

# A compartmental model of the cAMP/PKA/MAPK pathway in Bio-PEPA

Federica Ciocchetta

The Microsoft Research - University of Trento Centre  
for Computational and Systems Biology, Trento, Italy  
ciocchetta@cosbi.eu

Adam Duguid

University of Edinburgh, UK  
a.j.duguid@sms.ed.ac.uk

Maria Luisa Guerriero

University of Edinburgh, UK  
mguerrie@inf.ed.ac.uk

The vast majority of biochemical systems involve the exchange of information between different compartments, either in the form of transportation or via the intervention of membrane proteins which are able to transmit stimuli between bordering compartments.

The correct quantitative handling of compartments is, therefore, extremely important when modelling real biochemical systems. The Bio-PEPA process algebra is equipped with the capability of explicitly defining quantitative information such as compartment volumes and membrane surface areas. Furthermore, the recent development of the Bio-PEPA Eclipse Plug-in allows us to perform a correct stochastic simulation of multi-compartmental models.

Here we present a Bio-PEPA compartmental model of the cAMP/PKA/MAPK pathway. We analyse the system using the Bio-PEPA Eclipse Plug-in and we show the correctness of our model by comparison with an existing ODE model. Furthermore, we perform computational experiments in order to investigate certain properties of the pathway. Specifically, we focus on the system response to the inhibition and strengthening of feedback loops and to the variation in the activity of key pathway reactions and we observe how these modifications affect the behaviour of the pathway. These experiments are useful to understand the control and regulatory mechanisms of the system.

## 1 Introduction

Compartments are widely present in biological systems and play a major role in a lot of biological processes [1]; they are enclosed by membranes, which isolate biological species from the external environment. Biological species can be either located inside a compartment or reside on a membrane; in the latter case they can be attached to a side of the membrane (*receptors*, for instance) or span the entire membrane (i.e. *transmembrane proteins*). Species located in the same compartment can interact with each other through biochemical reactions; species can also move from one compartment to another via diffusion or through membrane channels, thus allowing the passing of information between adjacent compartments. Alternatively, the transmission of signals between compartments can involve receptors, which respond to the input of signalling molecules on one side of the membrane by triggering a cascade of events on the other side.

Due to the fundamental importance of compartments, their correct handling is often essential when studying biological systems, and an increasing number of modelling languages support their explicit definition [30, 6, 34, 32, 7, 41]. These languages differ from each other in the kind of abstraction, the assumptions made (for instance, static or dynamic compartments) and the features they are designed to represent. Most of them are supported by analysis tools [2, 19, 3, 14].

In this work we consider Bio-PEPA with locations as our modelling language [12]. Bio-PEPA locations abstract both biological compartments and membranes: they are defined by *names*, enriched with additional information allowing the modeller to express their type (i.e. compartment or membrane), their size, and their position with respect to the other locations of the system.

Bio-PEPA offers a high level of abstraction, similar to that used to describe biochemical networks in the literature and in biological databases (see, for instance, models in the *BioModels* database [28]). Only *static locations* can be represented (i.e. compartments cannot merge, split, or undergo any structural change); nevertheless, their size can vary with respect to time.

The main aim of this work is to demonstrate the power of Bio-PEPA with locations as a modelling language for multi-compartmental biochemical networks. Thus, we consider a moderately sized signalling pathway describing the cAMP/PKA/MAPK activity [29] with the explicit definition of compartments and membranes and we translate it into Bio-PEPA. In this pathway the spatial aspects are particularly relevant, since it involves interaction between molecular species residing in different biological compartments with different volume sizes.

The analysis of the model is performed using the Bio-PEPA Eclipse Plug-in [11], a novel simulation tool for the construction and analysis of Bio-PEPA models. The Bio-PEPA Eclipse Plug-in offers methods for numerical integration of Ordinary Differential Equations (ODEs) and stochastic simulation. We present here the latest version of the tool, which supports the definition of multi-compartmental models defined using the Bio-PEPA extension with locations and is able to handle multiple compartments with different sizes in the appropriate way.

Using the Bio-PEPA Eclipse Plug-in, we validate the model against the expected results in the literature and we perform in-silico experiments in order to investigate some properties of the pathway. Specifically, we study how the behaviour of the pathway is affected by varying the enzymatic activity of three key molecular species and by the removal of feedback loops with an important regulatory effect over particular species. These experiments are useful for a better understanding of the regulatory and control mechanisms of the system.

The structure of the paper is as follows. Section 2 reports a brief introduction of the Bio-PEPA language, while Sect. 3 is devoted to the description of the tool. In Sect. 4 we present a model of the cAMP/PKA/MAPK signalling pathway, and in Sect. 5 we show the validation of the model and some analysis results. Section 6 concerns some related work, whereas the last section reports some conclusive remarks.

## 2 Bio-PEPA with locations

In this section we describe the main features of the extension of Bio-PEPA [13] with locations [12], the modelling language used in this work.

A Bio-PEPA system representing a biochemical network consists of a *context*, defining information such as locations, kinetic rates, parameters and auxiliary information, a set of *species components* and a *model component*.

Locations represent biological compartments and membranes. They are described by *names*, enriched with additional information allowing the modeller to express their position with respect to the other locations of the system, their type (i.e. compartment or membrane), and their size (i.e. volume or surface area). The structure of the biological system is modelled as a *static hierarchy*, represented as a tree whose nodes represent locations; each node has one child for each of its sub-locations. The *location tree* allows us to keep track of the relative positions of locations and must be associated with the location definition.

The locations are defined as follows.

**Definition 1** *Each location is described by “ $L$ : size unit, type”, where  $L$  is the (unique) location name, “size” expresses the size and can be either a positive real number or a more complex mathematical*

expression depending on time  $t$ ; the (optional) “unit” denotes the unit of measure associated with the location size, and “type”  $\in \{\mathbf{M}, \mathbf{C}\}$  expresses if it is a membrane or a compartment, respectively. The set of locations is denoted by  $\mathcal{L}$ .

Functional rates are used to describe kinetic laws associated with a reaction. They have the form “ $f.\alpha = expr$ ”, where  $\alpha$  is the action type abstracting the reaction and  $expr$  is a mathematical expression describing the law. The kinetic parameters used in the mathematical expressions are defined by parameters in Bio-PEPA.

Species components describe the behaviour of the individual biological species, whereas the model component defines the interactions between the various species and the amount of them present in the system. The syntax of the Bio-PEPA components is defined as:

$$S ::= (\alpha, \kappa) \text{ op } S \mid S + S \mid C \mid S @ \text{location} \quad \text{with op} = \downarrow \mid \uparrow \mid \oplus \mid \ominus \mid \odot \quad P ::= S(x) \mid P \boxtimes_{\mathcal{I}} P$$

where  $S$  is the *species component* and  $P$  is the *model component*. In the prefix term  $(\alpha, \kappa) \text{ op } S$ ,  $\kappa$  is the *stoichiometry coefficient* of species  $S$  in reaction  $\alpha$ , and the *prefix combinator* “op” represents the role of  $S$  in the reaction. Specifically,  $\downarrow$  indicates a *reactant*,  $\uparrow$  a *product*,  $\oplus$  an *activator*,  $\ominus$  an *inhibitor*, and  $\odot$  a *generic modifier*. We can use “ $\alpha \text{ op } S$ ” as an abbreviation for “ $(\alpha, 1) \text{ op } S$ ”. The operator “+” expresses the choice between possible actions, and the constant  $C$  is defined by an equation  $C \stackrel{\text{def}}{=} S$ . The term  $S @ \text{location}$  indicates that the species represented by the process  $S$  is the location *location*. The parameter  $x \in \mathbb{R}^+$  in  $S(x)$  represents the number of molecules, which can be abstracted to a level [10] for analysis as a CTMC. Finally, the process  $P \boxtimes_{\mathcal{I}} Q$  denotes the cooperation between components: the set  $\mathcal{I}$  determines those activities on which the operands are forced to synchronise. The notation  $\boxtimes_*$  means that the species components are obliged to synchronize on the common action types.

