Some investigations concerning the CTMC and the ODE model derived from Bio-PEPA

Federica Ciocchetta [1,a], Andrea Degasperi [2,b], Jane Hillston [3,a] and Muffy Calder [4,b]

[a] *Laboratory for Foundations of Computer Science, The University of Edinburgh, Scotland*

[b] *Department of Computer science, University of Glasgow, Scotland*

Abstract

Bio-PEPA is a recently defined language for the modelling and analysis of biochemical networks. It supports an abstract style of modelling, in which *discrete levels of concentration* within a species are considered instead of individual molecules. A finer granularity for the system corresponds to a smaller concentration step size and therefore to a greater number of concentration levels. This style of model is amenable to a variety of different analysis techniques, including numerical analysis based on a CMTC with states reflecting the levels of concentration.
In this paper we present a formal definition of the CTMC with levels derived from a Bio-PEPA system. Furthermore we investigate the relationship between this CTMC and system of ordinary differential equations (ODEs) derived from the same model. Using Kurtz's theorem, we show that the set of ODEs derived from the Bio-PEPA model is able to capture the limiting behaviour of the CTMC obtained from the same system. Finally, we define an empirical methodology to find the granularity of the Bio-PEPA system for which the ODE and the CTMC with levels are in a good agreement. The proposed definition is based on a notion of distance between the two models. We demonstrate our approach on a model of the Repressilator, a simple biochemical network with oscillating behaviour.

Keywords: Systems Biology, process algebras, analysis, differential equations, Markov chains

1 Introduction

In the recent years there have been various applications of process algebras for the study of biochemical networks [16,15,2,6,1]. An attractive feature of process algebras is the simple abstraction they offer for representing biological entities. In the π-calculus and related calculi [16,15] each biochemical molecule is abstracted by a process and reactions are represented by means of communications between processes. In the recently defined Bio-PEPA formalism [4,5] a different view has been proposed: each component abstracts the behaviour of a species instead of a

[1] fciocche@inf.ed.ac.uk
[2] andrea@dcs.gla.ac.uk
[3] jeh@inf.ed.ac.uk
[4] muffy@dcs.gla.ac.uk

This paper is electronically published in
Electronic Notes in Theoretical Computer Science
URL: `www.elsevier.nl/locate/entcs`

single molecule. In particular, species concentrations are discretized into levels and the components capture concentration levels within a species. The *granularity* of the system is expressed in terms of the concentration step size h: when h decreases, the granularity increases. This modelling style, based on discrete concentration levels, gives rise to an underlying continuous time Markov chain (CTMC) which we will call the "*CTMC with levels*" to distinguish it from the CTMC which underlies a stochastic simulation based on individual molecules.

Since Bio-PEPA is an *intermediate, formal, compositional* representation of the biological model it supports different kinds of analysis, including stochastic simulation [9], analysis based on ordinary differential equations (ODEs), numerical solution of the CTMC with levels and stochastic model checking using PRISM [14]. It is worth noting that each of these analyses can aid understanding of different aspects of the behaviour of the system. Furthermore, when two analyses overlap in scope, the results obtained can used for verification.

This paper makes the following contributions:

- provide a formal definition of the CTMC with levels;
- present an investigation of the relationship between CTMC with levels and ODEs obtained from the same Bio-PEPA system;
- propose a methodology to find the granularity h for which these two underlying models are in good agreement.

The CTMC with levels was introduced in [3] and subsequently also used in the PEPA reagent-centric view [2]. One advantage of this approach is that it is *semi-quantitative*, allowing us to deal with incomplete information about molecular concentrations, as given in real experimental settings. Furthermore, in comparison to the CTMC underlying a stochastic simulation, it leads to a reduction of the state space, leading to models which may be amenable to numerical solution and approaches such as stochastic model checking. The authors of [3] focused on the case of reactions with mass-action kinetics and stoichiometry equal to one for all the reagents. Here we extend this approach to the general case and we investigate some properties of these Markov chains.

Such an approach does represent some loss of information compared to both the stochastic simulation (in which all molecules are represented individually) and ODE model (in which concentrations vary continuously rather than in discrete jumps). The second aspect of our work concerns an investigation of the relationship between the CTMC with levels and the set of ODEs obtained from the same Bio-PEPA system. Confidence in the compatibility between the two models is important since we can use them to perform different kinds of analysis. For instance we can check some properties of the system by using model checking before simulating the model by using ODEs. The validity of the results depends on the agreement between the two approaches. The relationship between ODEs and CTMC derived from a process algebra model has been previously investigated in [8], but in that case the authors focused on the pathway centric-view in PEPA. Here we adapt their approach to the reagent-centric style modelling supported by Bio-PEPA. Using Kurtz's theorem [11] we show that the set of ODEs derived from Bio-PEPA is able to capture the limiting behaviour of the CTMC with levels representing the discretised system.

This involves showing that the CTMC belongs to the family of *density dependent CTMCs*, i.e. the rates of the CTMC may depend on a scaled representation of states, in our case the step size of the species concentrations.

The last challenge is to determine a value for the step size h which gives good agreement between the two models. In other words, for a fixed error ϵ, we want to find a value h for which the two models differ by less that ϵ. This leads us to consider how to express the difference between models. A relation can be found from the probabilistic approach [12,13], but is too complex for practical use. We propose another approach based on the definition of a distance function between the models.

The rest of the paper is structured as follows. Section 2 reports a brief introduction to Bio-PEPA, a description of the definition of discrete levels and transition rates in Bio-PEPA. In Section 3 the CTMC with levels is defined. The maps from Bio-PEPA to ODEs and the CTMC with levels are described in Section 4 and the relationship between these models is discussed in Section 5. Firstly, Kurtz's theorem is applied to show the convergence of the two models in the limit, i.e. when the concentration step in the CTMC tends to zero. Secondly, we define the distance between the two models for a given granularity, in order to express a measure for the agreement between them. In Section 6 the repressilator model, a genetic network with oscillating behaviour, is considered to illustrate and test our approach. Finally, in Section 7, some conclusions and directions for future work are reported.

2 Bio-PEPA

In this section we present a short description of Bio-PEPA [4,5] and then we discuss the definition of discrete levels of concentration and how to derive the transition rates from the reaction kinetic laws. Some auxiliary definitions for Bio-PEPA are reported in the Appendix A.1.

The context of application is biochemical networks. A biochemical network is composed of n species that interact through m reactions in o compartments. The dynamics of reaction j is described by a kinetic law f_j. The stoichiometric coefficients of the reactions are assumed to be integer and bounded.

We make the following assumptions:

- only irreversible reactions are considered: reversible reactions can be seen as the union of a pair of forward and inverse reactions;
- the reactants of the reaction can only decrease their concentration whereas the products can only increase it. Enzymes and inhibitors do not change;
- the same species in different situations (e.g. phosphorylated, free, bound...) are regarded as different species and represented by distinct Bio-PEPA components;
- compartments are static and do not play an active role in reactions. Throughout this paper we assume that all reactions take place within a single compartment.

