The Concurrency Column

by

Luca Aceto

BRICS, Department of Computer Science
Aalborg University, 9220 Aalborg Ø, Denmark
luca@cs.auc.dk, http://www.cs.auc.dk/~luca/BEATCS

One of the most exciting recent applications of process calculi is in the field of *systems biology*. Paraphrasing the title of an excellent survey paper by Christos Papadimitriou, this is an example of the "extroverted" use of concurrency theory in a fast-developing science, where notions from our research area may, perhaps surprisingly, be even more appropriate than in software engineering. Possibly, the main "grand challenge" in this field is to define a set of basic and general primitives for modelling biological systems that are inspired by actual biological processes, and to use these primitives to develop an executable model based on process calculi of some (very) complex biological system. This model could ultimately be used to predict the behaviour of living matter.

This new direction of research in concurrency theory is being pursued by some of the most capable scientists in our field, and there is already a growing body of literature reporting on their work, and appearing in specialized conferences and workshops. I trust that newcomers to this exciting research area will find the present contribution to the Concurrency Column a very useful guide to the literature on the use of process calculi in systems biology. The authors of this survey are amongst the prime movers behind the surge of interest in the use of tools from concurrency theory in biology, and they offer a very thoughtful account of the process algebraic languages proposed so far for the description of biological phenomena, discuss their faithfulness to the biological context, and provide a list of interesting lines of research for the future. Enjoy it!

PROCESS CALCULI IN A BIOLOGICAL CONTEXT

D. PRANDI and C. PRIAMI and P. QUAGLIA
University of Trento

1 Introduction

The recent progress of molecular biology has made possible the detailed description of the components that constitute living systems, notably genes and proteins, as isolated entities. Biological molecules do not live alone. Rather, they participate in very complex networks that are involved in the maintenance as well as the differentiation of cellular systems (consider, e.g., regulatory networks for gene expression). The new challenge is to scale up from molecular biology to *systems biology*, namely to understand how these individual components integrate to take part into complex systems, and how they function, evolve, and interact together.

Molecular biologists use information technology to process, analyze, compare and share scientific knowledge. It is "easy" for a scientist to study a chosen sequence, archiving, editing it, comparing it with other sequences, and analyzing its various properties. The use of widely distributed databases is now common in the scientific community. The next step is to extend this approach for studying metabolic networks, signaling pathways, regulatory circuits or even an entire integrated cell. The goal is to combine experimental data with advanced formal theories from computer science to design formal languages for the specification of interacting molecular entities. Troubles come from the different "conventions" assumed by the life science and computer science communities.

The biological approach clarifies, for example, components, proteins and cells. It also gives graphical, and easily readable, representations of the interactions among the above entities. All these aspects, however, are far from being formally defined. Formal definitions and mathematical foundation of descriptions are mandatory requirements to enhance the understanding of complex biological systems and to allow automatic analysis and simulation. Computer science modelling, on its side, is specifically designed for formal analysis and simulation, but it heavily uses mathematical symbolism that is not easy to read for a neophyte of these formalisms (e.g. a biologist).

The above considerations result in two main principles:

1. there is the need to construct some formal tools to represent, simulate, and analyze biological experiments;

2. it is mandatory to hide from users of these formal tools as many technical details as possible, e.g., by means of suitable abstractions.

Various computational approaches have been proposed and are being actively developed to model and study molecular networks. According to some authors (see, e.g., [22]), it is becoming clearer and clearer that concurrency theory and process calculi could be useful for formally specifying the living matter and its behavior. Processes are the basic units of these languages: they have internal states and interaction capabilities. When a process receives an input its behaviour is based on its internal state and on the content of the input. Consequently it can change its internal state and hence its interaction capabilities. Complex entities, as protein complexes, are described hierarchically, and this allows for either top-down or bottom-up analysis. In this respect, process calculi offer two main advantages:

- compositionality, namely the ability of combining two independent processes with the language constructors obtaining a new process. This allows for a useful bottom-up approach in the design of experiments;

- equivalence relations that are typically well-assessed and could be powerful tools for biology. For example, the equivalence of related implementations of biological mechanisms in different organisms could be used as a measure of behavioral and structural similarity.

This paper offers a survey of the process calculi for systems biology. We avoid proposing a long and uninformative list of languages with their sintax and semantics. Rather we try to describe the fundamental choices behind the definitions of each language. For this reason we first of all set up a biological framework that is used to fix some important notions which will be then exploited to compare the various calculi (Section 2). In Section 3 we briefly overview various calculi for biology. We specifically deal with the biochemical stochastic π-calculus [23, 20], BioAmbients [21], CCS-R [5], Brane calculi [2], its projective variation [7], the κ-calculus [6], and Beta-binders [18]. Then the biological notions introduced in Section 2 are further used in Section 4 to comment on the calculi for biology which have been presented. Section 5 concludes the survey giving some hints on active line of research in the area.