Note that the same species in different locations is represented by distinct species components. Therefore, in order to avoid the possible duplication of actions in the model, we propose a notation to represent a species in multiple locations in a compact way. Given a species  $S$  which can be in  $L_1, \dots, L_n$ , one single component  $S$  is defined. This is simply a shorthand for a set of definitions  $S @ L_1, \dots, S @ L_n$ , each of which contains only those actions which can occur in the respective location. A reaction  $\alpha_1$  occurring only in  $L_i$  is defined as  $(\alpha_1, \kappa) \text{ op } S @ L_i$ , while  $\alpha_2$  occurring in all locations  $L_1, \dots, L_n$  is defined as  $(\alpha_2, \kappa) \text{ op } S$ .

No constraint is imposed on the location of interacting species. This implies that it is possible to describe reactions involving species located in non-adjacent locations. The reason behind this choice is that, even though interactions are generally local (i.e. involve molecules in the same or adjacent locations), imposing constraints on the relative position of species locations could prevent the modellers from defining some reactions (e.g. a reaction abstracting a sequence of reaction steps involving various locations).

The formal definition of the Bio-PEPA system is the following.

**Definition 2** A Bio-PEPA system  $\mathcal{P}$  is a 6-tuple  $\langle \mathcal{L}, \mathcal{N}, \mathcal{K}, \mathcal{F}_R, \text{Comp}, P \rangle$ , where:  $\mathcal{L}$  is the set of locations,  $\mathcal{N}$  is the set of (auxiliary) information for the species,  $\mathcal{K}$  is the set of parameters,  $\mathcal{F}_R$  is the set of functional rates,  $\text{Comp}$  is the set of species components and  $P$  is the model component.

A main feature of Bio-PEPA is that it can be seen as a formal, intermediate representation of biochemical systems, on which different kinds of analysis can be carried out, through the defined mappings into continuous-deterministic and discrete-stochastic modelling languages and CTMC with levels.

Bio-PEPA is given an operational semantics, which is used for the derivation of the labelled stochastic transition system and the associated CTMC with levels. The operational semantics is not considered for

the kinds of analysis on which we focus in this work (continuous-deterministic and discrete-stochastic) and, hence, it is not shown here. For details, see [12, 13].

Finally, it is important to remember that Bio-PEPA (and, in general, process algebras) offers a *compositional approach* for the definition of models: it is possible to build the full model starting from the definition of its subcomponents. This approach is particularly useful for biological systems, as they are generally large and complex. For these systems, a compositional definition of the model is desirable since it makes the construction of the full model easier and it helps in understanding the relationships between its various subcomponents and the structure of the network. The compositional approach has been used in the definition of the Bio-PEPA model for the cAMP/PKA/MAPK pathway. The compositional rules for Bio-PEPA models are standard and not reported here.

### 3 The Bio-PEPA Eclipse Plug-in

The Bio-PEPA Eclipse Plug-in [4] is one of two tools developed at Edinburgh University for the Bio-PEPA language. Built on top of the Eclipse platform [18], the plug-in offers an operating system agnostic editor and suite of time-series solvers for the construction and analysis of Bio-PEPA models. It provides the modeller with a rich modelling environment for Bio-PEPA, from the initial creation of the model through to the analysis and inspection of the results.

The features currently supported in the plug-in can be roughly grouped into those assisting the modeller in the writing of their models and features for the analysis of the model. Examples of the former include an editor (with Bio-PEPA-aware syntax highlighting) and static analysis of the model for detecting various syntactic errors. For the latter, the plug-in is capable of generating and displaying time-series results, from ODE numerical integration or stochastic simulation algorithms, with the option to export these results in both graphical and textual formats. Many of these features can be seen in Figure 1.

As with all previous releases of the plug-in, the language is augmented with keywords for particular definitions. These keywords increase the readability of the model, clearly signposting those definitions pertaining to locations, species and kinetic laws. These indicators are then emphasised further by the syntax highlighting built into the developing environment. Together they assist the modeller in recognising the purpose of various definitions found within any given model. As of version 0.1.0, the Eclipse Plug-in will contain initial support for multi-compartmental models as defined by the Bio-PEPA extension with locations. The following description details both this support along with several refinements to the implemented language. For the sake of brevity the description does not cover the entire language, instead concentrating on those required to model the system of interest.

Units of measures are currently unsupported, all values must be defined with respect to the same generic unit, with compartments referring to volumes ( $\text{unit}^3$ ) and membranes as surface area ( $\text{unit}^2$ ). Finally, the initial values for species are assumed to be in terms of molecular counts. With the introduction of locations with different sizes, concentrations become an inappropriate form of measurement and must be translated into molecular counts.

#### 3.1 Supported Language

Bio-PEPA does not impose a strict ordering on the definitions of context and species components, but these must all precede the model component. This final statement provides the structure of the system as well as the initial conditions. Since the modeller is relatively unconstrained with respect to the ordering we shall present the syntax for different definitions in the order that they are parsed within the plug-in.