2.1 The language

The syntax of Bio-PEPA is defined as:

$$S ::= (\alpha, \kappa)\ \mathtt{op}\ S \mid S{+}S \mid C \qquad P ::= P \underset{\mathcal{L}}{\bowtie} P \mid S(l) \quad \text{where } \mathtt{op} = \downarrow \mid \uparrow \mid \oplus \mid \ominus \mid \odot\,.$$

The component S is called a *sequential component* (or *species component*) and represents the species whereas the component P, called a *model component*, describes the system and the interactions among components. The parameter $l \in \mathbb{N}$ represents the discrete level of concentration. The prefix term $(\alpha, \kappa)\ \mathtt{op}\ S$ contains information about the role of the species in the reaction associated with the action type α: κ is the *stoichiometry coefficient* of the species and the *prefix combinator* "op" represents the role of the element in the reaction. Specifically, $\downarrow$ indicates a *reactant*, $\uparrow$ a *product*, $\oplus$ an *activator*, $\ominus$ an *inhibitor* and $\odot$ a generic *modifier*. The operator "+" expresses the choice between possible actions and the constant C is defined by an equation $C \stackrel{def}{=} S$. Finally, the process $P \underset{\mathcal{L}}{\bowtie} Q$ denotes the cooperation between components: the set $\mathcal{L}$ determines those activities on which the operands are forced to synchronize. We can define a Bio-PEPA system as follows:

Definition 2.1 A Bio-PEPA system $\mathcal{P}$ is a 6-tuple $\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}_R, Comp, P \rangle$, where: $\mathcal{V}$ is the set of compartments, $\mathcal{N}$ is the set of quantities describing each species, $\mathcal{K}$ is the set of parameter definitions, $\mathcal{F}_R$ is the set of functional rate definitions, $Comp$ is the set of definitions of sequential components, P is the model component describing the system.

Each element in the set $\mathcal{N}$ associates a species name with the step size, the initial concentration, the number of levels and the compartment where the species is. We denote the set of well-defined Bio-PEPA systems $\tilde{\mathcal{P}}$ (see [5] for more details).

The behaviour of the system is defined in terms of an operational semantics. The rules are reported in [5]. The derivation of the rate is reported in Section 2.4.

2.2 Dimerization example

Let us consider the system composed of the following two reactions, representing the dimerization of a protein and its inverse process:

$$2A \xrightarrow{f_{MA}(k_1)} B \qquad B \xrightarrow{f_{MA}(k_2)} 2A \tag{1}$$

The dynamics is mass-action kinetics($f_{MA}(k_1) = k_1 \cdot A^2$ and $f_{MA}(k_2) = k_2 \cdot B$, respectively). We assume that initially $A = 10\ mol/l$ and $B = 0\ mol/l$ and $k_1 = k_2 = 1.0$. In the following we show how to represent this system in Bio-PEPA. This simple network will be used as a running example throughout the remainder.

We define for each species the step size (h), the number of levels (N), the initial concentration (M_0) and the compartment containing the species (V):

$$A : h = 5,\ N = 2,\ M_0 = 10,\ V; \qquad B : h = 5,\ N = 1,\ M_0 = 0,\ V;$$

Note that the stoichiometry of A in the reactions is two and so we need at least two levels for A. This corresponds to the maximum granularity possible.

We define the functional rates: $f_{\alpha_1} = f_{MA}(k_1)$ and $f_{\alpha_2} = f_{MA}(k_2)$.
At this point we can define the set of species components:

$$A \stackrel{def}{=} (\alpha_1, 2)\downarrow A \; + \; (\alpha_2, 2)\uparrow A \qquad B \stackrel{def}{=} (\alpha_1, 1)\uparrow B \; + \; (\alpha_2, 1)\downarrow B;$$

The model component is: $A(2) \underset{\{\alpha_1,\alpha_2\}}{\bowtie} B(0)$.

2.3 Discrete levels of concentration

Each species is characterized by a number of levels, equidistant from each other, with distance equal to h. Specifically, we assume that all the species in the same compartment have the same step size [5]. This follows from the *law of conservation of mass*: there must be a "balance" between the number of molecules consumed (reactants) and the ones created (products). Note that a finer granularity of a Bio-PEPA system corresponds to a smaller step size.

We assign to each species different concentration levels, from 0 (corresponding to null concentration) to a maximum number N. This ensures that the underlying CTMC with levels has a finite state space — a condition which is necessary to make numerical analysis feasible. The maximum level N_i for each species i is defined according to prior knowledge and experimental evidence.

If l_i is the level for the species i, the concentration is taken to be $x_i = l_i \cdot h$. The initial concentration and the initial level of i are $x_{i,0}$ and $l_{i,0}$, respectively.

2.4 Derivation of rates

In the following we show how to derive the transition rates depending on stoichiometry when discrete concentrations are used. The transition rate is defined by $(\Delta t)^{-1}$, where Δt is the time taken to vary the concentration of reactants/products between states in the CTMC.

Stoichiometry equal to one

Let f_j be the kinetic law and let y be one product of the reaction j. The rate equation for that species with respect to the given reaction is $y' = f_j(t, \bar{x})$, where $\bar{x}$ is the set (or a subset) of the reactants/modifiers of the reaction. Applying the Taylor expansion (to two terms) we obtain:

$$y_{n+1} \approx y_n + f_j(t_n, \bar{x}_n) \cdot (t_{n+1} - t_n)$$

We define $y_{n+1} - y_n = 1 \cdot h$ and then derive the respective time interval $(t_{n+1} - t_n) = \Delta t$ as $\Delta t = \frac{h}{f_j(t_n, \bar{x}_n)}$. From this we obtain the transition rate $\frac{f_j(t_n, \bar{x}_n)}{h}$. Note that if stoichiometry is equal to one we have a variation of only one level from one state to the other.

[5] Elements that are only modifiers can have a different step size, as their concentration is not affected by the reaction.

Stoichiometry possibly different from one

We assume the kinetic law is mass-action. Let y be a product of the reaction and let κ be its stoichiometric coefficient with respect to that reaction. Applying the expansion again we obtain:

$$y_{n+1} \approx y_n + \kappa \cdot r \cdot \prod_{i=1}^{n_r} x_{i,n}^{\kappa_i} \cdot (t_{n+1} - t_n)$$

where r is the rate constant, x_i, with $i = 1, ..., n_r$ are the reactants of the reaction, κ_i are the associated stoichiometric coefficients and n_r is the number of distinct reactants in the reaction.

Now we can fix $y_{n+1} - y_n = \kappa \cdot h$. From this we can derive the rate as usual.

Summary

The rate associated with a transition from one state u to another state v can be calculated as: $r_j = \frac{f_j[u]}{h}$, where h is the step size of the reactants and $f_j[u]$ is the evaluation of the functional rate in the state u. When the stoichiometric coefficient of a reagent is κ then the reagent varies by κ levels as a result of the transition.

The kinetic laws for the reactions have are required to satisfy some properties.