2 The biological context

In this section we fix a few biological concepts which will be relevant for the following presentation. Later on, describing the main features of various process

calculi we will investigate whether, and possibly how, each formalism addresses the biological issues below.

Most of the phenomena that happen inside the cell are characterized by at least two phases: a *molecular* phase, and a *biochemical* one. First a necessary substance enters the cell, and this is a molecular phenomenon. In the second phase a biochemical reaction takes place. The biology literature typically does not differentiate between the two phases above. The issue is rather dealt with by completely abstracting from one phase or the other (e.g., it is typical to assume that the cell contains all the necessary reactants).

2.1 Molecular phenomena

Cell interactions, like endocytosis or exocytosis, are best described as processes at the molecular level. Here, abstractions like *membranes* or *cell compartments* can be used to provide suitable representations of the phenomena at hand. From the biological point of view, this level offers several open problems.

Molecular phenomena are mostly related to the motion of components. These phenomena can be classified into endocytosis, exocytosis, mitosis and meiosis.

1. *Endocytosis and exocytosis*. Endocytosis consists in absorbing substances from the external environment. Endocytosis can be further distinguished in *pinocytosis* (assumption of liquids: no particle is absorbed except those contained in the liquid), *phagocytosis* (absorption of another component of comparable size), and *generalized endocytosis* (absorption of an arbitrary number of smaller components). Exocytosis is opposite to endocytosis, i.e. it consists in the expulsion of sub-components.

2. *Mitosis and meiosis*. Mitosis, typical of viruses, consists in the exact duplication of the cell. Meiosis, typical of reproductive cells, is the separation of the cell and of the contained genetic material to yield two new "identical" cells. The phenomenon opposite to meiosis is called *merge*.

The division between molecular phenomena and biochemical reactions is a rough simplification. In fact, the activities of the cellular or intracellular membrane are regulated by a lot of *transmembrane proteins*. A transmembrane protein is a protein that passes once or more times through the lipid bilayer of a membrane. Therefore transmembrane proteins cannot freely float in the body, but they could interact with free proteins or with other transmembrane proteins.

Figure 1 exemplifies a process that involves the use of transmembrane protein: phagocytosis. Macrophages are cells belonging to our immune system and they can engulf a virus by phagocytosis. The macrophage uses transmembrane protein

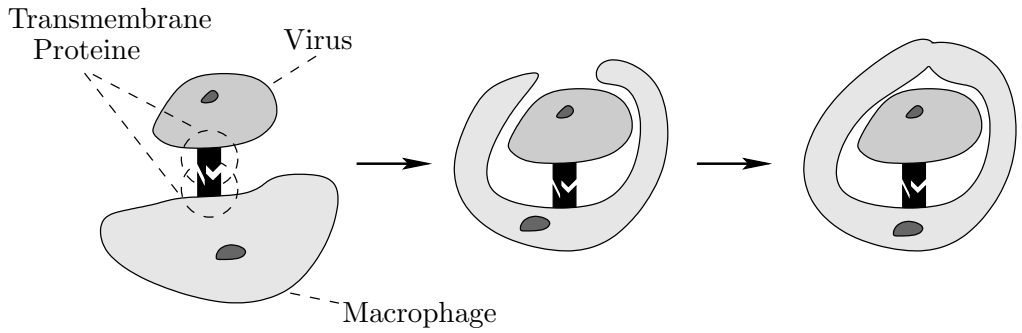

Figure 1: Use of transmembrane proteins

to recognize and block the virus. Only one transmembrane protein is drawn in Figure 1. In the real case hundreds of protein are involved in the event.

2.2 Biochemical reactions

Biochemical reactions focus on proteins. A protein is composed by several distinct structural parts called *domains*. The interaction capability of a protein depends on the structural and chemical complementarity of particular portions of a domain called *motifs* or *sites*. The term "motif" includes different factors, involving three dimensional structure, electrostatics and hydrogen bonds. Free proteins float in the living matter, they keep colliding one against the other and reactions take place whenever there is sufficient quantity of kinetic energy.

Consider, for example, the situation illustrated in Figure 2. A reaction between molecule A and molecule B may occur only if A and B are close enough, and correctly oriented. In Figure 2(a) protein A and B cannot interact due to the fact that their complementary motifs are not correctly oriented (see the two vectors $\vec{V}_A$ and $\vec{V}_B$ which represent orientation). The same two proteins do interact when correctly oriented (Figure 2(b)).

