

From the PEPA reagent-centric view to Bio-PEPA

Federica Ciocchetta * Jane Hillston[†]

Abstract

PEPA has been recently applied for the modelling and the analysis of some biological pathways. PEPA is useful in this field, but unfortunately not all the features of biological systems can be considered. In this work we show a first version of Bio-PEPA, a new language that extends PEPA in order to model biochemical networks.

1 Introduction

In the recent years there has been an increasing interest in the application of process algebras in the modelling and analysis of biological systems. PEPA [6], originally defined for the performance analysis of computer systems, has been recently applied in the context of signalling pathways [2, 3]. PEPA has been shown to be useful in the study of biological systems: it offers compositionality and defines a formal high-level representation of the biological model, on which different kinds of analysis can be carried out. Here we consider one of the possible approaches, the so-called *reagent-centric view*: the concentrations are discretised into levels, each level representing an interval of concentration values and processes represent the species at the different concentration levels. The level l can assume values between 0 and N_{max} (maximum level). Unfortunately, in the present version of PEPA not all the features of biochemical models can be represented. The main drawbacks are the definition of *stoichiometry* and the representation of *kinetic laws*. This second aspect is particularly relevant, as reaction with kinetic laws different from the basic mass-action (hereafter called *general kinetic laws*) are frequently found in models as abstractions of complex situations whose details are unknown. Reducing to the elementary steps is complex and often impracticable. This problem impacts also on other process algebras. Previous work concerning the use of general kinetic laws in process algebras has been proposed in [1].

Here we present a first version of Bio-PEPA, a language for the modelling and the analysis of biochemical networks. This is a modification of PEPA reagent-centric view [2] in order to represent explicitly some features of biochemical models, such as stoichiometry and the role of the different species in a given reaction. A major feature of Bio-PEPA is the introduction of functional rates to express general kinetic laws. The application context is the one of biochemical networks. Broadly speaking a biochemical system \mathcal{M} is composed of some species that interact each other through some

*Federica Ciocchetta, Laboratory of Foundation of Computer Science, University of Edinburgh, Scotland
Email: fciocche@inf.ed.ac.uk

[†]Jane Hillston, Laboratory of Foundation of Computer Science, University of Edinburgh, Scotland
Email: jeh@inf.ed.ac.uk

reactions. We consider only the case of static compartments these are not represented explicitly.

The idea underlying the new language is schematized in the following diagram:

$$\text{Biochemical networks} \longrightarrow \text{Bio-PEPA system} \longrightarrow \text{Analysis}$$

We start from a biological model and from it we derive the Bio-PEPA specification. This is an *intermediate, formal, compositional* representation of the biological model. At this point we can apply different kinds of analysis, from stochastic simulation by Gillespie, to analysis based on differential equations and continuous time Markov chains (CTMC).

2 Bio-PEPA

The syntax and the semantics. The PEPA syntax is modified in the following way:

$$S(l) := (\alpha, \kappa)\{l\} \text{ op } S(l) \mid S(l) + S(l) \mid C(l) \quad P := P \underset{\tau}{\bowtie} P \mid S(l)$$

where $\text{op} = \downarrow \mid \uparrow \mid \oplus \mid \ominus \mid \odot$. We suppose a countable set of components C and a countable set of action types \mathcal{A} .

The component $S(l)$ is called *sequential component* (or *species component*) and it is used to represent the species. It is parametrised by a parameter $l \in \mathbb{N}$, representing the level of concentration. The component P , called a *model component*, is used to describe the system and the interactions among components. The prefix term in PEPA is replaced by a new one, $(\alpha, \kappa)\{l\} \text{ op } S(l)$, containing the information about the impact of the species with respect to the reaction described by the action type $\alpha \in \mathcal{A}$: κ is the *stoichiometry coefficient* of the species in that reaction; l is the concentration level; the *prefix combinator* “op” represents the role of the element in the reaction. Specifically, \downarrow is used to indicate a *reactant*, \uparrow a *product*, \oplus an *activator*, \ominus an *inhibitor* and \odot a *generic modifier*.

For each species S_i we need to define the maximum concentration M_i (in molar) and the maximum level $N_i \geq 1$. With \mathcal{N} we indicate the list of all the components “ $S_i : M_i, N_i$ ” for each species S_i in the model.

In order to describe the dynamics of the system, we need to associate each action α_j with a functional rate f_{α_j} . The definition of the function is “ $f_{\alpha}(\bar{k}, \bar{C}) = \text{expression}$ ”, where \bar{k} is a set of parameters, \bar{C} is a set of name components and “*expression*” stands for an arithmetic expression that represents the kinetic law. The functional rates are defined externally to the components and are evaluated at the moment of the derivation of the system.

Concerning the semantics of Bio-PEPA, the rules are reported in Table 1.