Property 1 *For each reaction j the kinetic law function f_j has to satisfy the following properties:*

- *it is continuously differentiable;*
- *it is monotonically decreasing in terms of the reactant variables.*

We impose the first property, which is useful to prove some results about the CTMC, whereas the second property follows because we assume that the reactants can only decrease their concentration and products can only increase it. All the most well-known kinetic laws satisfy these properties.

3 CTMC with levels

The term *CTMC with levels* indicates a CTMC whose states capture levels of concentration of the species and the transitions from one state to another reflect some variations of these levels.

Definition 3.1 The states of a CTMC with levels are defined as vectors of levels $\sigma = (l_1, l_2, ..., l_n)$, where l_i , for $i = 1, 2, ..., n$, is the level of the species i.

The elements in a state represent the concentration levels of the different species.

Definition 3.2 The transitions of a CTMC with levels represent biochemical reactions. Each transition causes a change in the number of levels of one or more speci es and the variation in the number of levels depends on the stoichiometry. The transition rates are as defined in Section 2.3.

For the analysis, it is necessary to assume that the CTMCs are finite. Starting from a finite number of levels, is possible to obtain an infinite CTMC only if there

are some reactions of the kind "$\rightarrow A$" or "$C \rightarrow C + A$". We call these *creation reactions*. We term a biochemical network without creation reactions a *bounded chemical network*. We have the following result.

Proposition 3.3 *Let X_h be a CTMC corresponding to a bounded biochemical network with granularity h. Let $\sigma_0 = (l_{1,0}, l_{2,0}, ..., l_{n,0})$ be the vector describing the initial state. If the values $l_{i,0}$, for $i = 1, 2, ..., n$ are finite then X_h is finite and the maximum value of the level depends on the initial state σ_0 and the stoichiometric coefficients of the reactions. In particular, if all stoichiometric coefficients are equal to one, in each state σ each component l_i satisfies: $l_i \leq (\sum_{j=1}^{n} l_{j,0})$.*

We note that if we allow creation reactions, then in order to guarantee that the CTMC is finite we have to assume that there is a maximum concentration for each species i. Moreover, if there is a species without a limiting value, we consider a maximum level for the values greater than a certain (high) value. Note that this is an approximation and we have to pay attention to the results obtained from the analysis (e.g. model checking).

We now turn our attention to deriving CTMC with levels and ODEs.

4 From Bio-PEPA to CTMC with levels and to ODEs

In this section we outline how the CTMC with levels and the system of ODEs underlying a Bio-PEPA model are derived.

4.1 From Bio-PEPA to CTMC with levels (π_{CTMC})

Let π_{CTMC} be the function that derives a CTMC with levels from a Bio-PEPA system. We do not define this function formally, but states are derived in the obvious way from the model component, via the operational semantics, and transition rates are as described in Section 2.4. From any Bio-PEPA system we can apply the semantic rules to find the *derivatives* of the system, and regard the label transition system which results from the exhaustive application of the semantic rules as the *derivation graph* of the model.

We have the following result:

Theorem 4.1 *For any finite Bio-PEPA system $\mathcal{P} = \langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}_R, Comp, P \rangle$, if we define the stochastic process $X_h(t)$ such that $X_h(t) = P_i$ indicates that the system behaves as derivative P_i at time t, then $X_h(t)$ is a CTMC.*

The proof is not reproduced here but it is analogous the one presented for PEPA [10]. The rate associated with each activity is the rate obtained by evaluating the functional rate in the system.

The CTMC is characterised by a infinitesimal generator matrix $\mathbf{Q}$ whose off-diagonal entries are derived from the transition rates in the obvious way, and whose diagonal entries are the negative row sums.

For well-defined Bio-PEPA systems we assume that each action is associated with a reaction. From this observation and from the assumption that each action type represents a distinct reaction, we have the following fact: if two transitions

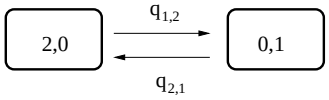

Fig. 1. CTMC with levels for the dimerization example ($h = 5$).

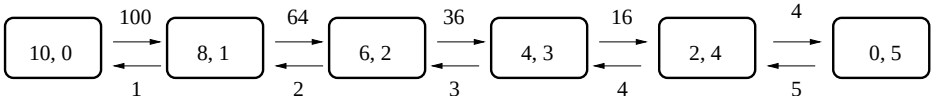

Fig. 2. CTMC with levels for the dimerization example ($h = 1$).

are possible between a pair of states, the actions involved are different and they represent reactions that differ only in the modifiers and/or the number of enzymes used. This is formalized in the following Proposition.

Proposition 4.2 *Let $\mathcal{P}$ be a well-defined Bio-PEPA system with model component P. Let P_u and P_v be two derivatives of P such that the latter is one-step derivative of the former. If there exist two action types α_1 and α_2 such that $P_u \xrightarrow{\alpha_1} P_v$ and $P_u \xrightarrow{\alpha_2} P_v$ then:*

(i) $\alpha_1 \neq \alpha_2$;

(ii) *the two action types refer to two transitions/biological reactions that differ only in the modifiers.*

The first point follows from the definition of well-defined Bio-PEPA systems. The second follows because the only possibility to have two transitions between two given states is that the associated reactions have the same reactants and products. We can see this by observing that the states depend on the levels and the reactions cause some changes in these levels. The only elements that do not change during a reaction are the modifiers.

4.1.1 Dimerization example (continued)

Consider the dimerization example; we derive the CTMC with levels for two values of h: $h = 5$ and $h = 1$, as illustrated in Figures 1 and 2.

When $h = 5$ there are 2 states, $(2, 0)$ and $(0, 1)$ and 2 transitions with transition rates:

$$q_{1,2} = k_1 \cdot A^2/h = 1 * (10)^2/5 = 20, \quad q_{2,1} = k_2 \cdot B/h = 1 * 5/5 = 1,$$

When h is smaller ($h = 1$) then there is a finer granularity. There are 6 states and 10 transitions (see Figure 2). As a further example (not illustrated), when $h = 0.1$ there are 51 states and 100 transtions.

4.2 From Bio-PEPA to ODEs

Let π_{ODE} be the definition of the set of ODEs from a Bio-PEPA model. A crucial part is the derivation of the stoichiometry matrix $D = \{d_{ij}\}$. The entries of the matrix are obtained in the following way: for each sequential component C_i consider the prefix subterms C_{ij} representing the contribution of the species i to the reaction

j. If the term represents a reactant we write the corresponding stoichiometry κ_{ij} as $-\kappa_{ij}$ in the entry d_{ij}. In the case of a product we write $+\kappa_{ij}$. All other cases are null.

π_{ODE} entails three steps: 1) definition of the stoichiometry $(n \times m)$ matrix D, where n is the number of species and m is the number of molecules; 2) definition of the *kinetic law vector* $(m \times 1)$ $\mathbf{v_{KL}}$ containing the kinetic laws of each reaction; 3) definition of the vector $(n \times 1)$ $\mathbf{x}$, with $\mathbf{x}^T = (x_1, x_2, ..., x_n)$.