Normally the frequency of reactions is quite low. If needed, enzymes may orientate the molecules in the right way so favoring reactions and speeding them up. Enzymes, most of which are proteins, act as special biological catalysts. They bind temporarily to one or more of the reactants of the reaction they catalyze and, in doing so, enzymes lower the amount of activation energy needed to activate the reaction. Referring to the example of Figure 2(a) an enzyme could change the orientation of protein B leading to the situation of Figure 2(b). Moreover, enzymes can change the structure of proteins to enable reactions. See, for example, the situation illustrated in Figure 3(a). Protein A has two motifs. One of them is *active* and ready for interaction, the other is *hidden* by the protein structure. As it is shown in Figure 3(b), an enzyme can unfold protein A, change its 3D structure,

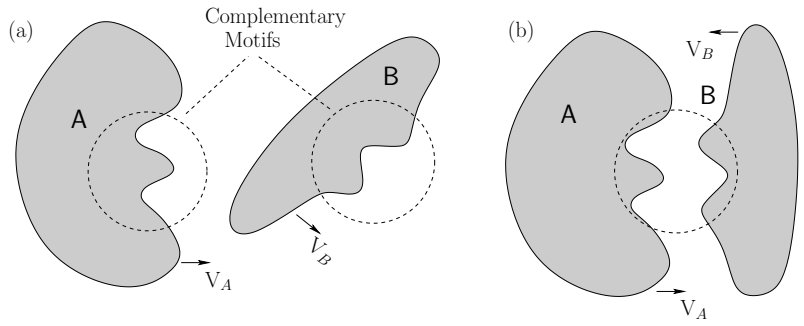

Figure 2: Example of proteins interaction

activate the hidden motif, and bind the active site. Notice that this modification could happen at the structural level as well as at the chemical or electric level (consider, e.g, phosphorylation).

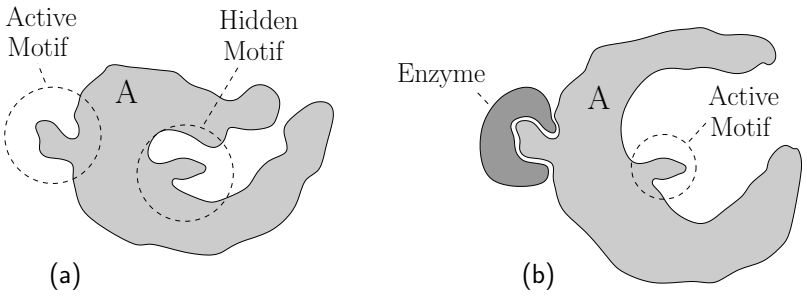

Figure 3: Protein unfolding

2.3 Further biological issues

In addition to biochemical and molecular phenomena, there are some other interesting features that could be represented by a calculus for biology. We present them in what follows, starting from "reversibility" that actually has already been tackled by one of the calculi presented in the next section.

In nature most reactions are *reversible*, meaning that bindings or activations could be undone. Reversibility is mainly governed by the kind of bonds involved and by the available energy. Covalent bonds need a very high quantity of energy to be undone and hence are rarely reverted in nature. Other kinds of bonds (ionic, van der walls) are less demanding w.r.t. energy, and then they can be undone. Reversibility is a basic regulation mechanism that, for example, prevents deadlocks (e.g. [5]).

The ability to reason about *quantitative information* plays a crucial role in biomolecular processes, and hence it also plays a fundamental role in modelling biological systems. For example, in order to correctly describe a reaction it is necessary to know the exact quantity of reactants involved, the affinity of the sites available for bindings, and the amount of energy which can be used. Moreover, interesting biological results come from *in vitro* or *in vivo quantitative* experiments.

Synchronous vs. asynchronous communication in information technology is a reason of disagreement between scientists. The synchronous communication paradigm is centered around the idea that, in order to communicate, both sender and receiver have to be ready and willing to do so. Asynchronous interactions substantially differ from the above model: any sort of agreement between sender and receiver is missing, and the sender just communicates its data which later on may be picked up by a receiver. An analogous dichotomy arises in biology. Depending on the circumstances, the interactions observed in nature follow either a synchronous or an asynchronous paradigm. For instance, DNA replication is made up of a set of synchronous events, that respect a precise order. Vice versa, when proteins are synthesized within the ER (Endoplasmic Reticulum), portions of the ER are pinched off, forming transport vesicles. This seems a spontaneous event (maybe activated by some unknown mechanism), and it would be convenient to abstract it as an asynchronous event.

