



ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Electronic Notes in
Theoretical Computer
Science

Electronic Notes in Theoretical Computer Science 194 (2008) 103–117

www.elsevier.com/locate/entcs

Bio-PEPA: An Extension of the Process Algebra PEPA for Biochemical Networks

Federica Ciocchetta¹ and Jane Hillston²

*Laboratory for Foundations of Computer Science
The University of Edinburgh
Edinburgh EH9 3JZ, Scotland*

Abstract

In this work we introduce Bio-PEPA, a process algebra for the modelling and the analysis of biochemical networks. It is a modification of PEPA to deal with some features of biological models, such as stoichiometry and the use of generic kinetic laws. Bio-PEPA may be seen as an intermediate, formal, compositional representation of biological systems, on which different kinds of analysis can be carried out. Finally, we show a representation of a model, concerning a simple genetic network, in the new language.

Keywords: Systems Biology, process algebras, biochemical pathways

1 Introduction

In recent years there has been an increasing interest in the application of process algebras in the modelling and analysis of biological systems [19,11,12,18,9,17,6]. Process algebras are appropriate for describing formally and analysing a biological system as a whole and for reasoning about the interactions among genes and proteins. Indeed, biological systems can be abstracted by concurrent systems described by process algebras: species may be seen as processes that can interact with each other and reactions may be modelled by using actions.

The process algebra PEPA, originally defined for the performance analysis of computer systems, has been recently applied in the context of signalling pathways [6,7]. Two different approaches have been proposed: one based on reagents and another based on pathways. 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*. Indeed, stoichiometry is not

¹ Email: fcioche@inf.ed.ac.uk

² Email: jeh@inf.ed.ac.uk

represented explicitly and the reactions are supposed to be elementary (with constant rate). The problem of extending to the domain of kinetic laws different from the basic mass-action (hereafter called *general kinetic laws*) is particularly relevant, as these kinds of reactions are frequently found in models as abstractions of complex situations whose details are unknown. Reducing to the elementary steps is, however, complex and often impracticable. This problem impacts also on other process algebras. Indeed, generally they rely on Gillespie's stochastic simulation for analysis and this considers only elementary reactions. Some recent works have extended the approach of Gillespie to deal with complex reactions [1,8] but these extensions are yet to be reflected in the work using process algebras. Previous work concerning the use of general kinetic laws in process algebras has been proposed in [3]. The authors presented a stochastic extension for Concurrent Constraint Programming (CCP) and show how to apply it in the case of biological systems. The analysis is limited to Gillespie's simulation. The possibility of considering different kinetic laws is also proposed in BIOCHAM [10], a programming environment for modeling biochemical systems, making simulations and querying the model in temporal logic.

In this paper 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 in order to represent explicitly some features of biochemical models, such as stoichiometry and the role of the different species in a given reaction. It allows us to model a general biochemical network. Broadly speaking, biochemical networks consist of some biological species, which interact with each other through chemical reactions. The dynamics is described in terms of some kinetic laws. The level of abstraction proposed is high, species are seen in terms of their interactions and the structure of the elements are unknown. A major feature of Bio-PEPA is the introduction of functional rates to express general kinetic laws. Each action type represents a reaction in the model and it is associated with a functional rate. The rate is evaluated at the moment of the reduction of the system.

We focus on the reagent-centric view: the components represent the levels of concentration of the species. This approach is different from other process algebras, in which there is a correspondence between molecules and processes.

The idea underlying our work is schematized in the following diagram:

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

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).

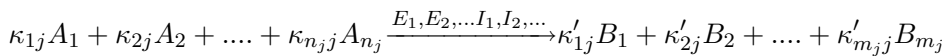
The rest of the paper is organised as follows. Section 2 reports a description of biochemical networks. In Section 3 there is an introduction to PEPA and a summary of the application of PEPA in the modelling of pathways. Section 4 describes Bio-PEPA in detail. In Section 5 we discuss the analysis techniques. After that, Section 6 shows an example of a genetic network represented in Bio-

PEPA. Finally, in Section 7, some conclusions and directions for future work are reported.