The ODE system thus obtained has the form:

$$\frac{d\mathbf{x}}{dt} = D \times \mathbf{v_{KL}}$$

where the vector of initial concentrations is $\mathbf{x_0}$, with $x_{i,0} = l_{i,0} \cdot h$, $i = 1, ..., n$.

4.3 Dimerization example (continued 2)

We define the vector $\mathbf{x}^T = (x_A, x_B)$ and the kinetic vector $\mathbf{v_{KL}}^T = (k_1 \cdot x_A^2, k_2 \cdot x_B)$. The stoichiometry matrix D associated with the system is

$$\begin{pmatrix} -2 & +2 \\ +1 & -1 \end{pmatrix}$$

The system of ODEs obtained by π_{ODE} is:

$$\frac{dx_A}{dt} = -2 \cdot k_1 \cdot x_A + 2 \cdot k_2 \cdot x_B$$
$$\frac{dx_B}{dt} = +k_1 \cdot x_A - k_2 \cdot x_B$$

with initial conditions $(x_{A,0}, x_{B,0})^T = (10, 0)$.

5 Comparison of CTMC with levels and ODEs

In this section we consider how to compare the two models derived from a Bio-PEPA system $\mathcal{P}$ and how to define h such that the difference between the two models is acceptable.

First we apply Kurtz's Theorem [11,13] to our case. This Theorem tells us that, unders some conditions, the ODE system is the limit of a sequence of density dependent CTMCs (the CTMCs with levels), as h approaches 0.

Second we consider how to define the difference between the two models. We define this as a distance measure and then we discuss the factors to consider when choosing h in order to ensure that the distance between the two models is less than a acceptable error measure ϵ.

In the following we introduce the notation used, then we show that the CTMC with levels derived from a Bio-PEPA system satisfies the conditions of Kurtz's Theorem. Finally, we observe that the set of ODEs extracted from the Bio-PEPA system coincides with those in the theorem.

5.1 Application of Kurtz's Theorem

Kurtz's Theorem applies to a sequence of *density dependent Markov chains*. In the original theorem the dependency is expressed in terms of the volume V, but we express the dependency in terms of the granularity h. Note that when h decreases, the granularity of the system increases. The formal definition of the Theorem with its conditions is reported in the Appendix A.2.

5.1.1 Notation

We consider Bio-PEPA systems representing bounded biochemical networks.

Let X_h be the CTMC describing the model with granularity h. Given a state of the CTMC σ, we denote by $\mathbf{h}\sigma$ the vector $(h \cdot l_1, h \cdot l_2, ..., h \cdot l_n)$, where h is the step size and l_i is the level of the species i, respectively.

Let D be the stoichiometry matrix obtained from the Bio-PEPA system and D^j the jth column of D. This vector represents the stoichiometric coefficients for all the species in a given reaction j.

5.1.2 Details of the proof

In order to apply Kurtz's Theorem, we have to see if the CTMC X_h is density dependent and if all the conditions of the theorem are satisfied (see A.2 for these conditions).

X_h is density dependent From the definition of CTMC with levels, we have that the entry $q_{u,v}$ of the infinitesimal generator matrix is

$$q_{u,v} = \sum_{\mathcal{A}(P_u|P_v)} f_j[u] \cdot h^{-1} \quad \text{if } u \neq v \qquad q_{u,u} = -\sum_{u \neq v} q_{u,v} \quad \text{otherwise.}$$

where $\mathcal{A}(P_u|P_v) = \{\alpha \mid P_u \xrightarrow{\alpha} P_v\}$ and $f_j[u]$ is the evaluation of the functional rate in the starting state P_u.

Using the notation above, we can write the rate as $f_j(\mathbf{h}\sigma, d^j) \cdot h^{-1}$, where d^j gives the information about the stoichiometry of the reagents of the reaction.

Conditions of Kurtz's theorem Let $\mathbf{x_0}$ be the initial concentration vector. We have that the initial level vector is by definition $\mathbf{l_0} = \lceil \mathbf{x_0}/h \rceil$. Therefore $\lim_{h \to 0} h * \mathbf{l_0} = \mathbf{x_0}$.

Consider the system of ODEs $\dot{X}(t) = F(X)$ where $F(\mathbf{x}) = \sum d^j f_j(\mathbf{x}, d^j)$ with initial condition $X(0) = \mathbf{x_0}$. By hypothesis, the trajectory of $X(t)$ is bounded, so we can include it in an open bounded set E. From the fact that each kinetic law is continuously differentiable (the first property of kinetic laws in 2.3), it follows that f is Lipschitz. This the first condition of Kurtz's Theorem. The second and third conditions of Kurtz's Theorem state that for each transition the rate of change is bounded and that there is a bound for the whole state space so that the impact of each transition is bounded. By the assumptions made for the kinetic laws and from the fact that the stoichiometry matrix contains values that are finite [6] it is clear that both these conditions are also satisfied. In particular, for

[6] Stoichiometric coefficients are assumed to be integer and bounded.

the third condition, we can observe that $f(\mathbf{x}, d^j)$ is equal to zero for all $|d| > C$ with $C = \sum_{i,j} d_{ij}$.

ODE systems Consider the ODE system $\pi_{ODE}(\mathcal{P})$, for a given Bio-PEPA system $\mathcal{P}$. We can observe that $F(\mathbf{x}) = D \times \mathbf{v_{KL}}$, as the kinetic law vector $\mathbf{v_{KL}}$ contains all the functions f_j for all the reactions. The ODE system $\pi_{ODE}(\mathcal{P})$ coincides with the one in Kurtz's Theorem, with initial condition $x_{i,0} = l_{i,0} \cdot h$, for $i = 1, 2, ..., n$.

5.2 Distances between the two models

The result in Section 5 confirm that, in the limit, the agreement between the ODEs and the CTMC with levels derived from a Bio-PEPA system is complete, but does not say anything about the relation between the two approaches for a given finite h. In [3] the authors showed experimentally that in some pathways the two models are indistinguishable for just few levels, for example when $h = 1$ and $h = 7$ (corresponding to h relatively high), but these results are not generalised. Here we investigate the relationship between the step size h of the CTMC and the agreement with ODEs.

In [12,13] Kurtz reported some estimates for the probability of convergence between the two models. However the estimation is complex and offers a poor guide to choosing V (or h, in our case).

In the following we propose a notion of distance between the two models. There are various possible ways to define this distance. One possibility is to define it in terms of the difference between the ODE simulation trajectory and the expected value (numerical solution) of the CTMC, for all the species in the biological network, with respect to a time interval. This gives the following definition of distance:

$$f_{dist} = \sum_{t=1}^{T_{sim}} \sum_{i=1}^{n} \left(X_i^h(t) \cdot h - x_i(t) \right)^2$$

where x_i is the ODE trajectory for the species i, $X_i^h(t)$ is the numerical solution of the CTMC for the species i at time t, n is the number of species in the network, T_{sim} is the simulation time and t indicates a simulation time point. We define a normalized distance by dividing each summand in the expression above by $x_i^2(t)$. In this last case we call the distance function, f_{dist}, normalized.