Equivalence relations, that are well-understood in concurrency theory, could be powerful tools for biology. In information technology two systems are considered equivalent if they show the same behaviour w.r.t. some chosen notion of observation. Different definitions of observation lead to distinct equivalence relations. The desired property is then that two equivalent components can be safely exchanged within a system without altering its overall behaviour. Analogous situations are found in biology. Up to a certain notion of observation, eukaryote cells and prokaryote cells might be seen as belonging to the same class of organisms. In order to relate distinct kinds of cells (e.g. macrophage vs. bacteria) it would be surely necessary a finer grained notion of equivalence.

3 Calculi for biology

In this section we survey the calculi for biology which have been proposed in the literature against the biological requirements presented in Section 2. There, biomolecular processes were described as networks of freely interacting proteins.

The first intuition behind the use of process calculi in modelling biological systems is that reactions can be abstracted by communications. In fact, various description languages were proposed after this first understanding. Chronologically speaking, the first calculi to be used to represent biological phenomena were

adaptations or extensions of languages already used for the specification of distributed systems. Later on, other calculi have been directly proposed for ease of representation of biological phenomena that are more general than biochemical interactions.

3.1 Biochemical stochastic π-calculus

Shapiro et al. proposed to represent molecular systems as mobile communicating processes of the π-calculus [13, 24]. The π-calculus is a name-passing process calculus where names are synonyms of both data and channels. The Biochemical stochastic π-calculus [23, 20], based on [17], extends the π-calculus to allow the quantitative representation of biochemical processes.

In the biochemical stochastic π-calculus, public channel names and co-names represent complementary motifs. Molecules are represented as computational processes, and molecular complexes as parallel compositions of processes sharing a private name. In this way, the scope of the private name represents the boundary of the complex, and a molecule that is external to the complex cannot access it. Movements between complexes and formations of new complexes are represented by name extrusions.

The biochemical stochastic π-calculus is provided with a quantitative simulation system, BioPSI [1], to perform in silico experiments (see, e.g., [11]). The actual rate of a reaction between two proteins is determined according to a basal rate and the concentration of reactants. Two distinct kinds of reactions are distinguished: those between two different proteins, and the so-called homodimerizations, namely the reactions which lead two identical proteins to bind together. The semantics of the biochemical stochastic π-calculus allows the rate of reactions to be traced, and the calculus permits the quantification of the number of molecules involved in a reaction. Recently, independent authors implemented a new simulator for the biochemical stochastic π-calculus [16] in order to model the stochastic simulation of biological processes correctly.

3.2 BioAmbients

BioAmbients [21] is an adaptation of Mobile Ambients [3], a calculus that was introduced to face the problem of representing a private ambient of computation in the study of mobile networks.

BioAmbients represents molecular interactions using processes enclosed within passive ambients. The authors argue that most existing models of biological systems are focused on chemical reaction and give little attention to the hierarchical organization of cellular compartments (e.g. nucleus, ER, Golgi apparatus, ...).

In [21], ambients are the boundaries of a set of π-calculus processes which can communicate with each other. Also, ambients can be nested and are organized in a hierarchy. Depending on the relative locations of the interacting processes, BioAmbients distinguishes three kinds of communications:

- local, namely between two processes in the same compartment, or
- s2s, namely between two processes located in siblings ambients, or
- p2c / c2p, namely between processes localed in ambients with a parent-child (or child-parent) relation.

Moreover, as far as the interpretation of movement is concerned, BioAmbients offers three pairs of primitives:

- enter n / accept n, for entering an ambient, and accepting the entrance, respectively;
- exit n / expel n, for exiting from a containing ambient, and expelling a contained ambient, respectively;
- merge+ n / merge- n, for merging two ambients together.

BioAmbients is stochastically extended following the same intuition as in the biochemical stochastic π-calculus. The definition however is more complex due to the compartment structure representable in the language and, to the best of our knowledge, no applications of the simulator to biological case studies are yet publically available.

Summing up, BioAmbients has a direct representation of molecular membrane, and, just as the biochemical stochastic π-calculus does, it abstracts the concept of transmembrane proteins by using names. Also notice that, like in Mobile Ambients, moving from one ambient to the other requires a synchronization between complementary actions.

3.3 CCS-R

CCS-R [5] is a CCS-like process algebra [12] that focuses on reversibility. The authors observe that the biological examples taken by the other computer science researchers always run forward, while biochemistry researchers state that practically all reactions are reversible.