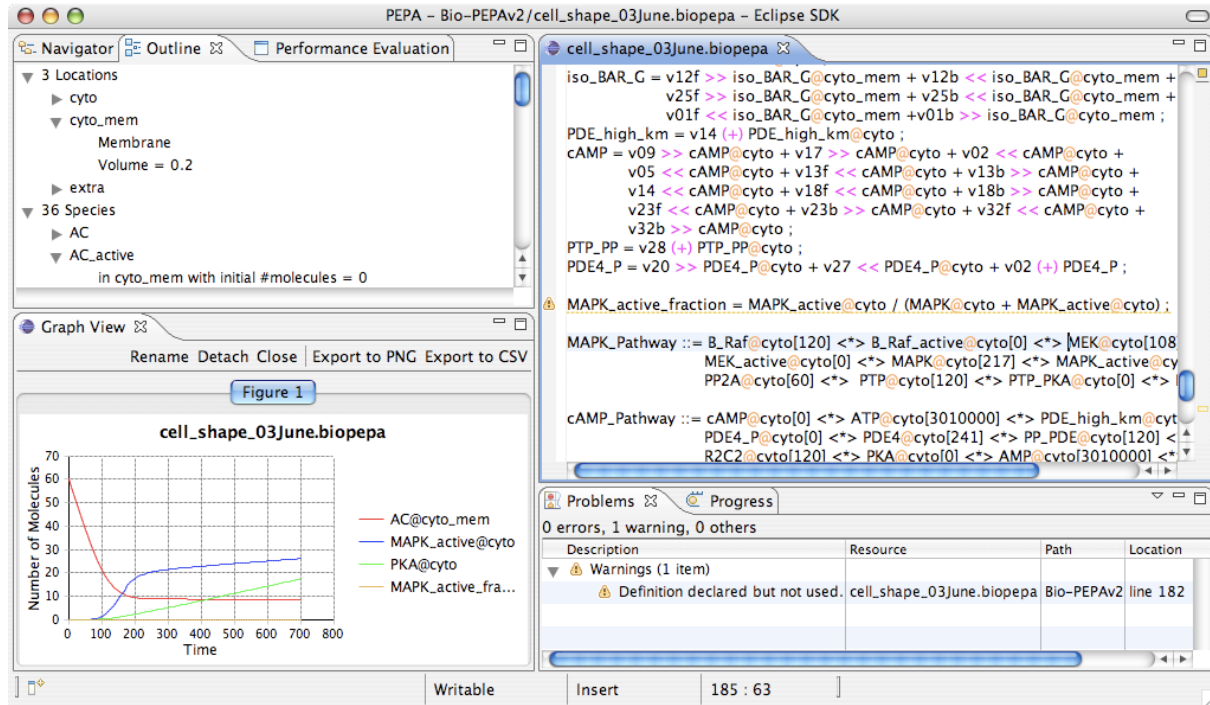


Figure 1: Screen shot of the Bio-PEPA Eclipse Plug-in. The editor can be seen in the top left corner, with the outline view in the top right. The problems tab (bottom right) displays warnings and errors associated with the open models and the ability to plot results can be seen in the bottom left.

The first is the location definition, an example of which can be seen below:

```
location cyto in cyto_mem : size = 1, type = compartment;
location cyto_mem : size = 0.2, type = membrane;
```

The required keyword `location` identifies this as a location statement and the optional keyword `in` is used for defining the position of the given location in the location tree. In the example above we have defined a compartment `cyto` to exist within a second location known as `cyto_mem`. Allowable parameters for these locations include the size and type, both of which can be seen. For reasons of legibility acceptable types are `compartment` and `membrane` rather than `M` and `C`. Based on the types in the example and the spatial hierarchy we can clearly see that `cyto_mem` is the membrane to the `cyto` compartment.

The next required definition in a Bio-PEPA model describes the kinetic rates of the reactions, with an example below:

```
kineticLawOf v1f : kf_activate_Gs*iso_BAR_G@cyto_mem;
```

where `kf_activate_Gs` is the constant rate of the reaction `v1f` and `iso_BAR_G@cyto_mem` is the reactant of the reaction (the species amount is given in number of molecules). Like location definitions, this definition is easily identifiable by the required keyword `kineticLawOf`, followed by its label for reference within the species component definitions. The defined rate can either be a custom rate, as seen in the

example, or make use of predefined laws such as mass-action or Michaelis-Menten. For example, if we know that the only reactant in the reaction associated with  $v1f$  is  $iso\_BAR\_G@cyto\_mem$ , then this custom rate describes mass-action. This could then be re-written as:

$$\text{kineticLawOf } v1f \quad : \quad \text{fMA}(kf\_activate\_Gs);$$

where  $\text{fMA}(r)$  stands for mass-action kinetic law with constant parameter  $r$ . The inclusion of predefined laws offers two clear benefits to the modeller. Mistakes in inputting rates can be reduced by removing the need to input the laws manually. Secondly, the use of predefined kinetic laws allows for additional static analysis of the model. If a reaction is defined as using Michaelis-Menten ( $\text{fMM}(v_M, k_M)$ , where  $v_M, k_M$  are the two constants in Michaelis-Menten kinetic law<sup>1</sup>), then the reaction must have a single species acting as a reactant or substrate, another species performing the role of the catalyst or enzyme and a third as the product of the reaction. If the list of species involved in this reaction differs from those expected for the kinetic law stated the tool will flag this as an error for the modeller.

For the species component definitions, acceptable symbols for the different possible behaviours are required. Below is the list of Bio-PEPA operators along with their ASCII equivalent for creating Bio-PEPA models.

Behaviour	Bio-PEPA symbol	ASCII representation
reactant	↓	<<
product	↑	>>
activator	⊕	(+)
inhibitor	⊖	(-)
modifier	⊙	(.)
uni-directional transportation	→	->
bi-directional transportation	↔	<->
cooperation	⊗ <sub>L</sub>	<L>

These operators are used within the species component definitions, an example of which can be seen below.

$$PDE4 = v27 >> PDE4@cyto + v20 << PDE4@cyto + v5 (+) PDE4@cyto;$$

The species PDE4 is involved in three interactions with different roles: it is a product in the reaction  $v27$ , a reactant in  $v20$  and an enzyme in  $v5$ . In the example above all three reactions are only possible when the species PDE4 is in the location  $cyto$ . From this example we can see shorthand notations for when the stoichiometric coefficient is one. As with all defined shorthands, their purpose is to keep the definitions free of superfluous details. This is not to suggest that stoichiometric information in general is unnecessary, but that through the use of this shorthand we can emphasise the less common instances (a stoichiometry of one being the majority of observed values) where this is not the case.

The final definition in a Bio-PEPA system is that of the model component (see Section 2). The model component describes the initial amount of each species in the system and the possible interactions involving these species. The plug-in however offers true compositionality. It does this by first augmenting the model component syntax to accept names and also allowing definitions written in the syntax of the

<sup>1</sup>Given an enzymatic reaction  $S + E \rightarrow P + E$ , where  $S$  is the substrate,  $P$  is the product and  $E$  is the enzyme, the Michaelis-Menten kinetic law is  $v_M \cdot E \cdot S / (K_M + S)$ , where  $v_M$  is the maximum rate constant and  $K_M$  is the Michaelis constant.

model component to be assigned names. These additional labelled compositional definitions can then be referred to in either the model component or other labelled compositional definitions. This allows the modeller to describe a complex system in terms of smaller pathways, or indeed any logical grouping that they wish to enforce.

$$\begin{aligned} \text{MAPK\_Pathway} ::= & B\_Raf@cyto[120] < * > B\_Raf\_active@cyto[0] < * > \dots \\ & < * > MAPK@cyto[217] < * > MAPK\_active@cyto[0] \\ & < * > \dots < * > PTP\_PP@cyto[60]; \end{aligned}$$

Here we have a labelled composition definition for the MAPK pathway. It should be noted that labelled compositional definitions use a different assignment operator to other definitions, using  $::=$  instead of  $=$ . As stated earlier, and as can be seen in the example above, labelled compositions also use the definition terminator. This leaves the model component as the final definition to present, which an example of can be seen below:

$$G\_Pathway < v09, v17 > cAMP\_Pathway < v08, v15 > MAPK\_Pathway$$

The benefits gained through compositional definitions are highlighted in this model component definition. Not only does the model component make it clear that the system is comprised of three pathways, but can also serve to highlight important reactions. We can see that interaction between the cAMP and MAPK pathways is limited to reactions  $v08$  and  $v15$ . If the MAPK pathway is reliant on the cAMP pathway, rather than any interaction being redundancy in the system, then the rates for these reactions are likely candidates for perturbation. This of course is not guaranteed, merely that the cooperation sets can be used to highlight limited interactions between groups of species. The compositional approach also makes it very easy to enable or disable entire pathways.