We propose the following empirical approach to find the value of h for which we have good agreement between the models.

- Let us consider a well-defined Bio-PEPA system $\mathcal{P}$, the CTMC $X_h = \pi_{CTMC}(\mathcal{P})$ and the ODEs solution X of the model $\pi_{ODE}(\mathcal{P})$.
- Let T_{sim} be the time of simulation (this depends on the model) and $\epsilon > 0$ the discrepancy admissible between the two models.
- Starting from an initial granularity we calculate the distance value in the simulation time. If the value is greater than ϵ then consider a smaller h and try again.

Some observations are due. A key point is the choice of the value ϵ. Furthermore, the calculation of the numerical solution for the CTMC is often impracticable for

non elementary models. Instead of considering the expected value of the CTMC, we can define the distance between the two models in terms of the average (mean) of some CTMC simulation runs. We have the following distance function:

$$f_{dist,avg} = \sum_{t=1}^{T_{sim}} \sum_{i=1}^{n} \left(\bar{X}_i^h(t) \cdot h - x_i(t) \right)^2$$

where $\bar{X}_i^h(t)$ is the average value for the CTMC over N_{run} runs for the species i at time t and the other variables are as before. In this approach the main challenge is the definition of the number of simulation runs needed to have a good approximation of the expected value. Increasing the number of simulation runs we obtain a better approximation of the expected value for the CTMC, however the calculations become more expensive. Generally we can obtain indistinguishable curves for a relative small number of runs.

In both the definitions, the distance between the two models decreases with the step size h. However, note that for very small h the number of states becomes large and furthermore the analysis of the CTMC may become prohibitively expensive. So there is a trade-off between accuracy (in terms of both number of runs and step size) and tractability. The resolution of this trade-off is left to the modeller.

5.3 Dimerization example (continued 3)

In Figure 3 we report some analysis results for the dimerization example. The ODE simulation is reported at the top. The other two graphs show the time evolution of the expected value of the CTMC with levels for $h = 5$, $h = 1$, $h = 0.1$ and $h = 0.01$, both for A and B.

By comparing the ODE trajectory and the numerical solutions, we can observe that for a large step sizes ($h = 5$) there is a discrepancy between the two curves, both for A and B. When we decrease the step size h, the discrepancy between the two curves becomes smaller and for $h = 0.1$ (corresponding to 100 levels) the expected value of the CTMC (almost) coincides with the ODE. This is as predicted by Kurtz's Theorem.

If we consider the average of some simulation runs instead of the expected value, we obtain similar results for 100 runs. However, in the case of $h = 5$ and $h = 1$ there is a large variability between the different simulation runs.

In the table below we report the distance f_{dist} between the ODE and CTMC for different values of h. If we fix the admissible distance between the models as $\epsilon = 1.05$, then we have $h = 0.1$. The choice of ϵ is supported by the graphical results reported above.

distance	h = 5	h = 1	h = 0.1	h = 0.01
f_{dist}	53.69	3.7	1.02	1.00

6 The repressilator

The *repressilator* is a synthetic genetic regulatory network with oscillating behaviour [7]. The repressilator consists of three genes (denoted *G1, G2, G3*) connected in a

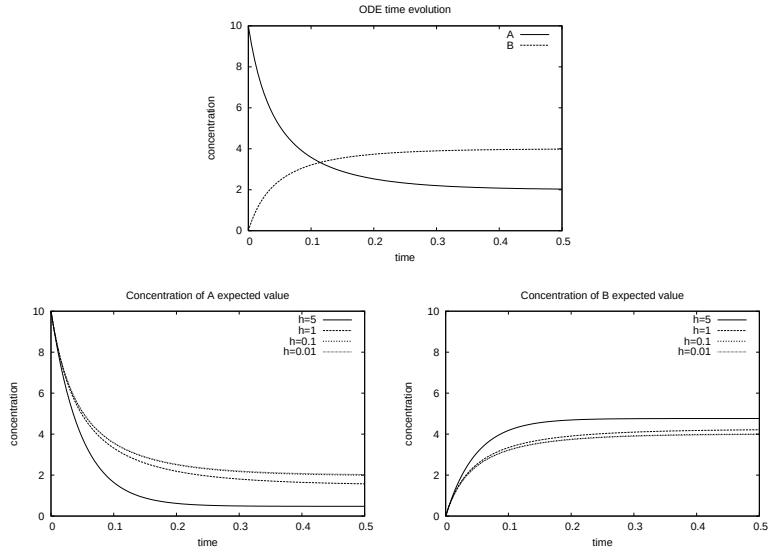

Fig. 3. Dimerization example. On the top: ODE simulation. The other two graphs represent the numerical solution for the CTMC for A (on the left) and B (on the right) for $h = 5$, $h = 1$, $h = 0.1$ and $h = 0.01$.

feedback loop, such that the transcription of a gene is inhibited by one of the other proteins (denoted *P1, P2, P3*). A schema of the network is reported in Figure 4.

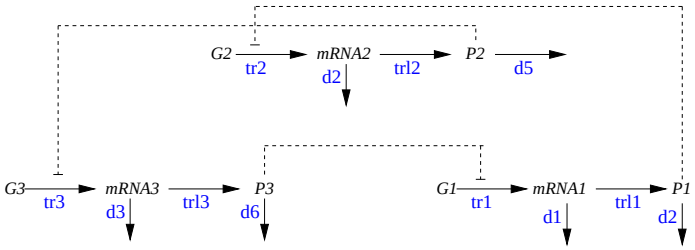

Fig. 4. Repressilator model.

The reactions are: the *transcription* of the three mRNAs with inhibition by one of the proteins (reactions $tr1$, $tr2$, $tr3$), the *translation* of mRNAs into the proteins (reactions $trl1$, $trl2$, $trl3$), *degradation* of both mRNAs and proteins (reactions d_i with $i = 1, ..., 6$).

The Bio-PEPA system corresponding to this network is detailed in the Appendix A.3. From the Bio-PEPA system we can derive the CTMC and the ODE model as usual (reported in the Appendix A.3). For each temporal point we show the mean and the standard deviation of the 100 runs. In Figure 5 we report some analysis results. The ODE simulation is reported at the top, left. The other graphs show the time evolution of the average of 100 simulation runs for the CTMCs with $h = 5$, $h = 0.1$ and $h = 0.01$. For the Repressilator the numerical calculation of the expected value of the CTMC is too expensive.

For h high, there is a great variability among the different simulation runs and the mean value is very different from the ODE results. For smaller h this variability decreases and the mean value approaches the ODE trajectory.