Based on the above motivations, reversibility is explicitly embedded in the syntax of CCS-R. The calculus is especially equipped with memories to trace communications and backtrack them following a by-need discipline. In order

to implement backtrack, each thread is assigned both a unique identifier and an individual memory stack where information for roll-back is stored. Two constraints have to be met to make roll-back work properly: reversing computations should not give access to unreachable states, and the storing scheme should not induce fake causal dependencies. Reversibility is implemented independently of the amount of energy needed to break bonds.

CCS-R is particularly suitable to represent biochemical interactions. Since it does not use a name-passing discipline, however, it is impossible to directly render the information flow from one process to another. Moreover, compartments cannot be represented via the scoping of names as it is done in the biochemical stochastic π-calculus.

3.4 κ-calculus

The κ-calculus [6] is a formal calculus of protein interactions.

It was particularly conceived to study the problem of complexation and decomplexation of proteins. In fact, the basic units of the κ-calculus are proteins, and its operators are meant to represent the creation and the division of protein complexes. Moreover, the calculus comes equipped with a very clear visual notation, and uses the concept of shared names to represents bonds. Proteins are drawn as boxes with sites on their boundaries. A site can be either visible or hidden or bound.

By a design policy of the authors, the κ-calculus is just a specialized language for proteins, and it does not offer a natural support for the molecular phenomena.

3.5 Brane calculi

Cardelli's Brane calculi presented in [2] are centered around membranes. The author notices that membranes are not just containers, but also active players that take care of coordinating specific activities. Membranes can be highly dynamic: for example, they can shift or merge. Molecules can communicate using their transmembrane proteins that are embedded into membranes and can be thought of as channels.

The primitives of Brane calculi are inspired by membrane properties. A system is viewed as a set of nested membranes, and a membrane as a set of actions depending on transmembrane proteins. The primitives related to movement to and from membranes are classified in two main groups, one for cytosis-like and the other for mitosis-like phenomena.

- General endocytosis is considered an uncontrollable process. More controllable interactions are considered: phagocytosis for engulfing just one

external membrane, and pinocytosis for engulfing zero external membrane. Their inverse, exocytosis, represents the expulsion of an internal membrane.

- Mitosis is considered an uncontrollable process, too. It can split a membrane at an arbitrary place. Alternatively to mitosis, two primitives are offered: budding and dripping. Budding renders the splitting off of one internal membrane, and dripping the splitting off of zero internal membrane. Eventually, a controlled merging of membranes can be done using the so-called mating.

Just as in BioAmbients, explicit coordination between interacting components is always required. In Brane calculi, however, communication is more structured than in both the π-calculus and BioAmbients. Communications are classified into communications *on-membrane* and *cross-membrane*, respectively, and are associated with distinct pairs of primitives.

- *On-membrane.* The primitives $\mathsf{p2p}_n/\mathsf{p2p}_n^{\perp}$ are for on-membrane communications only. These kinds of interaction follow a CCS communication style.

- *Cross-membrane.* The primitives $\mathsf{s2s}_n/\mathsf{s2s}_n^{\perp}$, $\mathsf{p2c}_n/\mathsf{p2c}_n^{\perp}$, and $\mathsf{c2p}_n/\mathsf{c2p}_n^{\perp}$ are for communications between processes in distinct membranes and follow a BioAmbients-like style.

3.6 Projective Brane calculus

Projective Brane calculus [7] is a refinement of Brane calculi. The purpose of this extension is to replace Brane calculi actions with directed actions. Directed actions tell whether an action is looking inwards or outwards the membrane. This abstracts the fact that transmembrane proteins could look inside and outside the membrane. Introducing directed actions takes the language a step further down in the details of biological phenomena.

3.7 Beta-binders

In Beta-binders [18] processes are encapsulated into boxes with interaction sites. These sites are represented by a special class of binders called *beta binders*. Processes inside boxes are standard π-calculus processes extended with a few primitives to manipulate the interaction sites over the surface of boxes and are called pi-processes.

Each beta binder is given a name, that can occur in the actions of the internal process, and a type, e.g. Γ, Δ, ..., that represents the interaction capability of the corresponding site. Γ and Δ are sets of names. Boxes can interact with each

other if they exhibit sites with non-disjoint type. Differently from BioAmbients and Brane calculi, in Beta-binders nesting of boxes is disallowed. The claim is made that, if needed, nesting could be emulated by properly defining the types of sites.

Further, processes can evolve using a limited number of primitives: add a binder to a box, hide a binder, or unhide a binder. These primitives represent the dynamical evolution of the interaction capabilities of the living matter. Also, join and split operations describe the evolution of the structure of boxes by relying on global functions f_{join} and f_{split} whose definition can be parametrically adapted to the biological phenomenon at hand.

