Hlavacek et al, 2006).

Other authors propose that the concurrency paradigm and the pi calculus theory are uniquely suited to model and study biomolecular processes in general and Signaling Transduction pathways in particular. Within the particular framework of the pi calculus, they set three principles for this correspondence; first, as primitive process, they choose the functional signaling domain. Second, they model the component residues of domains as communication channels that construct a process. Finally, molecular interaction and modification is modeled as communication and the subsequent change of channel names. Based on these three principles the pi calculus allows to fully represent complex molecular structures and signaling events (Table 1) (Shapiro et al, 2000).

<table>
<thead>
<tr>
<th>SI</th>
<th>Modeling analogy</th>
<th>pi calculus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular integrity</td>
<td>Reactive chemical reactions accessible only within the scope of the molecule</td>
<td>The non-operator ( \text{MULTIPLE} ) := new channels ( \text{DOM}<em>{1} ) : ( \text{DOM}</em>{2} )</td>
</tr>
<tr>
<td>Binding and complex formation</td>
<td>Two-way communicativity of global channel names and two-way passing of private channel names</td>
<td>The mixed-channel operator ( (.) ) := ( \text{DIHESION, DOM}<em>{1} ) ( \rightarrow ) ( \text{BINDING, MOUTH, dom}</em>{2} ) ( \rightarrow ) ( \text{BINDING, MOUTH, dom}<em>{2} ) ( \rightarrow ) ( \text{BINDING, MOUTH, dom}</em>{2} )</td>
</tr>
<tr>
<td>Ordered regulation</td>
<td>One communication event strictly preceding another</td>
<td>The prefix operation ( \text{INACTIVATION, DOM}<em>{1} ) := ( \text{INACTIVATION, DOM}</em>{1} ) ( \rightarrow ) ( \text{RESIDUE, dom}<em>{2} ) ( \rightarrow ) ( \text{RESIDUE, dom}</em>{2} ) ( \rightarrow ) ( \text{RESIDUE, dom}_{2} )</td>
</tr>
<tr>
<td>Enzymatic activity</td>
<td>A reacting communication sequence: binding, modification and recombination</td>
<td>A guarded recursive definition ( \text{ACTIVE, SITE} ) := ( \text{INACTIVATION, DOM}<em>{1} ) ( \rightarrow ) ( \text{SITE, dom}</em>{2} ) ( \rightarrow ) ( \text{SITE, dom}<em>{2} ) ( \rightarrow ) ( \text{SITE, dom}</em>{2} )</td>
</tr>
</tbody>
</table>

Table 1. Pi calculus modeling of typical molecular structures involved in Signaling Transduction Pathways and key signaling events.

### III. DETAILED DESCRIPTION OF THE G PROTEIN SIGNAL CASCADE

Most signal molecules targeted to a cell bind at the cell surface to receptors embedded in the plasma membrane. Only signal molecules able to cross the plasma

---

3 http://www.rpi.edu/dept/bcbp/molbiochem/MBWeb/mb1/part2/signals.htm#animat1
membrane (e.g., steroid hormones) interact with intracellular receptors. A large family of cell surface receptors have a common structural motif, 7 transmembrane α-helices. Rhodopsin was the first of these to have its 7-helix structure confirmed by X-ray crystallography.

Rhodopsin is unique. It senses light, via a bound chromophore, retinal. Most 7-helix receptors have domains facing the extracellular side of the plasma membrane that recognize and bind signal molecules (ligands). For example, the β-adrenergic receptor is activated by epinephrine and norepinephrine.

The signal is usually passed from a 7-helix receptor to an intracellular G-protein. Seven-helix receptors are thus called GPCR, or G-Protein-Coupled Receptors. Approx. 800 different GPCRs are encoded in the human genome.

G-protein-Coupled Receptors may dimerize or form oligomeric complexes within the membrane. Ligand binding may promote oligomerization, which may in turn affect activity of the receptor. Various GPCR-interacting proteins (GIPs) modulate receptor function. Effects of GIPs may include: altered ligand affinity, receptor dimerization or oligomerization, control of receptor localization, including transfer to or removal from the plasma membrane, promoting close association with other signal proteins.

G-proteins are heterotrimeric, with 3 subunits α, β, γ. A G-protein that activates cyclic-AMP formation within a cell is called a stimulatory G-protein, designated Gs with alpha subunit Gsα. Gs is activated, e.g., by receptors for the hormones epinephrine and glucagon. The β-adrenergic receptor is the GPCR for epinephrine.