2 Biochemical networks

Biological information is often collected in terms of biochemical networks. According to the different kinds of species and interactions involved, we can have *gene regulatory networks*, *signalling pathways* and *metabolic pathways*. In our work we consider a general approach that is valid for most of these networks. A biochemical system \mathcal{M} is composed of:

- (i) a set of species $\mathcal{S} = \{S_1, S_2, \dots, S_N\}$, where $N \geq 1$. These species may be genes, proteins, etc. For each S_i an initial concentration ic_i is given;
- (ii) a set of *compartments* \mathcal{C} (optional), to represent the places where the various species are;
- (iii) a set of reactions $\mathcal{R} = \{R_1, R_2, \dots, R_M\}$, with $M \geq 1$. The general form of a reaction R_j is given by:



where A_h $h = 1, \dots, n_j$, are the reactants, B_l , $l = 1, \dots, m_j$, are the products, E_v are the enzymes and I_u the inhibitors. The parameters κ_{hj} and κ'_{lj} are the stoichiometry coefficients. These refer to quantitative relationships of the reactants and products in reactions and express the degree to which species participate in a reaction. The dynamics associated with the reaction is described by a kinetic law $f_j = f(\bar{k}, \bar{S}_j)$, where \bar{k} is a vector of parameters and \bar{S}_j is a list of some species in \mathcal{S} .

The most well-known kinetic law is *mass action*: the rate of the reaction is proportional to the product of the reactants' concentrations. In models it is common to find *general kinetic laws*, which describe approximations of sequences of reactions. They are useful when it is difficult to derive certain information from the experiments, e.g. the reaction rates of elementary steps, or when there are different time-scales for the reactions. Generally these laws are valid under some conditions, such as the *quasi-steady-state assumption (QSSA)*. This describes the situation where one or more reaction steps may be considered faster than the others and so the intermediate elements can be considered to be constant. There is a long list of kinetic laws, for details see [20].

3 PEPA and biological systems

PEPA was originally defined for the performance modelling of systems with concurrent behaviour [15]. Systems are represented as the composition of components or agents which undertake actions. In PEPA actions are assumed to have a duration, that is represented by a random variable with an exponential distribution.

PEPA has a small set of combinators that allows the system description to be built up as the concurrent interaction of simple sequential components. We introduce informally the syntax of the language below. For more details see [15].

Prefix The basic term is the *prefix combinator* $(\alpha, r).S$. It denotes a component which has action of type α and an exponentially distributed duration with parameter r , and it subsequently behaves as S .

Choice The component $S + R$ represents a system which may behave either as S or as R . The activities of both S and R are enabled. The first activity to complete distinguishes one of them and the other is discarded.

Constant Constants are components whose meaning is given by a defining equation $C \stackrel{def}{=} S$. They allow us to assign names to patterns of behaviour associated with components.

Hiding In the hiding operator S/\mathcal{H} the set \mathcal{H} identifies those activities which can be considered internal or private to the component S .

Cooperation The term $S \bowtie_{\mathcal{L}} R$ denotes cooperation between S and R over the cooperation set \mathcal{L} , that determines those activities on which the cooperands are forced to synchronise. PEPA supports *multiway cooperations* between components: the result of synchronising on an activity α is thus another α , available for further synchronisation. For action types not in \mathcal{L} , the components proceed independently and concurrently with their enabled activities. In the context of performance evaluation the rate for the synchronised activities is the minimum of the rates of the synchronising activities.

PEPA has a structured operational semantics which generates a labelled transition system and from this a CTMC is derived.

Recently, PEPA has been applied to the modelling and analysis of signalling pathways. A first study concerns the influence of the Raf Kinase Inhibitor Protein (RKIP) on the Extracellular signal Regulated Kinase (ERK) [6], whereas in [7] the PEPA system for Schoeberl's model [13] involving the MAP kinase and EFG receptors is reported. In [6] two different modelling styles have been proposed, one based on the *reagent-centric view* and the other on the *pathway-centric view*. The former focuses on the variation in the concentrations of the reagents: the concentrations are discretised in levels, each level representing an interval of concentration values. The level l can assume values between 0 and N_{max} (maximum level). The granularity of the representation can vary; the coarsest possibility is $N_{max} = 1$, corresponding to the case of low and high levels. The pathway-centric style provides a more abstract view of the system and emphasises the subpathways. The two representations were shown to be equivalent. In addition to the standard analysis offered by process algebras, in [5] a way to map the system, specified in PEPA, to a system of differential equations, has been proposed.

From these works PEPA has been shown to be appropriate for the modelling of biological systems: it offers a high level of abstraction for the model and focuses on compositionality and on the interactions. Furthermore by using PEPA as a mod-

elling language it is possible to apply different kinds of analysis, not only stochastic simulation, but also differential equations and the study of properties by means of model checking. Differently from other process algebras, here we have the abstraction processes as levels of species concentrations, rather than processes as single molecules.

However not all the features of biochemical networks can be expressed by using the present version of PEPA: the various kinds of kinetics are not considered and stoichiometry is added by hand in the conversion of PEPA into ordinary differential equations (ODEs). With a few exceptions (as in [3]) and a few cases (dimerization), these features cannot be represented in other process algebras as well.