4 Calculi in the biological context

The languages viewed so far are quite different, and have been conceived for specifying entities at different levels of abstractions. In what follows, we test them against the biological framework introduced in Section 2, and further comment on their features.

The biochemical stochastic π-calculus, BioAmbients, and CCS-R are adaptations or extensions of calculi introduced to specify distributed systems. All of them represent motifs by names, and interactions between complementary motifs by communications over the corresponding channels.

The biochemical stochastic π-calculus is based on the π-calculus, which has been shown to encode a variety of paradigms and has a very well-assessed mathematical theory, including a wide range of studies on equivalence and congruence relations. The π-calculus semantics, however, may seem quite complex to a neophyte (e.g., to a biologist): using private names for representing a compartment hierarchy could be an uneasy read. Think for example of the macrophage-virus interaction in Figure 1. The macrophage would be represented as a process with a certain private name, say x, and both the virus and the macrophage processes would share yet another name, say z. Then phagocytosis would be rendered as the communication of x over the channel z. After this scope extrusion, the virus and the macrophage share the private name x that now represents their common outermost border. It would probably not be easy to convince a biologist, that this is the interpretation key. On the other hand, it has to be said that the biologist should not necessarily be involved in the formal specification of the system. He/she could just be requested the informal explanation of the interaction, then a computer scientist could carry on its modelling, and eventually the biologist would be given back the results of the simulation tools.

BioAmbients, with its representation of compartments, improves on the readability of process descriptions. For instance, phagocytosis seems more natu-

rally represented by the enter/accept primitives rather than by scope extrusion of names. In BioAmbients, however, a primitive for a biological phenomenon as typical as the splitting off of environments is missing. As a final remark on the calculus, we notice that the explicit use of compartments does not necessarily correspond to an advantage over languages without compartments. In fact, ambients also imply a more rigid structure and this, in turn, could diminish the flexibility of the language. This can be the case, for example, for those specifications which, when given in a π-calculus style, require a heavy use of scope extrusions of private names (see, e.g., the rolling on the endothelium of vessels of different diameters [11]). The BioAmbients representations of these biological phenomena, if possible, would probably be less readable than the π-calculus ones.

Both the biochemical stochastic π-calculus and BioAmbients are given a quantitative simulation system. Nonetheless, while the biochemical bases of stochastic simulations are well founded [9, 10], it is not yet clear how stochastic extensions should be interpreted when dealing with molecular phenomena. The authors of BioAmbients do not clarify, for instance, how Montecarlo processes could abstract cell interactions that, differently from biochemical interactions, require a real time extension. Finally we observe that different static analysis techniques, that could be the real innovation of the use of process calculi in systems biology, have recently been developed in, e.g., [15, 8].

CCS-R has been primarily developed to show how reversibility can be implemented in a process calculus for biology. For this reason CCS-R cannot be directly compared to either the biochemical stochastic π-calculus or BioAmbients. For example, due to the fact that CCS-R does not use a name-passing communication paradigm, it does not allow the description of compartments. Moreover, the authors of CCS-R interpret reversibility as the ability to backtrack from a reaction, then they conclude from the biological literature that it is a common phenomenon in nature. This is true, however, only to certain extents. Indeed the third principle of thermodynamics states that when reversing a reaction one obtains a system that is not exactly the same as the initial one. The interpretation of this principle could become crucial if reversibility would be investigated together with a quantitative analysis of the global energy of the biological system.

The κ-calculus, together with Brane calculi, its projective extension, and Beta-binders, builds on an effort opposite to the one of the above calculi. In each case the authors try to isolate the relevant biological primitives for their representation purposes, and model these primitives by the techniques and tools of concurrency theory. In particular, the κ-calculus was designed to represent complexation and decomplexation of proteins. These phenomena could be represented just as well in the π-calculus. It is, however, a good point of the κ-calculus that the original goal of the authors has been achieved in an effective and nice way. The visual language is quite intuitive, and the formal one rather simple to use. Eventually,

it has to be underlined that the κ-calculus aims at describing only a specific class of phenomena. This results in a well defined and highly comprehensible (for both biologists and computer scientists) language. In this respect the κ-calculus completely differs from all the other languages mentioned in this paper. Those calculi in fact integrate in a unique model biological phenomena which lie at completely distinct levels of abstraction. We claim that the κ-calculus could be a first building block of a successful bottom-up methodology: trying to construct correct models for each level of abstraction, and then integrating them in a greater picture.