In the Table below we report the distances between ODE and CTMC for different values of h. We consider the definition of distance in terms of the mean value. Note

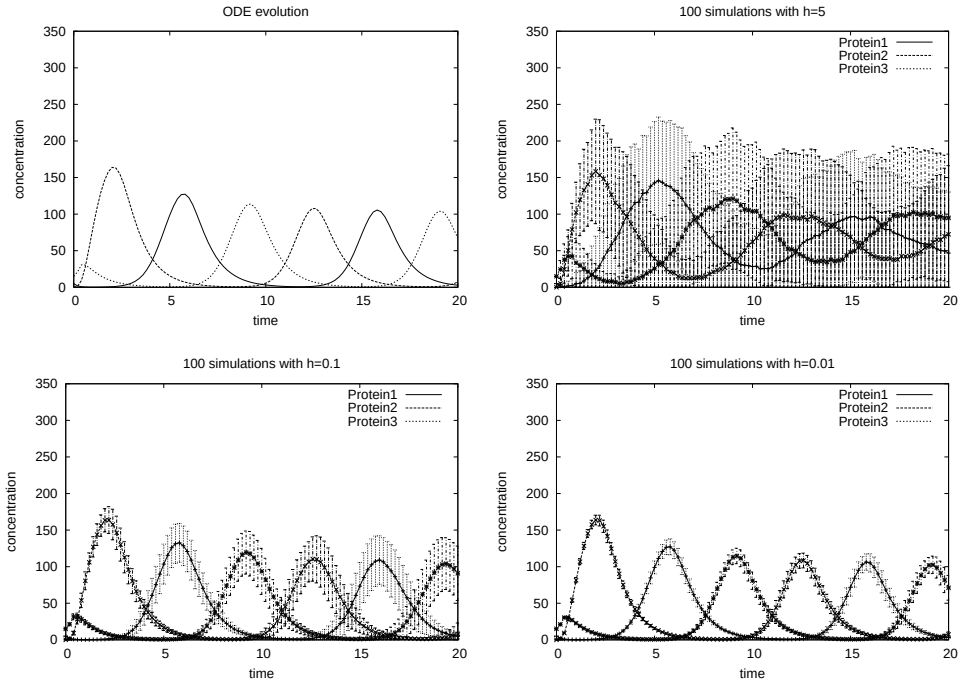

Fig. 5. Some analysis results for the Repressilator.

that these distances are not normalized and the sum is over all the six species of the system. By observing these values we can see that the distance between the two models decreases with smaller step sizes.

distance	h = 5	h = 1	h = 0.1	h = 0.01
$f_{dist,avg}$	969105	209817	21796	540

7 Discussion and conclusions

There are three main contributions of this work. Firstly, we gave a formal definition of the CTMC derived from a Bio-PEPA system. We called it the *CTMC with levels*, as its states are characterized in terms of the concentration levels for each of the species of the system. Secondly, we presented the maps from Bio-PEPA to CTMC and ODEs. Thirdly, we investigated the relationship between the ODE model and the CTMC obtained from the same Bio-PEPA system. Finally, we tested our approach against a simple example describing a dimerization reaction and the Repressilator network.

Based on our results, in the case of a low number of levels (i.e. rough granularity), the behaviour shown by the expected value of the CTMC might or might not agree from the ODE time evolution. We use a lower h in order to decrease the variability of the CTMC model, and as predicted by Kurtz's Theorem, obtain a global behaviour that is closer to that given by the deterministic approach. This can allow more flexibility to the modeller. For instance, in the presence of experimental observations that suggest a certain degree of uncertainty, we can choose the model that better agrees with those observations.

We proposed a distance measure between the CTMC and ODE models and this

has been used for finding a "good" granularity for the system. The definition of distance is based on the numerical solution of CTMC. As observed in the paper, the derivation of the numerical solution is often impracticable. In order to overcome this drawback we proposed an alternative definition of distance based on the average of a number of simulation runs. The selection of the appropriate number of runs remains an open problem. A deeper investigation of this point and the study of other definitions of distance between models is planned.

Finally, other future investigations concern the validation of the system. In this paper we focused on the comparison between two different representations of a biological model, obtained from the same Bio-PEPA system. An important aspect which remains to be considered is the validation of the system against experimental data and existing knowledge.

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, volume 4230 of LNCS, pages 1–23, Springer, 2006.

[3] Calder, M., V. Vyshemirsky, D. Gilbert and R. Orton. Analysis of Signalling Pathways using Continuous Time Markov Chains. *T. Comp. Sys. Biology*, VI, volume 4220 of LNCS, pages 44–67, Springer, 2006.

[4] Ciocchetta, F. and J. Hillston. Bio-PEPA: an extension of the process algebra PEPA for biochemical networks. Proc. of *FBTC 2007*, *Electronic Notes in Computer Science*, volume 194, number 3, pages 103–117, 2008.

[5] Ciocchetta, F. and J. Hillston. Bio-PEPA: a framework for the modelling and analysis of biological systems. Technical Report of the School of Informatics, University of Edinburgh, EDI-INF-RR-1231, 2008.

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

[7] Elowitz, M.B. and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, volume 403, number 6767, pages 335–338, 2000.

[8] Geisweiller, N., J. Hillston and M. Stenico. Relating continuous and discrete PEPA models of signalling pathways. To appear in *Theoretical Computer Science*, 2007.

[9] Gillespie, D.T.. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, volume 81, pages 2340–2361, 1977.

[10] Hillston, J.. *A Compositional Approach to Performance Modelling*, Cambridge University Press, 1996.

[11] Kurtz, T.G.. Limit Theorems for Sequences of Jump Markov Processes Approximating Ordinary Differential Processes. *Journal of Applied Probability*, volume 8, number 2, pages 344–356, 1971.

[12] Kurtz, T.G.. The Relationships between Stochastic and Deterministic Models for Chemical Reactions. *The Journal of Chemical Physics*, volume 57, number 7, pages 2976–2978, 1971.

[13] Kurtz, T.G.. Solutions of Ordinary Differential Equations as Limit of Pure Jump Markov Processes. *Journal of Applied Probability*, volume 7, number 1, pages 49–58, 1971.

[14] Prism web site. http://www.prismmodelchecker.org/

[15] Priami, C. and P. Quaglia. Beta-binders for biological interactions. *Proc. of CMSB'04*, 2004.

[16] 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.

A Appendix for the Reviewers

A.1 Auxiliary definitions for Bio-PEPA

In the following we report some auxiliary definitions. For more details see [4,5].

First of all we consider the definition of the set of action types.

Definition A.1 The set of current action types enabled in the model component P, denoted $\mathcal{A}(P)$, is defined as:

$$\mathcal{A}((\alpha,\kappa)\ \mathtt{op}\ S) = \{\alpha\} \quad \mathcal{A}(S_1 + S_2) = \mathcal{A}(S_1) \cup \mathcal{A}(S_2)$$
$$\mathcal{A}(S(l)) = \mathcal{A}(S) \qquad \mathcal{A}(C) = \mathcal{A}(S) \text{ where } C \stackrel{def}{=} S$$
$$\mathcal{A}(P_1 \underset{\mathcal{L}}{\bowtie} P_2) = \mathcal{A}(P_1)\backslash\mathcal{L} \cup \mathcal{A}(P_2)\backslash\mathcal{L} \cup (\mathcal{A}(P_1) \cap \mathcal{A}(P_2) \cap \mathcal{L})$$

If $\mathcal{P}$ is a Bio-PEPA system with model component P, the set of current action types enabled in $\mathcal{P}$ is $\mathcal{A}(\mathcal{P}) = \mathcal{A}(P)$.