4 Bio-PEPA

The aim of this work is to define a new process algebra in order to model some of the features of biochemical networks that are not possible to represent in PEPA. In particular the new language is able to represent all the reactions in a straightforward way and it deals with stoichiometry and general kinetic laws.

It refers to the *reagent-centric view* of PEPA and extends it in order to satisfy the above requirements. The reagent-centric view allows us to represent all kinds of reactions and it is based on a high level of abstraction similar to the one proposed in biological networks and formalisms such as SBML [2].

In the present work we limit our study to *static compartments*, i.e. compartments that are not actively involved in the reactions. They are basically containers where species are. We do not represent them explicitly. The transport of a species from one compartment to another one is modelled by introducing two components for representing the species in the two compartments. The reaction is simply abstracted by a transformation of one species into another one.

4.1 The syntax and the semantics

The PEPA syntax is modified in the following way in order to collect the biological information we need:

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

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

The component S is called *sequential component* (or *species component*) and it is used to represent the species. The component P , called a *model component*, is used to describe the system and the interactions among components. The parameter $l \in \mathbb{N}$ represents the level of concentration. The prefix term in PEPA is replaced by a new one, $(\alpha, \kappa) \text{ op } S$, containing the information of the role of the species in the reaction associated with the action type α :

- $\alpha \in \mathcal{A}$ is the *action type*;

- κ is the *stoichiometry coefficient* of the species in that reaction;
- 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*.

The choice operator, the cooperation and the definition of constant are as usual. We do not have the hiding operator, as it is not necessary for our purposes.

For each species S_i we need to define the maximum concentration M_i (in molar) and the maximum level $N_i \geq 1$. We use the notation “ $S_i : M_i, N_i$ ” to associate with the species the relative values. 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} . These functions represents the kinetic laws used to describe the dynamics of the associated reaction. The definition of the function is “ $f_{\alpha}(\bar{k}, \bar{C}) = \text{expression}$ ”, where \bar{k} is a set of parameters and \bar{C} is a set of name components, defined in the Bio-PEPA system and corresponding to the species involved in the kinetic law. The term “*expression*” stands for an arithmetic expression representing one of the kinetic laws in [20].

The functional rates are defined externally to the components and are evaluated at the moment of the derivation of the system.

Now we can define the semantics of Bio-PEPA. First of all, the transition labels must contain the quantitative information we need for the evaluation of the functional rate. We define the labels $\theta \in \Theta$ as:

$$\theta := (\alpha, v)$$

where v is defined as $v := [S : op(l, \kappa)] \mid v @ v$, with $S \in \mathcal{C}$, l the level and κ the stoichiometry coefficient of the components. The semantics rules are reported in Table 1.

The former three axioms describe the behaviour of the three different prefix terms. In the case of a reactant, the level decreases, in the case of products the level increases whereas in the case of modifiers the level remains the same. Concerning the level, it must be greater than 0 for the reactants, less than maximum level N for the products and it can have any possible value in the case of modifiers. In all the three cases the label collects the level and the stoichiometry of the associated component. 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 more prefix terms in summation. 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 rules **coop1** and **coop2** concern the case when the action enabled does not belong to the cooperation set. In this case the label in the conclusion contains only the information about the component that fires the action. The rule **coopFinal** describes the case in which the two components synchronise and the label reports the information from both the components. The concatenation operator of lists @ is used for this purpose.