Once the model has been built, the plug-in parses the model and performs static analysis to detect a variety of modelling errors, an example of which was given earlier in regards kinetic rate laws. Once the model is free of these errors the modeller has access to ODE solvers in the form of a Runge-Kutta implicit-explicit solver and a Dormand-Prince adaptive step-size solver [40], these being made available through the odeToJava library. The plug-in also supports analysis through stochastic simulators, with implementations of Gillespie's Stochastic Simulation Algorithm (SSA) [22], Gibson-Bruck [20] and the Tau-Leap algorithms [21] available. The ability to use different types of solvers has allowed us to discover errors in published modelling studies in computational biology [5].

## 4 The cAMP/PKA/MAPK pathway

The cAMP/PKA/MAPK network is a ubiquitous signalling pathway. In this work we consider the system as modelled in [29] and in particular the ODE model described by the authors and available in SBML format from [37].

The cAMP-dependent protein kinase (PKA) is responsible for performing several functions in the cell, such as regulation of glycogen, sugar, and lipid metabolism. Among the downstream targets of PKA there are mitogen-activated protein kinases (MAPKs), which are essential in regulating various cellular activities, including gene expression, mitosis, differentiation, and cell survival/apoptosis.

The major role of this pathway in various cellular processes, together with the availability of quantitative data such as the sizes of the involved compartments, make it an interesting case study for our purposes.

As is common in signalling pathways, spatial aspects are particularly relevant in this system, since it involves interaction between molecular species residing in different biological compartments. We consider molecules lying inside the cell (i.e. in the cytoplasm), and outside it (i.e. in the extra-cellular environment), specifically the receptor agonist which triggers the signalling cascade. These two compartments are separated by a membrane (i.e. the cytoplasmic membrane) where receptors and specific membrane-proteins are located.

The compartments have different sizes and this exercises a direct impact on the quantitative behaviour of the system. As in the original ODE model [29, 37], we consider the cell to be a sphere with a surface-to-volume ratio equal to 0.2. The kinetic parameters we use are taken from the original paper, which have been derived from the biochemical literature.

Before presenting the Bio-PEPA model in the following Sect. 4.1, we briefly describe the biological system. Figure 2 shows a schematic representation of the pathway, and the full model is reported in the supplementary material.

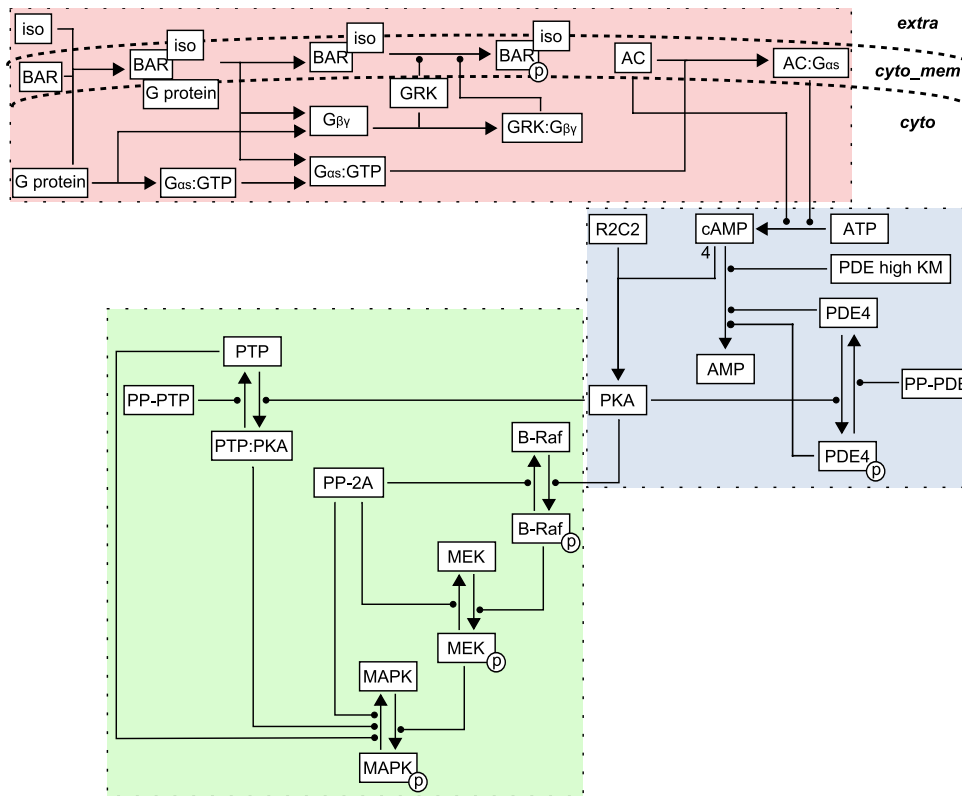


Figure 2: Schematic representation of the pathway. The coloured boxes represent a decomposition of the pathway into three subnetworks: the G-protein mediated activation of AC, the activation of PKA, and the MAPK signalling cascade. Arrows represent reactions, and lines terminating with circles represent enzymes participating in reactions.

The cAMP/PKA/MAPK network we consider is in neurons, and its biological action is initiated by isoproterenol (ISO) binding to specific cell-surface receptors called  $\beta$ -adrenergic receptors (BARs), which are a class of G-protein coupled receptors (GPCRs).

GPCRs reside in the cytoplasmic membrane and respond to a wide range of extracellular stimuli,



in our case to the synthetic drug isoproterenol. When GPCRs bind an extracellular ligand, they interact with G-proteins and propagate the signal across the membrane into the cytoplasm.

G-proteins are heterotrimeric complexes bound to the inside surface of the cell membrane, and consist of an active subunit  $G_\alpha$  and two tightly associated  $G_\beta$  and  $G_\gamma$  subunits.  $G_\alpha$  subunits alternate between an inactive guanosine diphosphate (GDP) and an active guanosine triphosphate (GTP) bound state. The transition from the inactive to the active state of  $G_\alpha$  and its simultaneous dissociation from the  $G_{\gamma\beta}$  subunit and the GPCR are stimulated by the binding of GPCRs with the ligand.

$G_s$  subunits are a subclass of  $G_\alpha$  subunits, and they stimulate the production of cyclic-AMP (cAMP) from ATP, by activating (by binding to it) the membrane-associated enzyme adenylate cyclase (AC) which, in turn, produces cAMP.

cAMP is a key signalling molecule which acts as a second messenger, activating the cAMP-dependent protein kinase A (PKA). PKA is a tetrameric holoenzyme, consisting of two regulatory and two catalytic subunits ( $R_2C_2$ ). Under low levels of cAMP, PKA is inactive; when the concentration of cAMP rises, cAMP molecules bind to each of the two binding sites on the two regulatory subunits, which leads to the dissociation and activation of the catalytic subunits (i.e. four cAMP molecules are needed to activate one PKA molecule). Upon activation, the catalytic subunits can phosphorylate a number of downstream targets.