We define the labels $\theta \in \Theta$ as $\theta := (\alpha, \nu)$, where ν is defined as $\nu := [] \mid [S : \text{op}(l, \kappa)]@v$, with $S \in C$, l the level and κ the stoichiometry coefficient of the components. The former three axioms describe the behaviour of the three different prefix terms. The rules `choice1` and `choice2` have the usual meaning. The rule `constant` is used to define the behaviour of the constant term, defined by one or by sum of more prefix terms. The label contains the information about the level and the stoichiometric coefficient related to the action α . The last three rules report the case of cooperation. The rule `coopFinal` describes the case in which the two components synchronise and the label reports the information from both the components.

prefixRec	$\frac{}{(\alpha, \kappa)\{l\}\downarrow S(l-1) \xrightarrow{(\alpha, [(\alpha, \kappa)\{l\}\downarrow S(l-1): \downarrow(l, \kappa))} S(l-1)}$
prefixProd	$\frac{}{(\alpha, \kappa)\{l\}\uparrow S(l+1) \xrightarrow{(\alpha, [(\alpha, \kappa)\{l\}\uparrow S(l+1): \uparrow(l, \kappa))} S(l+1)}$
prefixMod	$\frac{}{(\alpha, \kappa)\{l\}opS(l) \xrightarrow{(\alpha, [(\alpha, \kappa)\{l\}opS(l): op(l, \kappa))} S(l)} \quad \text{with } op = \odot, \oplus, \ominus$
Choice1	$\frac{S_1(l) \xrightarrow{(\alpha, v)} S'_1(l')}{S_1(l) + S_2(l) \xrightarrow{(\alpha, v)} S'_1(l')}$
Choice2	$\frac{S_2(l) \xrightarrow{(\alpha, v)} S'_2(l')}{S_1(l) + S_2(l) \xrightarrow{(\alpha, v)} S'_2(l')}$
Constant	$\frac{S(l) \xrightarrow{(\alpha, S': [op(l, \kappa)]} S'(l')}{C(l) \xrightarrow{(\alpha, C: [op(l, \kappa)]} S'(l')} \quad \text{with } C(l) \stackrel{def}{=} S(l)$
coop1	$\frac{P_1 \xrightarrow{(\alpha, v)} P'_1}{P_1 \underset{\mathcal{L}}{\boxtimes} P_2 \xrightarrow{(\alpha, v)} P'_1 \underset{\mathcal{L}}{\boxtimes} P_2} \quad \text{with } \alpha \notin \mathcal{L}$
coop2	$\frac{P_2 \xrightarrow{(\alpha, v)} P'_2}{P_1 \underset{\mathcal{L}}{\boxtimes} P_2 \xrightarrow{(\alpha, v)} P_1 \underset{\mathcal{L}}{\boxtimes} P'_2} \quad \text{with } \alpha \notin \mathcal{L}$
coopFinal	$\frac{P_1 \xrightarrow{(\alpha, v_1)} P'_1 \quad P_2 \xrightarrow{(\alpha, v_2)} P'_2}{P_1 \underset{\mathcal{L}}{\boxtimes} P_2 \xrightarrow{(\alpha, v_1 @ v_2)} P'_1 \underset{\mathcal{L}}{\boxtimes} P'_2} \quad \text{with } \alpha \in \mathcal{L}$

Table 1: Axioms and rules for Bio-PEPA.

In order to associate the rates with the transitions we need to consider a new relation $\leftrightarrow \subseteq C \times \Gamma \times C$, where the label $\gamma \in \Gamma$ is defined as $\gamma := (\alpha, r)$, with $r \in \mathbb{R}^+$. In this definition r represents the parameter of an exponential distribution. As usual, the dynamic behaviour of processes is determined by a *race condition*: all activities enabled attempt to proceed but only the fastest succeeds.

The relation \leftrightarrow is defined as the minimal relation satisfying the rule

$$\text{Final} \quad \frac{P \xrightarrow{(\alpha_j, v)} P'}{P \xrightarrow{(\alpha_j, f_\alpha(v, \mathcal{N}))} P'}$$

The second component in the label of the conclusion represents the rate associated with the transition. The notation $f_\alpha(v, \mathcal{N})$ means that the function f_α is evaluated over the list of quantitative information v and the set \mathcal{N} of maximum concentration/number of levels. We can define the *Quantitative Labelled Transition System QLTS* as $(C, \Gamma, \leftrightarrow)$. From QLTS it is possible to derive the CTMC as usual:

Theorem 1 Given $(C, \Gamma, \leftrightarrow)$, let $P \in C$ and $n = |ds(P)|$, where $ds(P)$ is the set of all the derivative of P . Then the generator matrix of the CTMC for P is a square matrix Q $n \times n$ whose elements $q_{h,k}$ are defined, for some action type $\alpha_j \in \mathcal{A}$, as

$$q_{h,k} = \sum_{\mathcal{A}(P_h|P_k)} f_{\alpha_j}(v_h, \mathcal{N}) \quad \text{if } h \neq k \quad \quad q_{h,h} = - \sum_{h \neq k} q_{h,k} \quad \text{otherwise}$$

where v_h and \mathcal{N} are defined above. The notation $\mathcal{A}(P_h|P_k)$ indicates all the actions enabled in the derivative P_h that lead to P_k .