| | | |
|-------------------|--|--|
| | - | |
| prefixReac | $\frac{}{((\alpha, \kappa) \downarrow S)(l) \xrightarrow{(\alpha, [S: \downarrow(l, \kappa)])} S(l-1)}$ | $0 < l \leq N$ |
| prefixProd | $\frac{}{((\alpha, \kappa) \uparrow S)(l) \xrightarrow{(\alpha, [S: \uparrow(l, \kappa)])} S(l+1)}$ | $0 \leq l < N$ |
| prefixMod | $\frac{}{((\alpha, \kappa) op S)(l) \xrightarrow{(\alpha, [S: op(l, \kappa)])} S(l)}$ | with $op = \odot, \oplus, \ominus$ and $0 \leq l \leq N$ |
| Choice1 | $\frac{S_1(l) \xrightarrow{(\alpha, v)} S'_1(l')}{(S_1 + S_2)(l) \xrightarrow{(\alpha, v)} S'_1(l')}$ | Choice2 $\frac{S_2(l) \xrightarrow{(\alpha, v)} S'_2(l')}{(S_1 + S_2)(l) \xrightarrow{(\alpha, v)} S'_2(l')}$ |
| Constant | $\frac{S(l) \xrightarrow{(\alpha, S': [op(l, \kappa)])} S'(l')}{C(l) \xrightarrow{(\alpha, C: [op(l, \kappa)])} S'(l')}$ | with $C \stackrel{def}{=} S$ |
| coop1 | $\frac{P_1 \xrightarrow{(\alpha, v)} P'_1}{P_1 \boxtimes_{\mathcal{L}} P_2 \xrightarrow{(\alpha, v)} P'_1 \boxtimes_{\mathcal{L}} P_2}$ | with $\alpha \notin \mathcal{L}$ |
| coop2 | $\frac{P_2 \xrightarrow{(\alpha, v)} P'_2}{P_1 \boxtimes_{\mathcal{L}} P_2 \xrightarrow{(\alpha, v)} P_1 \boxtimes_{\mathcal{L}} P'_2}$ | with $\alpha \notin \mathcal{L}$ |
| coopFinal | $\frac{P_1 \xrightarrow{(\alpha, v_1)} P'_1 \quad P_2 \xrightarrow{(\alpha, v_2)} P'_2}{P_1 \boxtimes_{\mathcal{L}} P_2 \xrightarrow{(\alpha, v_1 \oplus v_2)} P'_1 \boxtimes_{\mathcal{L}} P'_2}$ | 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 $\mapsto \subseteq \mathcal{C} \times \Gamma \times \mathcal{C}$, 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 \mapsto is defined as the minimal relation satisfying the rule

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

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. Specifically, for each component C_i in v we consider the maximum level N_i and the maximum concentration M_i and we derive the concentration as $l_i * \frac{M_i}{N_i}$. Then we replace each free occurrence of C_i with $(l_i * \frac{M_i}{N_i})^{\kappa_{ij}}$, where κ_{ij} is the stoichiometric coefficient of the species i with respect to the reaction R_j .

We can define the *Quantitative Labelled Transition System QLTS* as $(\mathcal{C}, \Gamma, \mapsto)$. The states of *QLTS* are unequivocally defined in terms of the concentration levels of the system components and the transitions from one state to another represent reactions that cause some changes in the concentration levels of some components.

From *QLTS* is it possible to derive the *CTMC* as usual. The following theorems

are simple adaptations of those proposed in [15].

Theorem 4.1 *For any finite bio-PEPA model P , if we define the stochastic process $X(t)$ such that $X(t) = P_i$ indicates that the system behaves as component P_i at time t , then $X(t)$ is a Markov Process.*

We consider finite models to ensure that a solution for the CTMC is feasible. This is equivalent to supposing that each species in the model has a maximum level of concentration.

Theorem 4.2 *Given $(\mathcal{C}, \Gamma, \mapsto)$, let $P \in \mathcal{C}$ and $n = |ds(P)|$, where $ds(P)$ is the set of all the derivatives 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 types $\alpha_j \in \mathcal{A}$, as*

$$q_{h,k} = \sum_{\{\alpha_j\}_h^k} f_{\alpha_j}(v_h, \mathcal{N}) \quad \text{if } h \neq k \quad \quad q_{h,h} = - \sum_{h \neq k} q_{h,k} \quad \text{otherwise}$$

where v_h and \mathcal{N} are defined above and $\{\alpha_j\}_h^k$ stands for all the types of actions enabled from the state h to the state k .

4.2 From biochemical networks to Bio-PEPA

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

- (i) each species $S_i \in \mathcal{S}$ in the network is described by a species component. We have the following correspondence:

$$S_i \leftrightarrow C_i$$

The constant component C_i 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 . A unique definition can express the behaviour of the species at any level. At level 0 only the actions that involve the element as a product are possible, while in the case of the maximum level, only the actions involving the element as reactant or modifier are enabled.

- (ii) each reaction R_j is associated with an action type α_j and its dynamics is described by a specific function f_{α_j} .
- (iii) *compartments* are not described explicitly.

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

Finally, the definition of a model in Bio-PEPA is based on the following steps:

- (i) definition of the maximum concentration M_i and the maximum number of levels N_i for each species S_i ;
- (ii) definition of the functional rates for each reaction/action type;

- (iii) definition of constants to describe the behaviour of each species;
- (iv) composition of the system.

4.3 Some examples

In the following we report some simple examples in order to show how some biological situations can be specified by using Bio-PEPA.