PKA is an enzyme which participates in several metabolic and signalling pathways, leading to a wide range of biological responses. Here we focus on its ability to initiate the pathway leading the activation of the Mitogen-Activated Protein Kinase (MAPK).

The downstream pathway in our network represents an instance of the common MAPK signalling chain. It starts with the phosphorylation of the MAPKKK B-Raf, which phosphorylates (i.e. activates) the MAPKK MEK, which in turn phosphorylates MAPK.

The amounts of PKA and MAPK in cells are important elements to monitor when analysing the behaviour of the system, because of their essential role in several signalling pathways and other cellular processes.

An interesting aspect of this pathway is the presence of numerous feedback mechanisms, which are carried out mainly by kinases and phosphatases, regulating the amount of the involved proteins.

Two kinds of phosphatases are involved: first, the serine/threonine protein phosphatase 2 (PP-2A), which is a ubiquitous enzyme able to dephosphorylate a wide range of proteins among which there are the components of the MAPK signalling cascade (B-Raf, MEK and MAPK); second, protein tyrosine phosphatases (PTPs), which are also involved in the dephosphorylation of MAPK.

In addition to its positive role carried out by activating B-Raf, PKA is further involved in the regulation of the pathway, by exercising both a positive and a negative regulatory role: respectively, it phosphorylates PTPs (in the phosphorylated form, PTPs are less effective in dephosphorylating MAPK) and PDE4 (in the phosphorylated form, PDE4 enhances the transformation of cAMP into AMP, making it unavailable for activating PKA).

The relative strength of the different PKA enzymatic activities, and particularly the opposed positive and negative feedbacks, is therefore an interesting issue to be considered when studying PKA regulatory role.

#### 4.1 A Bio-PEPA Model of the Network

In the schematic representation of the pathway in Fig. 2 we have subdivided the system into three different subnetworks:

- upstream pathway: G-protein mediated activation of AC

- intermediate pathway: activation of cAMP-dependent PKA
- downstream pathway: MAPK signalling cascade

The Bio-PEPA language allows the definition of compositional models and, therefore, the above subdivision can be straightforwardly employed in the model itself. This helps in improving the readability of the model. The full Bio-PEPA model (using the syntax and the shorthand notation supported by the Bio-PEPA Eclipse plug-in) is reported in the supplementary material.

Three locations are defined in the Bio-PEPA model: *cyto*, *cyto\_mem*, and *extra*. Each of them represents one of the involved biological compartments, and is assigned a size expressing its volume (3D compartments) or its surface area (2D membranes).

Kinetic parameters and rate laws are specified. Some of the reaction rates follow the mass-action law, while others are based on the Michaelis-Menten kinetics. The quantitative data we adopt are those defined in the original ODE model [37]. The reversible reactions are decomposed into two irreversible (mass-action) reactions describing the forward and the backward directions, respectively. The full set of parameters used is reported in the supplementary material.

Each molecular species in the pathway is abstracted by a species component describing the reactions in which it is involved and its role in each of them (i.e. whether it is a reactant, a product, or an enzyme). The Bio-PEPA model component describes how the species components interact with each other, and defines the initial amounts for each species.

There are two noteworthy differences with respect to the original model. These are related to the conversions needed to switch to molecular counts as units of measure for the species amount. In the ODE model the species amounts are specified in terms of concentrations (numbers of molecules over volume/area), while in our Bio-PEPA model they must be specified in terms of numbers of molecules. Consequently, the rate laws must be rescaled accordingly to take into account the sizes of the compartments in which the species involved in the reactions are located.

In this definition of the model, compartment volumes and membrane surface areas are assumed to be constant. This constraint is not present in Bio-PEPA, which allows location sizes to be defined as arbitrary functions (e.g., they can change over time or depend on the amount of given species). Here we do not consider generic location sizes in order to be able to compare our results with the original model. With this assumption, the system could be modelled using languages which are not equipped with a notion of location (provided reaction rates are carefully computed taking into account volumes). However, using a language, such as Bio-PEPA, which is equipped with a notion of location, has also the obvious advantage that the explicit definition of information such as locations of species, volumes and areas brings a greater clarity of the model.

## 5 Model Analysis

### 5.1 Validation

In order to verify that the Bio-PEPA model is a faithful representation of the original ODE model [37], we compared the results we have obtained from the Bio-PEPA model with the results obtained through the BioModels online simulator on the original model.

Figure 3(a) reports the comparison of the results for the ratio of active MAPK (i.e. phosphorylated over total MAPK). We show both the time series obtained using one of the stochastic simulators and the one obtained using one of the ODE solvers available within the Bio-PEPA Eclipse Plug-in. We observe a good agreement between the results from the Bio-PEPA model and the ones from the original ODE

model. In particular, the two ODE-based analyses produce coincident results. The SSA plot, obtained as the average over 500 individual stochastic simulation runs, is in good agreement with the deterministic analysis and also shows the variability of the system due to the presence of some stochastic noise.

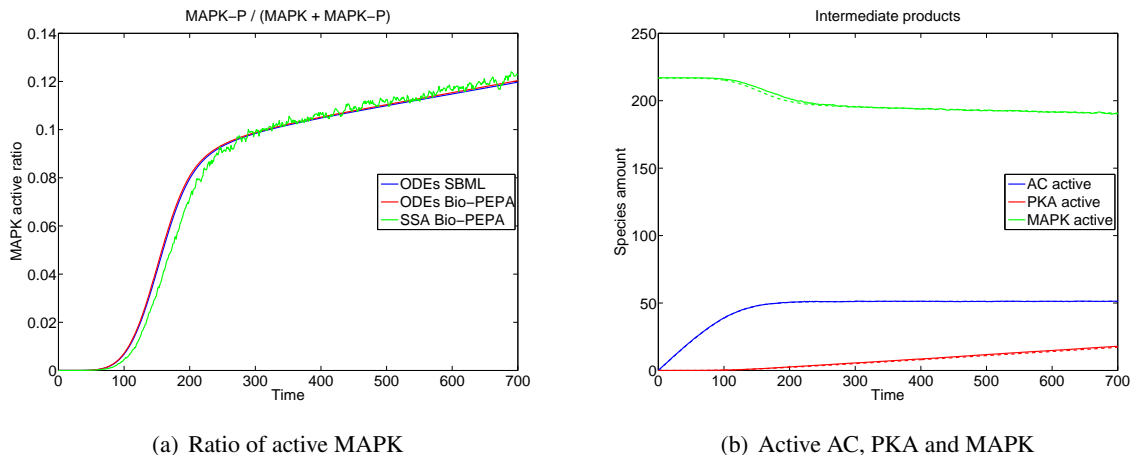


Figure 3: Comparison between the simulation results obtained by the Bio-PEPA Eclipse Plug-in and the results of the original ODE model. SSA results are the average over 500 simulation runs.

A similar agreement was also found for all the other molecular species involved in the pathway. In Fig. 3(b) we report the results for three of the key molecular species involved in the pathway (AC, PKA, and MAPK), each of which is the final product of one of the three subnetworks. For each of these proteins, we compare the amount of its active form in our SSA results (solid lines) with the results from the original model (dashed lines).