The α subunit of a G-protein (Gsα) binds GTP, and can hydrolyze it to GDP + P_i. α

Fig. 3 The G Protein Signal Cascade

and γ subunits have covalently attached lipid anchors that bind a G-protein to the plasma membrane cytosolic surface. Adenylate Cyclase (AC) is a transmembrane protein, with cytosolic domains forming the catalytic site.

The sequence of events by which a hormone activates cAMP signaling include the following phases:

Phase 1. Initially G_α has bound GDP, and α β and γ subunits are complexed together. G_βγ, the complex of β and γ subunits, inhibits G_α.

Phase 2. Hormone binding, usually to an extracellular domain of a 7-helix receptor (GPCR), causes a conformational change in the receptor that is transmitted to a G-protein on the cytosolic side of the membrane. The nucleotide-binding site on G_α becomes more accessible to the cytosol, where [GTP] > [GDP]. G_α releases GDP and binds GTP (GDP-GTP exchange).

Phase 3. Substitution of GTP for GDP causes another conformational change in G_α. G_α-GTP dissociates from the inhibitory βγ complex and can now bind to and activate Adenylate Cyclase.

Phase 4. Adenylate Cyclase, activated by the stimulatory G_α.-GTP, catalyzes synthesis of cAMP.

Phase 5. Protein Kinase A (cAMP Dependent Protein Kinase) catalyzes transfer of phosphate from ATP to serine or threonine residues of various cellular proteins, altering their activity.

The turn off of the signal involves these kind of possibilities:

P1. G_α hydrolyzes GTP to GDP + P_i. (GTPase). The presence of GDP on G_α causes it to rebind to the inhibitory bg complex. Adenylate Cyclase is no longer activated.

P2. Phosphodiesterases catalyze hydrolysis of cAMP → AMP.

P3. Receptor desensitization varies with the hormone. In some cases the activated receptor is phosphorylated via a G-protein Receptor Kinase. The phosphorylated receptor then may bind to a protein β-arrestin. β-Arrestin promotes removal of the receptor from the membrane by clathrin-mediated endocytosis. β-Arrestin may also bind a cytosolic
Phosphodiesterase, bringing this enzyme close to where cAMP is being produced, contributing to signal turnoff.

P4. Protein Phosphatase catalyzes removal by hydrolysis of phosphates that were attached to proteins via Protein Kinase A.

The signal amplification is an important feature of signal cascades. One hormone molecule can lead to formation of many cAMP molecules. Each catalytic subunit of Protein Kinase A catalyzes phosphorylation of many proteins during the life-time of the cAMP. Different isoforms of $G_\alpha$ have different signal roles. For example, the stimulatory $G_{s\alpha}$, when it binds GTP, activates Adenylate cyclase. An inhibitory $G_{i\alpha}$, when it binds GTP, inhibits Adenylate cyclase. Different effectors and their receptors induce $G_{i\alpha}$ to exchange GDP for GTP than those that activate $G_{s\alpha}$. The complex of $G_{B\alpha}$ that is released when $G_\alpha$ binds GTP is itself an effector that binds to and activates or inhibits several other proteins. For example, $G_{B\alpha}$ inhibits one of several isoforms of Adenylate Cyclase, contributing to rapid signal turnoff in cells that express that enzyme.

IV. WHY TO DEVELOP A MODEL BY USING NTCC CALCULUS?

Partial information arises naturally in the description of biological systems. It is possible to distinguish two main kinds of partial information when modeling those systems: quantitative and behavioral. While partial quantitative information usually involves incomplete information on the state of the system (e.g., the set of possible values that a variable can take), partial behavioral information refers to the uncertainty associated to behavior of interactions (e.g., the unknown relative speeds on which two systems interact). Finding precise ways of expressing these kinds of partial information can help to better understand complex pattern behaviors, frequent in biological systems. Partial information is a central feature of Concurrent Constraint Programming (CCP), a well-established formalism for concurrency. In CCP, processes interact with each other by telling and asking partial information represented as constraints (e.g., $x < 42$). Perhaps the most appealing and distinctive feature of CCP is that it combines the traditional operational view of process calculi with a declarative one based upon logic. In other words, the process terms can be viewed at the same time as
computing agents and logic formulas. This combination allows CCP to benefit from the large body of techniques of both process calculi and logic. For these reasons CCP can be a convenient framework to describe and reason about biological systems. In this paper we propose works with ntcc, a timed process calculus based on CCP, as a suitable language for analyzing biological systems. In ntcc the above-mentioned kinds of partial information are naturally captured. On the one hand, partial quantitative information is captured by the notion of constraint system, a structure that gives coherence and defines (logic) inference capabilities over constraints. Since constraint systems are parametric to ntcc, by choosing the appropriate constraint system(s) several kinds of conditions, at different levels of detail, can be stated. This could be particularly useful in the description of quantitative information. For instance, one could think of a constraint system over differential equations interacting with others over, say, integers or real intervals. On the other hand, partial behavioral information is represented by non-deterministic and asynchronous operators available in ntcc. The interplay of these operators in the discrete time of ntcc allows to explicitly describe and reason about the uncertainty in the occurrence time of many biological phenomena (Gutierrez et al, 2006).