The behaviour of the system is defined in terms of an operational semantics. We defined two relations over the processes. The former, called the *capability relation* (indicated with $\rightarrow_c$), supports the derivation of quantitative information and it is auxiliary to the latter which is called the *stochastic relation* (indicated with $\rightarrow_s$). The stochastic relation gives us the rates associated with each action. a formal definition of these relations and their rules is reported **??**.

The following definitions concern the derivative of a component, the derivative set and the derivative graph. We refer to the relation $\rightarrow_s$. The case of $\rightarrow_c$ is analogous.

Definition A.2 If $\mathcal{P} \xrightarrow{(\alpha,r)}_s \mathcal{P}'$ then $\mathcal{P}'$ is a one-step $\rightarrow_s$ system derivative of $\mathcal{P}$.
If $\mathcal{P} \xrightarrow{(\alpha_1,r_1)}_s \mathcal{P}_1 \xrightarrow{(\alpha_2,r_2)}_s .... \xrightarrow{(\alpha_n,r_n)}_s \mathcal{P}'$ then $\mathcal{P}'$ is a system derivative of $\mathcal{P}$.

We can indicate the sequence $\xrightarrow{\gamma_1}_s \xrightarrow{\gamma_2}_s .... \xrightarrow{\gamma_n}_s$ with $\xrightarrow{\mu}_s$, where μ denotes the sequence $\gamma_1\gamma_2, ...\gamma_n$ (possibly empty).

Definition A.3 A system α-derivative of $\mathcal{P}$ is a system $\mathcal{P}'$ such that $\mathcal{P} \xrightarrow{(\alpha,r)}_s \mathcal{P}'$. For each $\alpha \in \mathcal{A}$ we have at most one system α-derivative of a system $\mathcal{P}$.

Definition A.4 The system derivative set ds($\mathcal{P}$) is the smallest set such that:

- $\mathcal{P} \in ds(\mathcal{P})$;
- if $\mathcal{P}' \in ds(\mathcal{P})$ and there exists $\alpha \in \mathcal{A}(\mathcal{P}')$ such that $\mathcal{P}' \xrightarrow{(\alpha,r)}_s \mathcal{P}''$ then $\mathcal{P}'' \in ds(\mathcal{P})$.

Definition A.5 The system derivative graph $\mathcal{D}(\mathcal{P})$ is the labelled directed multi-graph whose set of nodes is ds($\mathcal{P}$) and whose multi-set of arcs are elements in $ds(\mathcal{P}) \times ds(\mathcal{P}) \times \Gamma$.

In the derivation of the CTMC (Section 3) we need to identify the actions describing the interactions from one state to another.

Definition A.6 Let $\mathcal{P}$ be a Bio-PEPA system and let $P = \pi_P(\mathcal{P})$. Let P_u, P_v be two derivatives of a model component P with P_v a one-step derivative of P_u.

The set of action types associated with the transitions from the process P_u to the process P_v is denoted $\mathcal{A}(P_u|P_v)$.

A.2 Kurtz's theorem

In the following we report the main theorem described in [11].

First of all we give the definition of *Density Dependent Markov Chain*, as here we limit our attention to this kind of CTMC.

Definition A.7 A family of CTMCs X_V, for some parameter V, is called *density dependent* if and only if there exists a continuous function $f(x,s)$, $x \in \mathbb{R}^n$, $s \in \mathbb{Z}^n$, such that the entries of the infinitesimal generators are given by:

$$q_{k,k'} = f(\sigma V^{-1}, s) \cdot V \qquad s \neq 0$$

with σ the state vector and s a transition vector that contains the modifications for each state of each species (i.e. the number of copies to add or substract) when the transition is taken.

Theorem A.8 *Let X_V be a family of density dependent CTMCs with the infinitesimal generator matrix as in the definition above. Assume $X(t)$ is the solution of the ODE system $\dot{X} = F(X)$, where $F(X) = \sum_s s f(x,s)$ and let $X(0) = x_0$.*
If there exists an open set $E \subset \mathbb{R}^n$ such that $X(t) \in E$ and

(i) $\exists M,\ \forall x, y \in E \quad \mid F(x) - F(y) \mid < M \mid x - y \mid$;

(ii) $\sup \sum_s \mid s \mid f(x,s) < \infty$;

(iii) $\lim_{d\to\infty} \sup_{x\in E} \sum_{|s|>d} \mid s \mid f(x,s) = 0$

then

$$\lim_{V\to\infty} V^{-1} X_V(0) = x_0 \implies \forall \delta > 0,\ \forall t > 0 \quad \lim_{V\to\infty} \mathbb{P}(\sup_{z<t} \mid V^{-1} X_V(z) - X(z) \mid > \delta) = 0$$

The theorem above claims that, under some conditions, the system of ODEs can be defined as a limit of a sequence of density dependent CTMCs. In the version of the theorem reported above the states represent number of individuals and are normalized with respect to a parameter V (in this case the volume). Therefore $V^{-1}X_V(z)$ represents the scaled Markov process with concentrations.

A.2.1 Inequality

In [13] the author showed an inequality expressing an estimate for the convergence probability of the CTMC to the associated ODE system. The inequality is:

$$\mathbb{P}(\sup_{z<t} \mid V^{-1} X_V(z) - X(z) \mid \geq \delta) \leq \frac{t\Gamma}{V\theta^2}$$

where we have:

- $K_\delta = \{x : \inf_{z\leq t} \mid x - X(z) \mid \leq \delta\}$, that is the set of points within a δ distance of the trajectory $X(z)$, for $z \leq t$;

- $\Gamma(x) = \sum_{j=1}^{m} \sum_{i=1}^{n} d_{ij}^2 f_j(x)$ and $\Gamma = \sup_{x \in K_\delta} \Gamma(x)$;
- $M = \sup_{x,y \in K_\delta} [\mid F(x) - F(y) \mid / \mid x - y \mid]$
- $\mu = \sup_{i/V \in K_\delta} [\mid F(i/V) - F_V(i) \mid]$, where i indicates the CTMC states (here defined in terms of number of molecules/elements) and F_V is F in the case of CTMC depending on V.
- $\theta = \delta e^{-Mt} - \mid V^{-1} X_V(0) - x_0 \mid - t\mu > 0$.

In our case the scaling factor is in terms of h instead of V. Therefore the estimation of the error is $\frac{t\Gamma h}{\theta^2}$, where Γ and θ are as defined above. Note that the step size h appears explicitly in this estimate and also in the definition of μ. In particular in our case μ is defined as

$$\mu = \sup_{x \in K_\delta} [\mid F(\lceil x/h \rceil \cdot h) - F(x) \mid]$$

that is the difference between the continuous function $F(x)$ and the step function $F(\lceil x/h \rceil \cdot h)$, considered in the derivation of the rates for the CTMC with levels. This difference is expected to decrease with h.