Note that, in the SBML model, the amounts of species are given in terms of numbers of molecules over compartment sizes and, hence, the amounts must be rescaled by multiplying by the compartment sizes in order to be comparable with our results, which are expressed in terms of molecular counts. AC is located on the cell membrane and, therefore, it must be multiplied by the area of the membrane ( $0.2 \mu\text{m}^2$ ); in the case of PKA and MAPK the amounts result the same, since the volume of the cytoplasm is  $1 \mu\text{m}^3$ .

## 5.2 Experimentation

In this section we report some of the computational experiments we have performed on the model. First, we have studied the effect, on the final product of the pathway (MAPK-P, i.e. phosphorylated MAPK), of changes in the network structure: we have removed some of the feedback loops, in particular the ones involving PKA, which is one of the proteins mostly responsible for both positive and negative regulation of various pathway components.

Figure 4 shows the effect of inhibiting some important steps on the behaviour of the system. In Fig. 4(a) we compare the behaviour of the original system (blue line) with the behaviour of three modified systems: one with weaker enzymatic activity of PKA (10-fold increase of  $K_{m\_v08}$ ) in the phosphorylation of B-Raf (red line), one with slower activation of AC (10-fold decrease of  $K_f\_AC\_activation$  in the binding reaction between AC and  $G_{\alpha s} : GTP$ ) (green line), and one with slower cAMP/R2C2 binding (10-fold decrease of  $K_{f\_v13}$ ,  $K_{f\_v18}$ ,  $K_{f\_v23}$ ,  $K_{f\_v32}$ ) (black line). Each of these modified systems corresponds to slowing down one reaction rate in each of the subnetworks; Fig. 4(a) shows that a reduction

in the rate of cAMP binding has the strongest effect: an expected result, since the cAMP binding reaction must occur four times for a single molecule of PKA to become active.

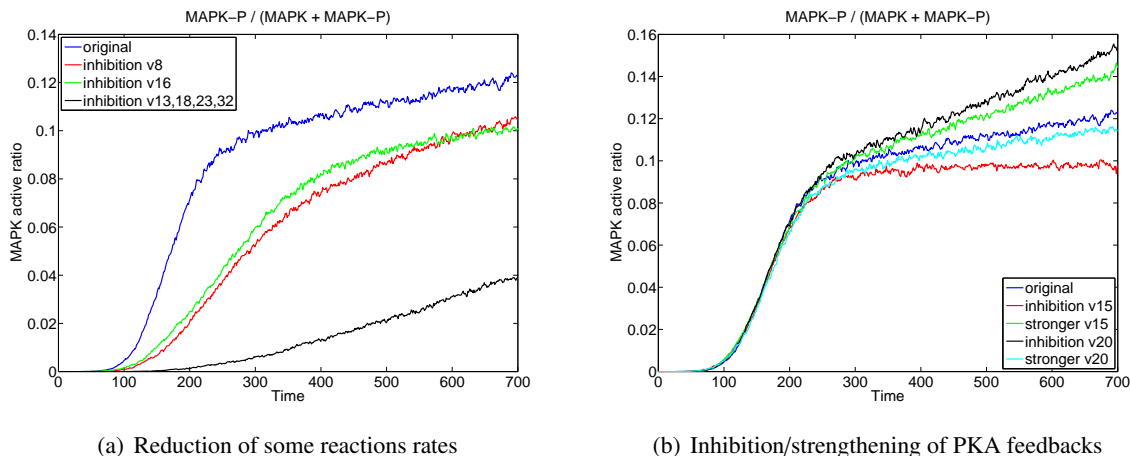


Figure 4: Experimenting on reaction rates and feedback loops (each line is the average behaviour over 500 independent stochastic simulation runs).

In Fig. 4(b), instead, we experiment with the model by changing the parameters relative to the enzymatic activity of PKA and studying their effect on the system behaviour. As described in Sect. 4, PKA takes part in a positive feedback loop by phosphorylating PTPs, and in a negative feedback loop by phosphorylating PDE4.

The results shown in Fig. 4(b) confirm the presence of these feedback loops: a weaker enzymatic activity of PKA (100-fold increase of  $K_m$ -v15) in the phosphorylation of PTP (red line) leads to a decrease of MAPK-P (due to the lower presence of phosphorylated PTP, which is less effective than its unphosphorylated form in dephosphorylating MAPK), while a stronger enzymatic activity of PKA (100-fold decrease of  $K_m$ -v15) (green line) leads to an increase of MAPK-P. The opposite occurs for the phosphorylation of PDE4: a weaker enzymatic activity of PKA (100-fold increase of  $K_m$ -v20) in the phosphorylation of PDE4 (black line) leads to an increase of MAPK-P (due to the lower presence of phosphorylated PDE4, which is more effective than its unphosphorylated form in transforming cAMP into AMP), while a stronger enzymatic activity of PKA (100-fold decrease of  $K_m$ -v20) (cyan line) leads to a decrease (though smaller) of MAPK-P.

## 6 Related work

Several languages have been defined to model biological systems with compartments and membranes [30, 6, 34, 32, 7, 41, 27]. In the following we report a brief overview of the most well-known languages and the associated analysis tools.

*Membrane systems* or *P systems* [30, 9] are computational models that are based on a notion of membrane structure. In particular, multisets of objects are enclosed in a nested hierarchy of membranes and their behaviour is described by local rewriting rules.

The analysis of P systems is supported by the *P System Modelling Framework* [19]. This framework provides stochastic simulation based on the multi-compartmental Gillespie algorithm [31] and offers the

mapping to PRISM [25, 33] for model checking. Furthermore, translations to *SBML* [36] and *CellDesigner* [8] have been implemented in order to export P system models into other formats for further kinds of analysis and for graphical representation.

The variant of membrane systems with peripheral and integral membrane proteins defined by Cavaliere and Sedwards [7] is at the basis of *Cyto-Sim* [38, 14]. The analysis is limited to stochastic simulation by Gillespie's algorithm [22], but it is possible to export the corresponding ODE model into MATLAB format [15] for further analyses. In addition to mass-action kinetics, the simulator supports chemical reactions with arbitrary kinetic laws based on functions of the reactants.

One of the first process calculi with an explicit notion of compartments is *BioAmbients* calculus [34]. A BioAmbient system is seen as a hierarchy of nested *ambients*, generally abstracting compartments, containing communicating processes whose actions specify the evolution of the system. Various kinds of action involving compartments can be easily represented, such as transport of small molecules across compartments. BioAmbients models are supported by two tools: *Bio-SPI* [39], a stochastic simulator based on Gillespie's algorithm and originally implemented for models in stochastic  $\pi$ -calculus, and the *BioAmbient machine (BAM)* [43, 2], a more efficient and user-friendly simulator written in Java. BAM is based on the stochastic version of BioAmbients, defined by Vigliotti and Harrison [42] and can be used to simulate bio-regulatory pathways as well as membrane interaction. BAM automatically produces both a simulation graph and a debugging trace of the program. Furthermore, it offers a graphical user interface for entering program code and visualising simulation results.

A more active role for membranes can be described using *Brane calculi* [6]. In Brane calculi, a system is represented as a set of nested membranes, and a membrane is represented as a set of actions. Membranes can move, merge, split, enter into and exit from other membranes.