Brane calculi is mainly concerned about the description of processes relative to transmembrane proteins. For example, Brane calculi has a specialized primitive to represent membrane division, and another to render phagocytosis. Also, membrane interactions preserve the nesting parity, a principle observed in biological systems and referred to as *bitonality*. The fully synchronous interaction paradigm used in Brane calculi might be an issue for the language. Maybe all the biological activities of membranes are synchronous events that respect a precise order. As a matter of fact however, so far most of these mechanisms are unknown (think, e.g, of the formation of transport vesicles in the endoplasmic reticulum). Then in some circumstances the synchronization mechanism underpinning Brane calculi could be overly specified to allow an easy abstraction from unknown details. Eventually, like in the case of the biochemical stochastic π-calculus, and at least to non experts, the syntax of Brane calculi can make the intended behaviour of systems difficult to understand.

Projective Brane calculus is a variation of Brane calculi based on the observation that transmembrane proteins are oriented. Taking orientation into the calculus leads to simpler definitions, at the level of both syntax and semantics. Phagocytosis and budding elegantly collapse into the same primitive, with a simple difference in the orientation of the transmembrane proteins involved. Like Brane calculi, the projective language adopts a fully synchronous interaction paradigm. Also, despite of the simplifications induced by taking into account orientation of molecules, probably the intended behaviour of processes cannot be easily understood by life science researchers.

The formalism called Beta-binders slightly departs from those mentioned above in the interpretation of the interaction mechanism underlying both the binding of sites and the management of the compartment structure. Beta-binders builds on the following intuitions:

1. a user-driven coordination of the actions leading to the transformation of compartments seems to be too rigid for either abstracting from unknown details or predicting new behaviours;

2. using just names for representing motifs can be too detailed to ensure complete compositionally of specifications. A relevant target of the formal-

ization of biological behaviours could be filling in distributed databases of specifications. This would enable *in silico* experiments: biologists could take here and there the needed components, put all of them in parallel, and then launch the simulation. Using names for motifs would imply that the researchers writing the specifications should agree in advance on the names (and actions vs co-actions) used for certain sites. This is unfeasible in a really distributed research environment.

Given the above observations, Beta-binders is not fully synchronous. For example, the f_{split} functions allow the definition of spontaneous divisions of biological entities. Also, motifs are described not only by names but also by types which can be thought of as representing the shape of sites. Communication between two entities is allowed if they have sites with non-disjoint types, meaning that the shape of the two motifs are complementary to each other. This gives more flexibility to the communication paradigm. Following this same principle of flexibility, molecular phenomena are not directly represented but rather left for definition via the global functions f_{join} and f_{split}. So, for instance, phagocytosis becomes an instantiation of the f_{join} function [19]. A first criticism to the approach comes from the latest statement: it is not easy for a biologist to define a function for phagocytosis. It would be extremely more intuitive to use pairs of enter/accept actions. Moreover, the authors of Beta-binders, although claiming that it could be simulated via types, have chosen to disallow nesting of compartments. This is in contrast with the actual "trend" and could sound odd to a biologist. From the computational point of view, however, this choice could result in a simplification for possible implementations.

5 Future directions

This section concludes the survey by commenting on some active research directions in the area: stochastic extension of calculi, investigation of equivalence relations, and analysis of the represented biological systems.

Up to now stochastic extensions have been developed only for the biochemical stochastic π-calculus and BioAmbients, where the properties of the molecular interactions are still to be studied. For the other languages, stochastic extensions are just planned. The availability of those extensions could help in the cooperation with biologists, whose experiments are based on quantitative valuations. For biologists, the definition of suitable extensions of the various languages and the development of corresponding simulators could result in:

- early feedback, due to the possibility to test the semantics on real in vivo or in vitro experiments;

- better confidence, due to the familiarity with the experiments performed.

Stochastic extensions with general distribution and real time extensions should also be considered.

There is yet no study on specific equivalences in systems biology. In classical process calculi theory, two processes are equivalent if they show the same behaviour w.r.t. some chosen notion of observation. Also, equivalences can typically be characterized by specialized logics and well-assessed techniques can be used to check whether processes are equivalent or not. New notions of observation have to be studied to apply this same approach sensibly to calculi for biology.

All the calculi presented here describe the behaviour of a system by relying on the representation of its transition system. The size of the transition system may be exponential in the size of the program, making the exploration, and thus the analysis, unfeasible. A valid alternative is *static analysis* [14]. The guiding principle is to infer suitable properties only from the text of the program. The use of these techniques in calculi for biology could be a relevant added value. It could be possible, for example, to discover some low probability events that could not be detected by applying simulations. Some of the techniques presented for the π-calculus and the Mobile Ambients could be adapted to the biological extensions of the corresponding calculi (see, e.g., Control Flow Analysis [15]). Moreover, following biological requirements, other techniques could be introduced.