Note that the use of this inequality in order to derive a good h can be quite complex.

A.3 The Repressilator model in Bio-PEPA

In the following we describe the Bio-PEPA system corresponding to the Repressilator network. The parameters and the initial concentrations are the ones defined in the paper [7].

There are no compartments defined explicitly in the model. So we consider the default compartment $v_{Cell} : 1$. The step size, the number of levels, the initial concentration and location of species are declared as:

$mRNA1 : h = 1, N = 10, M_0 = 0, \_, v_{Cell}$; $mRNA2 : h = 1, N = 10, M_0 = 0, \_, v_{Cell}$;

$mRNA3 : h = 1, N = 10, M_0 = 0, \_, v_{Cell}$; $P1 : h = 5, N = 50, M_0 = 5, \_, v_{Cell}$;

$P2 : h = 5, N = 50, M_0 = 0, \_, v_{Cell}$; $P3 : h = 5, N = 50, M_0 = 15, \_, v_{Cell}$;

$Res : h = 1, N = 1, \_, \_, v_{Cell}$; $CF : h = 1, N = 1, \_, \_, v_{Cell}$;

It is worth noting that in the original model the genes are not represented explicitly. In Bio-PEPA we introduce the auxiliary component CF to define the transcription. The component *Res* is used to describe degradation. For all the species we consider the step size $h = 5$. The numbers of levels are derived in terms of the concentration in the biological model. The set of functional rates is:

$$f_{tr1} = \frac{\alpha}{1+P3^2} + \alpha_0; \qquad f_{tr2} = \frac{\alpha}{1+P1^2} + \alpha_0; \qquad f_{tr3} = \frac{\alpha}{1+P2^2} + \alpha_0;$$

$$f_{trl1} = fMA(\beta); \quad f_{trl2} = fMA(\beta); \quad f_{trl3} = fMA(\beta);$$

$$f_{di} = fMA(1)\ i = 1,2,3,4,5,6;$$

$fMA(r)$ stands for mass-action with rate constant r. All the three repressors have same behaviour except for their DNA-binding specificities. We assume that all the degradation reactions have rate 1. The other parameters are: $\alpha = 250$; $\alpha_0 = 0$; $\beta = 5$.

The species components are defined as:

$$\begin{aligned}
mRNA1 &\stackrel{def}{=} (d1,1)\downarrow mRNA1 + (tr1,1)\uparrow mRNA1 + (trl1,1) \oplus mRNA1;\\
mRNA2 &\stackrel{def}{=} (d2,1)\downarrow mRNA2 + (tr2,1)\uparrow mRNA2 + (trl2,1) \oplus mRNA2;\\
mRNA3 &\stackrel{def}{=} (d3,1)\downarrow mRNA3 + (tr3,1)\uparrow mRNA3 + (trl3,1) \oplus mRNA3;\\
P1 &\stackrel{def}{=} (d4,1)\downarrow P1 + (trl1,1)\uparrow P1 + (tr3,1) \ominus P1;\\
P2 &\stackrel{def}{=} (d5,1)\downarrow P2 + (trl2,1)\uparrow P2 + (tr1,1) \ominus P2;\\
P3 &\stackrel{def}{=} (d6,1)\downarrow P3 + (trl3,1)\uparrow P3 + (tr2,1) \ominus P3;\\
CF &\stackrel{def}{=} (tr1,1) \odot CF + (tr1,1) \odot CF + (tr1,1) \odot CF;\\
Res &\stackrel{def}{=} (d1,1) \odot Res + (d2,1) \odot Res + (d3,1) \odot Res\\
&\quad +(d4,1) \odot Res + (d5,1) \odot Res + (d6,1) \odot Res;
\end{aligned}$$

Finally, the model is defined as:

$$\Big(\Big(((((mRNA1(0) \bowtie mRNA2(2)) \bowtie mRNA3(0)) \underset{\{trl1,tr3\}}{\bowtie} P1(1)) \underset{\{trl2,tr1\}}{\bowtie} P2(0)) \underset{\{trl3,tr2\}}{\bowtie} P3(3)\Big) \underset{\{tr1,tr2,tr3\}}{\bowtie} CF(1)\Big) \underset{\{d1,d2,d3,d4,d5,d6\}}{\bowtie} Res(0)$$

The initial levels are defined according to the initial values of the model.

A.3.1 Derivation of CTMC and ODE system

From the Bio-PEPA system above we obtain a CTMC with 50^6 states, where 50 is the number of levels considered and 6 is the number of species considered.

The stoichiometry matrix D associated with the Bio-PEPA system described above is

$$\left(\begin{array}{cccccccccccc|c}
tr1 & tr2 & tr3 & trl1 & trl2 & trl3 & d1 & d2 & d3 & d4 & d5 & d6 & \\
\hline
+1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & x_{m1} \\
0 & +1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & x_{m2} \\
0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & x_{m3} \\
0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & x_{p1} \\
0 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & x_{p2} \\
0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & -1 & x_{p3}
\end{array}\right)$$

Each row describes the stoichiometric coefficients for a given species in all the reactions. The last column reports the name of the variables associated with the species in the network (m_i stands for $mRNAi$ and p_i for $P1$, $i = 1, 2, 3$). The kinetic vector $\mathbf{v_{KL}}$ is:

$$\left(\frac{\alpha}{1+x_{p3}^2} + \alpha_0, \frac{\alpha}{1+x_{p1}^2} + \alpha - 0, \frac{\alpha}{1+x_{p2}^2} + \alpha_0, \beta \cdot x_{m1}, \beta \cdot x_{m2}, \beta \cdot x_{m3}, 1 \cdot x_{m1}, 1 \cdot x_{m2}, 1 \cdot x_{m3}, 1 \cdot x_{p1}, 1 \cdot x_{p2}, 1 \cdot x_{p3}\right)^T$$

The ODE system is:

$$\begin{aligned}
\frac{dx_{m1}}{dt} &= +\frac{\alpha}{1+x_{p3}^2} + \alpha_0 - 1 \cdot x_{m1} \\
\frac{dx_{m2}}{dt} &= +\frac{\alpha}{1+x_{p1}^2} + \alpha_0 - 1 \cdot x_{m2} \\
\frac{dx_{m3}}{dt} &= +\frac{\alpha}{1+x_{p2}^2} + \alpha_0 - 1 \cdot x_{m3} \\
\frac{dx_{p1}}{dt} &= +\beta \cdot x_{m1} - 1 \cdot x_{p1} \\
\frac{dx_{p2}}{dt} &= +\beta \cdot x_{m2} - 1 \cdot x_{p2} \\
\frac{dx_{p3}}{dt} &= +\beta \cdot x_{m3} - 1 \cdot x_{p3}
\end{aligned}$$

with the initial conditions $(x_{m1}, x_{m2}, x_{m3}, x_{p1}, x_{p2}, x_{p3})^T = (0, 0, 0, 5, 0, 15)$.