The *bio $\kappa$ -calculus* [27] is a language for describing proteins and cells based on Brane calculi and  $\kappa$ -calculus [16]. The atomic elements of the calculus are *proteins* and two constructors for representing *solutions* and *cells*. Proteins are characterised by sites, that can be in different states: *bound* to another site of a protein, *visible* (i.e not connected to other sites) and *hidden* (i.e not connected to other sites but not available for other interactions). Cells abstract compartments and consist of a membrane and a cytoplasm. Reactions can be interactions between two proteins or between two membranes.

In *Beta-binders* [32] and the associated *BlenX* language [17], systems are modelled as a composition of boxes representing biological entities. In the original version of the language compartments are not defined explicitly, but a virtual form of nesting is rendered by appropriate typing for sites. Explicit static compartments and transport of biological across them have been added to Beta-binders in [24]. The analysis of BlenX models is possible in the *Beta Workbench (BetaWB)* [3]. The BetaWB is composed of a stochastic simulator based on an efficient variant of Gillespie's algorithm, the BetaWB designer, a graphical editor for developing models and the BetaWB plotter, a tool to analyse the results of a stochastic simulation run. Furthermore, BlenX models can be exported in SBML.

Finally, there are some variants of the stochastic  $\pi$ -calculus enriched with the notion of locations (see [41, 26], for instance). *S $\pi$ @* calculus [41] extends the stochastic  $\pi$ -calculus syntax with the explicit addition of compartments. This language handles varying volumes and dynamic compartments by defining the compartment volume as the sum of the volumes occupied by all the molecules it contains.

Another extension of the  $\pi$ -calculus with compartments is *SpacePi* [26]. Pi processes are embedded into a vector space and move individually. Only processes that are sufficiently close can communicate. The operational semantics of SpacePi defines the interplay between movement, communication, and time-triggered events.

All the languages listed above differ from Bio-PEPA in various aspects: they are based on different levels of abstraction or focus on dynamical compartments or handle volumes in a distinct way. In

Bio-PEPA, locations have a fixed structure and compartments are essentially containers for biological species: interactions between molecules in different compartments is allowed, but the main evolution is given by interactions of molecular species within compartments. This concept of location is that usually considered in biochemical networks in databases and in the literature. For these systems Bio-PEPA offers a direct and formal representation, allowing an intuitive representation of both intra-compartment and inter-compartment reactions.

Concerning the Bio-PEPA Eclipse Plug-in, a prominent feature of the tool is its support of both stochastic simulators and ODE numerical solvers. The ability to utilise different analysis techniques is very important; this can aid in understanding different behavioural aspects of the system and discovering possible errors due to the use of a specific solver/simulator [5]. Furthermore, the user can select the most appropriate analysis method based on the model under study. Most of the other tools support stochastic simulation by Gillespie's algorithm only and in a few cases mappings to other languages or tools, such as SBML or PRISM.

## 7 Conclusions

Interactions of molecular species located in compartments with different sizes occur typically in biological systems. In signalling pathways, for instance, a biochemical signal is transferred from the extra-cellular environment into the cell through the intervention of membrane receptors and other membrane-bound proteins.

In this work we have presented a Bio-PEPA compartmental model of the cAMP/PKA/MAPK pathway. Our main aim was to demonstrate, on a real case study, the effectiveness and the correctness of the representation of multi-compartment models in Bio-PEPA.

The Bio-PEPA language allows us to explicitly define locations representing compartments and membranes, to express the location of each species and to indicate the relative position of a location with respect to the others. Furthermore, Bio-PEPA correctly deals with the location sizes (volumes and surface areas). It is generally useful to collect these kinds of information, in order to have a better view of the system under consideration. Moreover, the information about the size of locations is necessary to compare the results obtained in Bio-PEPA with the ones in the literature and experiments, when given in concentrations instead of number of molecules. Another benefit is although the system in question can be written as a set of ODEs, high-level languages like Bio-PEPA can offer a cleaner view of the model through a more restricted (but focussed) number of operators. Through these operators the modeller must explicitly state any behaviours he or she wishes the components of the system to exhibit. A restricted set of operators also allows the tool to perform more extensive checking over any legal Bio-PEPA model, something not possible if manually coding a set of ODEs in a general tool such as Matlab.

In order to be able to validate our results, we developed our Bio-PEPA model as a faithful representation of an existing ODE model.

The analysis of the pathway was performed using the stochastic simulator available within the Bio-PEPA Eclipse Plug-in, which is able to handle multi-compartmental models. The comparison of our results with the ones obtained from the original ODE model demonstrate the correctness of our approach. We also perform a number of simulation experiments in order to investigate specific properties of the system such as its response to the inhibition of feedback mechanisms and to the variation of key molecular species.

In this model we have found it convenient to describe the behaviour of the pathway in terms of three functional elements, a feature supported by the tool. Currently such modularisation is supported only

at the level of model definition as a convenience for the modeller. However it is an interesting area for future work to explore the extent to which such structuring might be exploited during model analysis.

While the locations extension for Bio-PEPA does not restrict species interaction by location, future improvements within the tool will flag such reactions. Once implemented this will allow the modeller to clearly see which reactions involve species not existing in adjacent locations. Other future work in the Bio-PEPA Eclipse Plug-in includes improved support for locations as well as the implementation for automatic import/export of existing specification languages supporting multi-compartment models (e.g. [36, 35, 23]).

Finally, we would like to extend the analysis of this model (and of multi-compartmental models in general) by applying some of the CTMC-based analysis methods (e.g. model-checking) which are supported by the Bio-PEPA language.

## Acknowledgements

The authors thank Jane Hillston for her helpful comments. At the time of writing this paper, Federica Ciocchetta was a research fellow at the University of Edinburgh, supported by the EPSRC grant EP/C543696/1 “Process Algebra Approaches to Collective Dynamics”. Adam Duguid is supported by the EPSRC Doctoral Training Grant EP/P501407/1. Maria Luisa Guerriero is supported by the EPSRC grant EP/E031439/1 “Stochastic Process Algebra for Biochemical Signalling Pathway Analysis”.

## Supplementary Material

The Bio-PEPA source files and additional information on the model are available online on the Bio-PEPA web page at [http://homepages.inf.ed.ac.uk/jeh/Bio-PEPA/CDG\\_MecBIC09\\_Supplement.zip](http://homepages.inf.ed.ac.uk/jeh/Bio-PEPA/CDG_MecBIC09_Supplement.zip).