Signal-transduction pathways can be viewed as a molecular circuit. We begin by examining the challenges posed by transferring extracellular information to a cell's interior:

Phase 1. Membrane receptors transfer information from the environment to the cell's interior. A few nonpolar signal molecules such as estrogens and other steroid hormones are able to diffuse through the cell membranes and, hence, enter the cell. Once inside the cell, these molecules can bind to proteins that interact directly with DNA and modulate gene transcription. Thus, a chemical signal enters the cell and directly alters gene-expression patterns.

However, most signal molecules are too large and too polar to pass through the membrane, and no appropriate transport systems are present. Thus, the information that signal molecules are present must be transmitted across the cell membrane without the molecules themselves entering the cell. A membrane-associated receptor
protein often performs the function of information transfer across the membrane.

Such a receptor is an intrinsic membrane protein that has both extracellular and intracellular domains. A binding site on the extracellular domain specifically recognizes the signal molecule (often referred to as the ligand). Such binding sites are analogous to enzyme active sites except that no catalysis takes place within them. The interaction of the ligand and the receptor alters the tertiary or quaternary structure of the receptor, including the intracellular domain. These structural changes are not sufficient to yield an appropriate response, because they are restricted to a small number of receptor molecules in the cell membrane. The information embodied by the presence of the ligand, often called the primary messenger, must be transduced into other forms that can alter the biochemistry of the cell.

Phase 2. Second messengers relay information from the receptor-ligand complex. Changes in the concentration of small molecules, called second messengers, constitute the next step in the molecular information circuit. Particularly important second messengers include cyclic AMP and cyclic GMP, calcium ion, inositol 1,4,5-trisphosphate, (IP3), and diacylglycerol.

The use of second messengers has several consequences. First, second messengers are often free to diffuse to other compartments of the cell, such as the nucleus, where they can influence gene expression and other processes. Second, the signal may be amplified significantly in the generation of second messengers. Enzymes or membrane channels are almost always activated in second-messenger generation; each activated macromolecule can lead to the generation of many second messengers within the cell. Thus, a low concentration of signal in the environment, even as little as a single molecule, can yield a large intracellular signal and response. Third, the use of common second messengers in multiple signaling pathways creates both opportunities and potential problems. Input from several signaling pathways, often called cross talk, may affect the concentrations of common second messengers. Cross talk permits more finely tuned regulation of cell activity than would the action of individual independent pathways. However, inappropriate cross talk can cause second messengers to be misinterpreted.
Phase 3. Protein phosphorylation is a common means of information transfer. Many second messengers elicit responses by activating protein kinases. These enzymes transfer phosphoryl groups from ATP to specific serine, threonine, and tyrosine residues in proteins.

We previously encountered the cAMP-dependent protein kinase. This protein kinase and others are the link that transduces changes in the concentrations of free second messengers into changes in the covalent structures of proteins. Although these changes are less transient than the changes in secondary-messenger concentrations, protein phosphorylation is not irreversible. Indeed, protein phosphatases are enzymes that hydrolytically remove specific phosphoryl groups from modified proteins.

Phase 4. The signal is terminated. Protein phosphatases are one mechanism for the termination of a signaling process. After a signaling process has been initiated and the information has been transduced to affect other cellular processes, the signaling processes must be terminated. Without such termination, cells lose their responsiveness to new signals. Moreover, signaling processes that fail to be terminated properly may lead to uncontrolled cell growth and the possibility of cancer.

As we shall see, the use of protein modules in various combinations is a clear, even dominant, theme in the construction of signal-transduction proteins. Signal-transduction proteins have evolved by the addition of such ancillary modules to core domains to facilitate interactions with other proteins or cell membranes. By controlling which proteins interact with one another, these modules play important roles in determining the wiring diagrams for signal-transduction circuits.