4.3.1 Example 1: mass action kinetics

A first example concerns a reaction with the Mass Action kinetic law. Consider the reaction $2A + C \rightarrow 3B$, described by the law $r.A^2.C$. The three species can be specified by the syntax:

$$A = (\alpha, 2)\downarrow A \quad C = (\alpha, 1)\downarrow C \quad B = (\alpha, 3)\uparrow B$$

The system is described by $(A(l_{A0}) \boxtimes_{\{\alpha\}} C(l_{C0})) \boxtimes_{\{\alpha\}} B(l_{B0})$, where l_{A0} , l_{B0} and l_{C0} denote the initial level of the three components. The functional rate is $f_\alpha = fMA(r, [A, C])$.

4.3.2 Example 2: Michealis-Menten kinetics

One of the most used kinetic laws is Michaelis-Menten. It describes a 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 involved in the reaction. This reaction is an approximation of a sequence of two reactions, under the quasi-steady state assumption. The whole sequence of reactions is described by the kinetic law $fMM((v, K), S, E) = \frac{v * E * S}{(K + S)}$ (hereafter the symbol “*” stands for the product over real numbers). For more details about the derivation of this kinetic law and the meaning of parameters see [20].

The three species can be specified in Bio-PEPA by the following components:

$$S = (\alpha, 1)\downarrow S \quad P = (\alpha, 1)\uparrow P \quad E = (\alpha, 1) \oplus E$$

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

4.3.3 Example 3: degradation and synthesis of a species

Two particular reactions are the ones describing the degradation and the creation of a species. In this case we need to add two auxiliary species components to represent respectively the *residue* (*Res*) of the reaction and the *creation factor* (*CF*), that can abstract DNA or genes.

Let us consider the degradation reaction $A \rightarrow \emptyset$. We describe this reaction in Bio-PEPA by introducing the component *Res* as the residue/product of the reaction. The two species A and *Res* are defined as:

$$A = (\alpha, 1)\downarrow A \quad Res = (\alpha, 1) \odot Res$$

The component Res is described by one or more sub-terms each of which describes a different degradation reaction.

The synthesis of a species $\emptyset \rightarrow A$ is instead described by a new component CF . In this case the two species A and CF are described by:

$$A = (\alpha, 1)\uparrow A \quad CF = (\alpha, 1)\odot CF$$

In the definition of the components Res and CF we use the symbol \odot to indicate they do not change with the reaction.

5 Analysis

A model in Bio-PEPA is an *intermediate, formal, compositional* representation of the biological model. From it we can derive different kinds of analysis. In Section 4.1 we have described how to derive the CTMC from the transition system associated with a Bio-PEPA model. In this section we discuss briefly the derivation of the *Ordinary Differential Equations (ODEs)* and *Gillespie simulation*.

5.1 ODEs

The derivation of ODEs from the Bio-PEPA model is straightforward. We follow the approach proposed in [5] and extend it to our case.

Starting from a Bio-PEPA model, we define a *stoichiometry matrix* A in which we have a positive entry for products and a negative entry for reactants. A vector v (*kinetic law vector*) is introduced to contain the kinetic laws of each reaction. Concerning the initial conditions, for each species S_i the initial concentration ic_i is derived as $ic_i = l_{i0} * \frac{M_i}{N_i}$, where l_{i0} is the initial level. The ODE system is obtained by defining for each species S_i the equation $\frac{ds_i}{dt} = Av$ (where s_i is the concentration of the species S_i), with the initial condition $s_{i0} = ic_i$. In the case of general kinetics, there are not problems in the derivation of ODEs, as they are derived in this context.

5.2 Gillespie simulation

The translation of a Bio-PEPA model for Gillespie simulation is similar to the case of PEPA and to the approach for ODEs. The main drawbacks are the definition of the rates and the correctness of the approach in the case of general kinetic laws.

Indeed Gillespie's algorithm [14] supposes elementary reactions and constant rates (mass-action kinetics). In the case that the model contains only this kind of reactions the translation is as usual. In the case of non-elementary reactions, and therefore general kinetic laws, it is a widely-used approach to consider them translated directly into a stochastic context. However this is not always a valid approach and some counterexamples have been demonstrated [4]. The authors of [4] showed that, when Gillespie's algorithm is applied to Hill kinetics in the context of the transcription initiation in autoregulated genes, the magnitude of fluctuations is overestimated. The application of Gillespie in the case of general kinetics laws

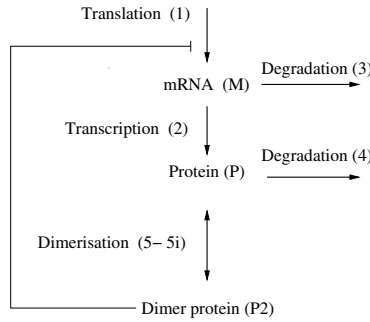


Fig. 1. Genetic network model

is discussed by several authors [1,8]. Rao and Arkin [1] showed that this approach is valid in the case of some specific kinetic laws, such as Michaelis-Menten and inhibition. However it is important to remember that these laws are approximations, are based on some assumptions that conditions (such as “ $S \gg E$ ” in the case of Michaelis-Menten) hold. The approach followed here is as in [16]: we can apply Gillespie in any case, but particular attention must be paid to the interpretation of the simulation results and to their validity.