## References

- [1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts & P. Walter (2002): *Molecular biology of the cell (IV ed.)*. Garland Science.
- [2] *The BioAmbients Machine Home Page*. <http://aesop.doc.ic.ac.uk/tools/bam/>.
- [3] *The BetaWB Home Page*. <http://www.cosbi.eu/index.php/research/prototypes/beta-wb>.
- [4] Bio-PEPA Home Page. <http://www.biopepa.org/>.
- [5] M. Calder, A. Duguid, S. Gilmore & J. Hillston (2006): *Stronger computational modelling of signalling pathways using both continuous and discrete-state methods*. In: *Proc. of CMSB'06, LNCS 4210*. pp. 63–77.
- [6] L. Cardelli (2005): *Brane Calculi - Interactions of Biological Membranes*. In: *Proc. of CMSB'04, LNCS 3082*. pp. 257–278. Available at <http://lucacardelli.name/Papers/Brane%20Calculi%20CMSB%20LNCS.pdf>.
- [7] M. Cavaliere & S. Sedwards (2006): *Modelling Cellular Processes Using Membrane Systems with Peripheral and Integral Proteins*. In: *Proc. of CMSB'06, LNCS 194*. pp. 108–126.
- [8] *Cell Designer Home Page*. <http://celldesigner.org>.
- [9] G. Ciobanu, M.J. Pérez-Jiménez & G. Păun, editors (2006): *Applications of Membrane Computing*. Natural Computing Series. Springer.
- [10] F. Ciocchetta, A. Degasperri, J. Hillston & M. Calder (2009): *Some Investigations Concerning the CTMC and the ODE Model Derived from Bio-PEPA*. In: *Proc. of FBTC'08, ENTCS 229*. pp. 145–163.
- [11] F. Ciocchetta, A. Duguid, S. Gilmore, M.L. Guerriero & J. Hillston (2009, to appear): *The Bio-PEPA Tool Suite*. In: *Proc. of QEST'09*.

- [12] F. Ciocchetta & M.L. Guerriero (2009): *Modelling Biological Compartments in Bio-PEPA*. In: *Proc. of MeCBIC'08, ENTCS 227*, pp. 77–95.
- [13] F. Ciocchetta & J. Hillston (2009): *Bio-PEPA: a Framework for the Modelling and Analysis of Biological Systems*. *Theoretical Computer Science* 410(33-34), pp. 3065–3084.
- [14] *Cyto-Sim*. <http://www.cosbi.eu/index.php/prototypes/cyto-sim>.
- [15] *MATLAB*. <http://www.mathworks.it/>.
- [16] V. Danos & C. Laneve (2004): *Formal molecular biology*. *Theoretical Computer Science* 325(1), pp. 69–110.
- [17] L. Dematté, C. Priami & A. Romanel (2008): *Modelling and simulation of biological processes in BlenX*. *SIGMETRICS Performance Evaluation Review* 35(4), pp. 32–39.
- [18] Eclipse Home Page. <http://www.eclipse.org/>.
- [19] *P systems Modelling Framework*. <http://www.dcs.shef.ac.uk/~marian/PSimulatorWeb/Tools.htm>.
- [20] M.A. Gibson & J. Bruck (2000): *Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels*. *The Journal of Chemical Physics* 104, pp. 1876–1889.
- [21] Daniel T. Gillespie & Linda R. Petzold (2003): *Improved leap-size selection for accelerated stochastic simulation*. *The Journal of Chemical Physics* 119(16), pp. 8229–8234. Available at <http://link.aip.org/link/?JCP/119/8229/1>.
- [22] D.T. Gillespie (1977): *Exact stochastic simulation of coupled chemical reactions*. *J Phys Chem* 81(25), pp. 2340–2361.
- [23] M.L. Guerriero, J.K. Heath & C. Priami (2007): *An Automated Translation from a Narrative Language for Biological Modelling into Process Algebra*. In: *Proc. of CMSB'07, LNCS 4695*, pp. 136–151.
- [24] M.L. Guerriero, C. Priami & A. Romanel (2007): *Modeling Static Biological Compartments with Beta-binders*. In: *Proc. of AB 2007, LNCS 4545*, pp. 247–261.
- [25] A. Hinton, M. Kwiatkowska, G. Norman & D. Parker (2006): *PRISM: A tool for automatic verification of probabilistic systems*. In: *Proc. of TACAS'06, LNCS 3920*, pp. 441–444.
- [26] M. John, R. Ewald & A.M. Uhrmacher (2008): *A Spatial Extension to the  $\pi$  Calculus*. In: *Proc. of QAPL 2006, ENTCS 194*, pp. 133–148.
- [27] C. Laneve & F. Tarissan (2008): *A simple calculus for proteins and cells*. *Theoretical Computer Science* 404(1-2), pp. 127–141.
- [28] N. Le Novère, B. Bornstein, A. Broicher, M. Courtot, M. Donizelli, H. Dharuri, L. Li, H. Sauro, M. Schilstra, B. Shapiro, J.L. Snoep & M. Hucka (2006): *BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems*. *Nucleic Acids Research* 34(Database issue), pp. D689–D691.
- [29] S.R. Neves, P. Tsokas, A. Sarkar, E.A. Grace, P. Rangamani, S. Taubenfeld, C.M. Alberini, J.C. Schaff, R.D. Blitzer, I.I. Moraru & R. Iyengar (2008): *Cell Shape and Negative Links in Regulatory Motifs Together Control Spatial Information Flow in Signaling Networks*. *Cell* 133, pp. 666–680.
- [30] G. Păun & G. Rozenberg (2002): *A guide to membrane computing*. *Theoretical Computer Science* 287(1), pp. 73–100. Available at <http://scienceserver.cilea.it/pdflinks/06090213054602466.pdf>.
- [31] M.J. Pérez-Jiménez & F.J. Romero-Campero (2006): *P Systems, a New Computational Modelling Tool for Systems Biology*. *Transactions on Computational Systems Biology* 6, pp. 176–197.
- [32] C. Priami & P. Quaglia (2005): *Operational patterns in Beta-binders*. *Transactions on Computational Systems Biology* 1, pp. 50–65. Available at <http://springerlink.metapress.com/openurl.asp?genre=article&issn=0302-9743&volume=3380&page=50>.
- [33] PRISM Home Page. <http://www.prismmodelchecker.org>.
- [34] A. Regev, E.M. Panina, W. Silverman, L. Cardelli & E.Y. Shapiro (2004): *BioAmbients: an Abstraction for Biological Compartments*. *Theoretical Computer Science* 325(1), pp. 141–167. Available at <http://lucacardelli.name/Papers/BioAmbients%20An%20Abstraction%20for%20Biological%20Compartments.pdf>.



- [35] *SBGN Home Page*. <http://www.sbgn.org>.
- [36] *SBML Home Page*. <http://www.sbml.org>.
- [37] *BioModels Database model 182*. <http://www.ebi.ac.uk/biomodels-main/BIOMD00000000182>.
- [38] S. Sedwards & T. Mazza (2007): *Cyto-Sim: A formal language model and stochastic simulator of membrane-enclosed biochemical processes*. *Bioinformatics* 23(20), pp. 2800–2802.
- [39] W. Silverman. *BioSpi home page*. <http://www.wisdom.weizmann.ac.il/~biopsi>.
- [40] J. Stoer & R. Bulirsch (1993): *Introduction to Numerical Analysis*. New York: Springer-Verlag.
- [41] C. Versari & N. Busi (2008): *Efficient Stochastic Simulation of Biological Systems with Multiple Variable Volumes*. In: *Proc. of FBTC'07, ENTCS 194*. pp. 165–180.
- [42] M.G. Vigliotti & P.G. Harrison (2007): *Stochastic Ambient Calculus*. In: *Proc. of QAPL 2006, ENTCS 164*. pp. 169–186.
- [43] A.M. Vinod, A. Phillips & M.G. Vigliotti (2008): *BAM: BioAmbient Machine*. In: *Proc. of ACSD 08, IEEE Computer Society*. pp. 45–49.