(Bockaert, 1999) G protein coupled receptors (GPCRs; 7TM receptors; seven transmembrane domain receptors; heptahelical receptors; G protein-linked receptors [GPLR]) are the largest family of transmembrane receptors in humans, accounting for more than 1% of the protein coding capacity of the human genome. All known GPCRs share a common architecture of seven membrane spanning helices connected by intra and extracellular loops. The extracellular loops contain two highly conserved cysteine residues that
form disulphide bonds to stabilize the structure of the receptor. They recognize diverse messengers such as light, odorants, small molecules, hormones and neurotransmitters. Most GPCRs act as guanine nucleotide exchange factors; activated by ligand binding, they promote GDP GTP exchange on associated heterotrimeric guanine nucleotide binding (G) proteins. These in turn activate effector enzymes or ion channels. GPCRs are involved in a range of physiological roles which include the visual sense, smell, behavioural regulation, functions of the autonomic nervous system and regulation of the immune system and inflammation. GPCRs are divided into 6 classes based on sequence homology and functional similarity; the main mammalian families are classes A/1 C/3. (From www.reactome.com)

Class A/1 (Rhodopsin like)
Class B/2 (Secretin receptor family)
Class C/3 (Metabotropic glutamate/pheromones)
Class D/4 (fungal mating pheromone receptors)
Class E/5 (cAMP receptors)
Class F/6 (Frizzled/Smootherened)

Signal transduction pathways allow cells to respond to environmental signals. In these pathways, a signal is amplified such that each step in the pathway results in a large number of activated components than in the previous step. This phenomenon, called signal amplification, caused the liver cell for example, to release a significant number of glucose molecules after detecting just a single molecule of epinephrine. Amplification can occur at many points in the pathway. For example, as long as epinephrine remains bound to the receptor, the receptor can activate a succession of G proteins. In addition, each adenylyl cyclase enzyme can convert numerous ATPs into cyclic AMP molecules. Other activated enzymes in the pathway can also continually catalyze reactions. One component that activates just a single enzyme, however, is the G protein. A G protein must remain attached to the adenylyl cyclase enzyme in order for the enzyme to remain activated.

Termination of the cellular response is an important as its initiation. In order for a cell to respond only when a signal is present, the many players in the pathway have to regulated so that they are activate for only a short period of time.
We are planning, gain a comprehensive and predictive understanding of the dynamic, interconnected processes underlying living systems, in these case, the G Protein Signal Cascade. For these purpose, is necessary to have information about dynamics, molecular structure and biochemical detail of each interaction, to explore and apply formal semantics to simulate, analyze and compare the biomolecular G Protein Signal Cascade System. So in these first draft, we not yet include, the computational model, but we now, that the molecules involves in the signal pathways have interaction capability, reactions/interactions and modification, with the same principles came from chemistry, organic chemistry, enzymatic reactions, metabolic pathways, signal-transduction pathways and ultimately the entire cell. Biology is driven by quantities (e.g., energy, time, affinity, distance, amount of components) so we would need to consider this.

The goal 1 would be, identify and characterize the biomolecular machinery of G Protein Signal Cascade as concurrent computation, using Process algebra: the NTCC calculus. The second one, develop the computational capabilities to advance understanding of complex biological systems and predict their behavior.

Previous abstractions shows, that in concurrent computational processes, each biological entity is a process that may carry some state and interacts with other processes. Prior proposals\(^5\) based on process algebras, such as the pi calculus can be applied for make an appropriate reasoning about this kind of systems.

The benefits of the NTCC approach should be an unified view of the system, the simulation and analysis and a comparative power and scalability to enrich the model with experimental (quantitative) data. The idea in this sense, is to explore and combine the methodology with implementations than permit an interplay between collecting data in experiments (experimental biology) and the NTCC model, with the aim to capture some mechanistic understanding of how the systems under study works, by validating the model under various conditions that correspond to the experiments and by comparing the outcomes to the experimental data, one can identify discrepancies between hypothetical

mechanisms and the experimental observations. These differences can be used to suggest new hypotheses, which serve to adjust the model and need to be validated experimentally, or new experiments, which can confirm or falsify the modeling hypotheses.

The goal will be find an appropriate model for G Protein Signal Cascade that include molecular structure, behavior and biological formal semantics.

Expected results we are thinking to obtain: a unified view of structure and dynamics of G Protein Signal Cascade, a detailed molecular information (complexes, molecules, domains, residues) in visible form, a complex dynamic behavior (feedback, cross-talk, split and merge), a modular system.

We want construct a NTCC model that could resolve some of these biological queries: describe the sequence of events by which a hormone such as epinephrine or glucagon activates production of cyclic AMP within a cell; include the roles of the receptor (GPCR), the different subunits of the stimulatory G protein, and Adenylate Cyclase. How is the signal turned off at each step?. How we can simulate the role of β-arrestin?. How we can describe the activation of cAMP-Dependent Protein Kinase (Protein Kinase A)?. What causes the enzyme to be inhibited in the absence of cyclic AMP?. How is activation by cyclic AMP turned off?. What reaction is catalyzed by the enzyme Protein Phosphatase?.

REFERENCES