The languages presented in this survey could be considered as a machine model for modelling biological phenomena. They have in fact a narrow set of instructions. In some sense this is not in line with the proposal of hiding from users as many technical details as possible. A characteristic solution of information technology is to design high-level modeling languages that are later translated in a machine language. A first proposal in this direction is PML [4], a high level language for modelling pathways. PML uses the π-calculus as underlying machine model and thus it makes it possible to use all the analysis techniques and the simulation systems developed for π-calculus.

References

[1] Biopsi home page: `http://www.wisdom.weizmann.ac.il/~biopsi/`.

[2] L. Cardelli. Membrane interactions. In *BioConcur '03, Workshop on Concurrent Models in Molecular Biology*, 2003.

[3] L. Cardelli and A.D. Gordon. Mobile ambients. *Theoretical Computer Science*, 240(1):177–213, 2000.

[4] B.E. Chang and M. Sridharan. PML: Toward a high-level formal language for biological systems. In *BioConcur '03, Workshop on Concurrent Models in Molecular Biology*, 2003.

[5] V. Danos and J. Krivine. Formal molecular biology done in CCS-R. In *BioConcur '03, Workshop on Concurrent Models in Molecular Biology*, 2003.

[6] V. Danos and C. Laneve. Graphs for core molecular biology. In C. Priami, editor, *Proc. 1st Int. Workshop on Computational Methods in Systems Biology (CMSB 2003)*, volume 2602 of *Lecture Notes in Computer Science*. Springer, 2003.

[7] V. Danos and S. Pradalier. Projective Brane Calculus. In V. Danos and V. Vincent Schächter, editors, *Proc. 2nd Int. Workshop on Computational Methods in Systems Biology, CMSB '04*, Lecture Notes in BioInformatics. Springer, 2005. To appear.

[8] D. Schuch da Rosa F. Nielson, H. R. Nielson and C. Priami. Static analysis for systems biology. In *Proc. of workshop on Systeomatics - dynamic biological systems informatics*, number to appear. Computer Science Press, Trinity College Dublin, 2005.

[9] D.T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical species. *Journal of Computational Physic*, 22:403–434, 1976.

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

[11] P. Lecca, C. Priami, P. Quaglia, B. Rossi, C. Laudanna, and G. Costantin. A Stochastic Process Algebra Approach to Simulation of Autoreactive Lymphocyte Recruitment. *SIMULATION: Transactions of The Society for Modeling and Simulation International*, 80(4), 2004.

[12] R. Milner. *Communication and Concurrency*. International Series in Computer Science. Prentice hall, 1989.

[13] R. Milner. *Communicating and mobile systems: the π-calculus*. Cambridge Universtity Press, 1999.

[14] F. Nielson, H. R. Nielson, and C. L. Hankin. *Principles of Program Analysis*. Springer, 1999.

[15] F. Nielson, H. R. Nielson, C. Priami, and D. Schuch da Rosa. Control flow analysis for bioambients. In *BioConcur '03, Workshop on Concurrent Models in Molecular Biology*, 2003.

[16] A. Phillips and L. Cardelli. A correct abstract machine for the stochastic pi-calculus. In *BioConcur '04, Workshop on Concurrent Models in Molecular Biology*, 2004.

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

[18] C. Priami and P. Quaglia. Beta binders for biological interactions. In V. Danos and V. Vincent Schächter, editors, *Proc. 2nd Int. Workshop on Computational Methods in Systems Biology, CMSB '04*, Lecture Notes in BioInformatics. Springer, 2005. To appear.

[19] C. Priami and P. Quaglia. Operational patterns in Beta-binders. *Transactions on Computational Systems Biology*, 2005. To appear.

[20] C. Priami, A. Regev, W. Silverman, and E. Shapiro. Application of a stochastic name-passing calculus to representation and simulation of molecular processes. *Information Processing Letters*, 80(1):25–31, 2001.

[21] A. Regev, E.M. Panina, W. Silverman, L. Cardelli, and E. Shapiro. Bioambients: An abstraction for biological compartments. *Theoretical Computer Science*, 2005. To appear.

[22] A. Regev and E. Shapiro. Cells as computations. *Nature*, 419:343, 2002.

[23] A. Regev, W. Silverman, and E. Shapiro. Representation and simulation of biochemical processes using the pi-calculus process algebra. In *Proceedings of the Pacific Symposium of Biocomputing 2001*, volume 6, pages 459–470, 2001.

[24] D. Sangiorgi and D. Walker. *The π-calculus: a Theory of Mobile Processes*. Cambridge Universtity Press, 2001.