The abstraction. The translation of a biochemical network \mathcal{M} into Bio-PEPA is based on the following abstraction:

1. each species $S_i \in \mathcal{S}$ in the network is described by a species component $C_i(l)$. This is defined by the “sum” of *elementary components* describing the interaction capabilities of the species S_i . We suppose that there is at most one term in each species component with an action of type α ;
2. each reaction R_j is associated with an action type α_j and its dynamics is described by a specific function f_{α_j} .

The species components are then composed together to describe the behaviour of the system and the different interactions. The initial levels of the components describe the initial situation.

Let’s consider a simple example to show how the translation work. The basic enzymatic reaction from the substrate S to the product P and it is written as $S + E \rightarrow P + E$, where E is the enzyme. The dynamics is described by the law $fMM((v, k), S, E) = \frac{v * E * S}{(K + S)}$. The three species can be specified in Bio-PEPA by the following components:

$$S(l) = (\alpha, 1)\{l\}\downarrow S(l-1) \quad P(l) = (\alpha, 1)\{l\}\uparrow P(l+1) \quad E(l) = (\alpha, 1)\{l\} \oplus E(l)$$

The system is described by $(S(l_{S0}) \bowtie_{\{\alpha\}} E(l_{E0})) \bowtie_{\{\alpha\}} P(l_{P0})$.

Comparison between PEPA and Bio-PEPA. Bio-PEPA has some differences with respect to the original version of PEPA. Some features are not considered as not necessary in the context of biological systems and some new ones are introduced to model some specific biological aspects. First of all, the PEPA activity is replaced by a more complex prefix containing various information about the role of the biological species in the reaction and the stoichiometry. Secondly, we do not consider the hide operator, there are not silent actions and there is not a definition of passive rate. In Bio-PEPA the rates are positive real number (possibly zero) and are expressed externally by using a function (*functional rate*). Finally, all the reactions are abstracted by a cooperation.

Example: Goldebeter cyclin model This example describes a minimal model for the cascade of post translational modifications that modulate the activity of *cdc2* kinase during the cell cycle, as described by Goldbeter [5]. There are three different variables involved:

- *cyclin*, the protein protagonist of the cycle;
- *cdc2 kinase*, in both active and inactive form;
- *cyclin protease*, in both active and inactive form.

The model is globally described by 7 reactions, whose dynamics are in most cases of the Michaelian type. The translation into Bio-PEPA is straightforward: constant processes are introduced to represent the species and the kinetic laws are expressed by functional rates. The role of each element with respect to the various reactions and the stoichiometric coefficients are shown explicitly. From the Bio-PEPA model we can analyze the model by using ODEs, Gillespie and CTMCs.

The direct translation into PEPA is not possible, as in the model we have some complex reactions described by Michaelis-Menten law. In order to translate the model into PEPA we should replace the complex reactions with simple elementary steps. In this way the model increases in complexity (we would obtain 13 species and 14 reactions) and there is the problem of finding appropriate rates. Generally this transformation is difficult and complex.

3 Conclusions

Bio-PEPA is a modification of the process algebra PEPA for the modelling and the analysis of biochemical networks. Bio-PEPA allows us to represent explicitly some features of biological networks, such as stoichiometry and general kinetic laws. The possibility to consider various kinds of kinetic laws permits us to model a vast number of biochemical networks.

One topic for the future concerns a deeper investigation into the relation between Gillespie simulation and general kinetic laws. Furthermore it is necessary to study the properties of CTMC with levels and see the possible relation with the other kinds of analysis. Another future work will concern the possible application of model checking techniques for the analysis of the properties of a Bio-PEPA system. Finally we need to define some bisimulations and equivalences for the language.

References

- [1] Bortolussi, L. and A. Policriti, *Modeling Biological Systems in Stochastic Concurrent Constraint Programming*, Proc. of WCB 2006, 2006.
- [2] Calder, M., S. Gilmore, and J. Hillston, *Modelling the influence of RKIP on the ERK signalling pathway using the stochastic process algebra PEPA*, T. Comp. Sys. Biology, VII, **4230** (2006), 1–23, Springer.
- [3] Calder, M., A. Duguid, S. Gilmore and J. Hillston, *Stronger computational modelling of signalling pathways using both continuous and discrete-space methods*, Proc. of CMSB06, 2006.
- [4] Gillespie, D.T. *Exact stochastic simulation of coupled chemical reactions*, Journal of Physical Chemistry **81** (1977), 2340–2361.
- [5] Goldbeter, A. *A Minimal Cascade Model for the Mitotic Oscillator Involving Cyclin and Cdc2 kinase*. Proc. Nat. Acad. Sci. **8** (1991), 9107–9111.
- [6] Hillston, J., “A Compositional Approach to Performance Modelling”, Cambridge University Press, 1996.