Some modifications are necessary to translate a Bio-PEPA model for Gillespie.

- (i) We need to derive the initial amounts in terms of number of components. If the concentrations are given in terms of molar, the initial amount ia_i for the species S_i is derived as:

$$ia_i = l_{i0} * \frac{M_i}{N_i} * Na * V \text{ molecules}$$

where Na is the Avogadro number and V the volume of the compartment.

- (ii) The rates must be modified in the appropriate way. In the case of mass action, the *basal rates* necessary in Gillespie’s algorithm are derived from constant rates by using the relations given in [14]. In the case of general kinetic laws the rates correspond to the kinetic laws evaluated in terms of number of elements instead of concentrations as shown in [1].

6 A simple genetic network with negative feedback

In order to show how to model a biological system in our formalism, we consider a model from [4]. The model describes a general genetic network with negative feedback through dimers, such as the one representing the control circuit for the λ repressor protein CI of λ -phage in *E.Coli*.

A schema of the model is reported in Fig.1. The model is composed of three biological entities that interact with each other through five reactions. The biological entities are the mRNA molecule (M), the protein in monomer form (P) and the protein in dimeric form ($P2$).

The first reaction (1) is the translation of the mRNA (M) from the genes/DNA (not considered explicitly). The protein P in the dimer form ($P2$), that is the

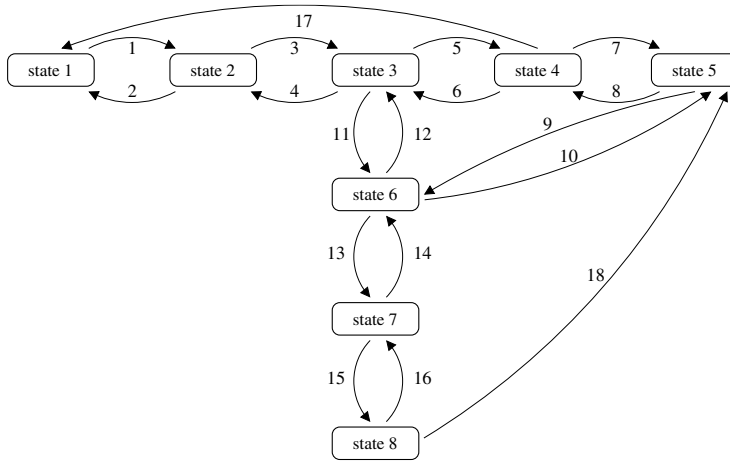


Fig. 2. Transition system in the case of genetic network for two levels of concentration.

final result of the network, has an inhibitory effect on this process. The second reaction (2) is the transcription of the protein P from M . Another two reactions represent the degradation of M (3) and the degradation of P (4). Finally there is the dimerization of P and its inverse process (5,5i). All the reactions are described by mass action kinetics with the exception of the first reaction, that has an Inhibition/Michaelis Menten kinetics. In order to translate the degradation reaction and the translation, the two auxiliary components CF and Res are introduced, as described in Section 4.3.

Let us see how to translate the model into Bio-PEPA. First of all we need to define the maximum concentration and the number of levels for each species by using the list $[M : N_M, M_M; P : N_P, M_P; P2 : N_{P2}, M_{P2}]$. After that, a functional rate is associated with each action/reaction in the following way:

$$\begin{aligned}
 f_{\alpha_1} &= fI((v, K_M), [P2, CF]) = \frac{v * CF}{K_M + P2}; \\
 f_{\alpha_2} &= fMA(k2, [M]); \quad f_{\alpha_3} = fMA(k3, [M]); \quad f_{\alpha_4} = fMA(k4, [P]); \\
 f_{\alpha_5} &= fMA(k5, [P]); \quad f_{\alpha_{5i}} = fMA(k5i, [P2]);
 \end{aligned}$$

where fMA describes mass action kinetics and the suffix of the action type α refers to the number of the reaction as reported in Fig.1. The parameter values are $K_M = 356nM$ and $v = 2.19s^{-1}$, $k2 = 0.043s^{-1}$, $k3 = 0.0039s^{-1}$, $k4 = 0.0007s^{-1}$, $k5 = 0.025s^{-1}nM^{-1}$ and $k5i = 0.5s^{-1}$.

The components to describe species are defined as:

$$\begin{aligned}
 M &= (\alpha_2,1) \oplus M + (\alpha_3,1) \downarrow M + (\alpha_1,1) \uparrow M; \\
 P &= (\alpha_4,1) \downarrow P + (\alpha_5,2) \downarrow P + (\alpha_{5i},1) \uparrow P + (\alpha_2,0) \uparrow P; \\
 P2 &= (\alpha_1,1) \ominus P2 + (\alpha_{5i},1) \downarrow P2 + (\alpha_5,1) \uparrow P2; \\
 Res &= (\alpha_3,1) \odot Res + (\alpha_4,1) \odot Res; \\
 CF &= (\alpha_1,1) \odot CF; \\
 &(((CF(1) \boxtimes_{\{\alpha_1\}} M(0)) \boxtimes_{\{\alpha_2\}} P(0)) \boxtimes_{\{\alpha_5, \alpha_{5i}\}} P2(0)) \boxtimes_{\{\alpha_3, \alpha_4\}} Res(0)
 \end{aligned}$$

The transition system in the case of two levels consists of 8 states and 18 transitions. A schema of the transition system is reported in Fig. 2.

The states are described by the levels of the single components. Specifically, we

can define a state by using a vector $(CF(l_1), M(l_2), P(l_3), P2(l_4), RES(l_5))$, where l_i represents the level of each component. As here we assume only two levels, the parameters l_i can assume only the values 0 and 1. The states reported in Fig. 2 are:

| | |
|--|--|
| state 1 = (CF(1), M(0), P(0), P2(0), RES(0)) | state 2 = (CF(1), M(1), P(0), P2(0), RES(0)) |
| state 3 = (CF(1), M(1), P(1), P2(0), RES(0)) | state 4 = (CF(1), M(0), P(1), P2(0), RES(0)) |
| state 5 = (CF(1), M(0), P(0), P2(1), RES(0)) | state 6 = (CF(1), M(1), P(0), P2(0), RES(0)) |
| state 7 = (CF(1), M(1), P(1), P2(1), RES(0)) | state 8 = (CF(1), M(0), P(1), P2(0), RES(0)) |

whereas the labels θ_t , with the suffix t indicating the transitions in Fig. 2, are:

$$\begin{aligned} \theta_1 = \theta_6 &= (\alpha_1, [CF : \odot(1, 1); P2 : \ominus(0, 1)]); & \theta_9 = \theta_{16} &= (\alpha_1, [CF : \odot(1, 1); P2 : \ominus(1, 1)]); \\ \theta_3 = \theta_{13} &= (\alpha_2, [M : \oplus(1, 1)]); & \theta_7 = \theta_{11} &= (\alpha_5, [P : \downarrow(1, 2)]); & \theta_8 = \theta_{12} &= (\alpha_{5i}, [P2 : \downarrow(1, 1)]); \\ \theta_2 = \theta_5 = \theta_{10} = \theta_{15} &= (\alpha_3, [M : \downarrow(1, 1)]); & \theta_4 = \theta_{14} = \theta_{17} = \theta_{18} &= (\alpha_4, [P : \downarrow(1, 1)]); \end{aligned}$$

The labels γ_t related to the transition system contain the kind of action α_j and the rate, calculated by applying the associated function f_{α_j} to the quantitative information collected in the previous labels θ_t . The values obtained are rates associated with the CTMC derived from the labelled transition system. In the case of two levels, these are:

$$\begin{aligned} q_{1,2} = q_{4,3} = \frac{v^*(1)}{K_M} & & q_{8,7} = q_{5,6} = \frac{v^*(1)}{K_M + M_{P2}} & & q_{2,3} = q_{6,7} = k_2 * M_M & & q_{3,2} = q_{7,6} = k_4 * M_P \\ q_{2,1} = q_{3,4} = q_{7,8} = q_{6,5} = k_3 * M_M & & q_{4,5} = q_{3,6} = k_5 * (M_P)^2 & & q_{5,4} = q_{6,3} = K_5i * M_{P2} \end{aligned}$$

It is worth noting that in this case the suffix h, k of q indicates the initial state (h) and the final one (k) of the associated transition.

7 Conclusions

In this paper we have presented Bio-PEPA, 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. Stoichiometric coefficients are added to the syntax as additional information. They are used in the calculation of the rate and are necessary for the further analysis of the system. However, stoichiometry does not influence the number of levels. Indeed stoichiometry is at a different level of abstraction with respect to the concentration levels.

In Bio-PEPA not only elementary reactions with constant rates, but also complex reactions described by general kinetic laws can be considered. Indeed each reaction in the model is associated with an action type and with a functional rate. Functional rates are defined externally to the components and are evaluated in the system at the moment of the derivation. The possibility to consider various kinds of kinetic laws permits us to model a vast number of biochemical networks. Indeed complex reactions are frequently found in models as abstractions of sequences of elementary steps and reducing to elementary reactions is often impossible and undesirable.

In this work we have shown how to derive a CTMC from the transition system of a Bio-PEPA model and we discussed the derivation of ODEs and the use of Gillespie. The major problem is the application of Gillespie with general kinetic laws. The approach proposed in this work is to use Gillespie also in this context,

but to pay attention to the interpretation of the results. One topic for the future concerns a deeper investigation into the relation between Gillespie simulation and general kinetic laws. Another area for future work will concern the application of model checking techniques for the analysis of the properties of a Bio-PEPA system.

Acknowledgement

The authors are supported by the EPSRC under the CODA project “Process Algebra Approaches for Collective Dynamics” (EP/c54370x/01 and ARF EP/c543696/01). They also thank the anonymous referees for the useful comments.

References

- [1] Arkin, A.P. and C.V. Rao, *Stochastic chemical kinetics and the quasi-steady-state assumption: application to the Gillespie algorithm*, *Journal of Chemical Physics* (2003), **11**, 4999–5010.
- [2] Bornstein, B.J., J.C. Doyle, A. Finney, A. Funahashi, M. Hucka, S.M. Keating, H. Kitano, B.L. Kovitz, J. Matthews, B.E. Shapiro and M.J. Schilstra, *Evolving a Lingua Franca and Associated Software Infrastructure for Computational Systems Biology: The Systems Biology Markup Language (SBML) Project*, *Systems Biology*, **1** (2004), 41–53.
- [3] Bortolussi, L. and A. Policriti, *Modeling Biological Systems in Stochastic Concurrent Constraint Programming*, Proc. of WCB 2006, 2006.
- [4] Bundschuh, R., F. Hayot and C. Jayaprakash, *Fluctuations and Slow Variables in Genetic Networks*, *Biophys. J.*, **84**, 1606–1615, 2003.
- [5] Calder, M., S. Gilmore and J. Hillston, *Automatically deriving ODEs from process algebra models of signalling pathways*, Proc. of CMSB’05, 2005.
- [6] 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.
- [7] 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.
- [8] Cao, Y. and D.T. Gillespie and L. Petzold, *Accelerated Stochastic Simulation of the Stiff Enzyme-Substrate Reaction*, *J. Chem. Phys.*, **123** (2005), 144917–144929.
- [9] Cardelli, L., E. M. Panina, A. Regev, E. Shapiro and W. Silverman, *BioAmbients: An Abstraction for Biological Compartments*, *Theoretical Computer Science*, **325(1)** (2004), 141–167, Elsevier.
- [10] Chabrier-Rivier, N., F. Fages and S. Soliman, *Modelling and querying interaction networks in the biochemical abstract machine BIOCHAM*, *Journal of Biological Physics and Chemistry* **4** (2004), 64–73.
- [11] Costantin, G, C. Laudanna, P. Lecca, C. Priami, P. Quaglia and B. Rossi, *Language modeling and simulation of autoreactive lymphocytes recruitment in inflamed brain vessels*, *SIMULATION: Transactions of The Society for Modeling and Simulation International*, **80** (2003), 273–288.
- [12] Danos, V. and J. Krivine, *Formal molecular biology done in CCS-R*, Proc. of Workshop on Concurrent Models in Molecular Biology (BioConcur’03), 2003.
- [13] Eichler-Jonsson, C., E.D. Gilles, G. Muller and B. Schoeberl, *Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors*, *Nature Biotechnology*, **20** (2002), 370–375.
- [14] Gillespie, D.T. *Exact stochastic simulation of coupled chemical reactions*, *Journal of Physical Chemistry* **81** (1977), 2340–2361.
- [15] Hillston, J., “A Compositional Approach to Performance Modelling”, Cambridge University Press, 1996.
- [16] Kierzek, A.M. and J. Puchalka, *Bridging the gap between stochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks*, *BIOPHYS J.*, **86** (2004), 1357–1372.

- [17] Kuttler, C. and J. Niehren, *Gene regulation in the π -calculus: simulating cooperativity at the lambda switch*, Transactions on Computational Systems Biology VII, **4230** (2006), 24–55.
- [18] Priami, C. and P. Quaglia, *Beta-binders for biological interactions*, Proc. of CMSB'04, 2004.
- [19] Priami, C., 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** (2001), 25–31.
- [20] Segel, I.H. “Enzyme Kinetics: Behaviour and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems”, Wiley-Interscience, New-York, 1993.