Equivalence and Discretisation in Bio-PEPA

Vashti Galpin[1] and Jane Hillston[1,2]

[1] LFCS, School of Informatics, University of Edinburgh
[2] Centre for Systems Biology (CSBE), University of Edinburgh
Vashti.Galpin@ed.ac.uk, Jane.Hillston@ed.ac.uk

Abstract. Bio-PEPA is a process algebra for modelling biological systems. An important aspect of Bio-PEPA is the ability it provides to discretise concentrations resulting in a smaller, more manageable state space. The discretisation is based on a step size which determines the size of each discrete level and also the maximum number of levels. This paper considers the relationship between two discretisations of the same Bio-PEPA model that differ only in the step size and hence the maximum number of levels, by using the idea of equivalence from concurrency and process algebra. We present a novel behavioural semantic equivalence, compression bisimulation, that equates two discretisations of the same model and we show that this equivalence is a congruence with respect to the synchronisation operator.

1 Introduction

The use of process algebras for modelling biological systems has become a popular technique [1–5]. Some approaches use the process algebra as originally defined for description of computer systems and in others a process algebra is tailored to be specific to systems biology. One of the latter class is Bio-PEPA [6] which was developed from the stochastic process algebra PEPA [7] and has been successfully used to model Goldbeter's model of cyclin oscillation [8, 9], the Repressilator [10], genetic networks [6], the MAPK model [11], the Neurospora circadian clock [12] and the gp130/JAK/STAT pathway [13]. This paper investigates a semantic equivalence for Bio-PEPA.

An important aspect of Bio-PEPA is the ability it provides for the discretisation of concentrations. Instead of working with a "process-as-molecule" approach, it uses a "process-as-species" approach whereby a process can either be parameterised by concentration or by a discrete level which is obtained from dividing the concentration into a discrete number of chunks. Typically, there is a fixed step size and each species has a maximum number of levels dependent on its maximum concentration. For a given step size, we call a system with levels a discretisation.

Bio-PEPA distinguishes itself from many other process algebras for modelling biological systems by providing multiple analyses including continuous-time Markov chains (CTMCs), ordinary differential equations (ODEs), stochastic simulation and model checking. By developing a semantic equivalence for Bio-PEPA, we have a new type of analysis based on behaviour that can be used to compare models.

Semantic equivalences are an important technique in process algebra for specifying notions of equivalent behaviour. They equate processes that have the same behaviour,

and can be divided into qualitative equivalences which only consider the behaviour in terms of which actions can be performed and quantitative equivalences which consider the rates at which actions can happen as part of the behaviour. In this paper we consider a qualitative equivalence. Adding quantitative aspects is ongoing research.

Semantic equivalences typically have two important properties – they are equivalence relations (hence the name) and congruence relations. A congruence relation is a relation that is preserved by the operators of the process algebra. For example, if ϕ is a binary operator, then an equivalence $\equiv$ is preserved by this operator if $P \equiv Q$ implies that $P \phi R \equiv Q \phi R$ and $R \phi P \equiv R \phi Q$ for every process R.

These properties are important when modelling and evaluating the concurrent behaviour of computer systems as they let us substitute like for like thereby exploiting the compositionality provided by a process algebra. This allows for the substitution of a system with a smaller state space or other desirable properties, and make the analysis of the system easier. In applying the idea of a semantic equivalence in systems biology, similar advantages will be gained hence the importance and timeliness of this research. Moreover, since the semantic equivalence we define is based upon what reactions can be observed, it is well suited for biological modelling.

In searching for a suitable equivalence, there are at least two approaches that can be taken. One is to consider existing equivalences from the literature. The other is to consider what behaviours we want to consider as identical and to develop an equivalence from this starting point. This is the approach taken here.

We have a immediate candidate for what we want to consider the same. For a Bio-PEPA system, we can consider two different discretisations of that system. Since they both represent the same system, we want their behaviours to be identified (assuming neither have few enough levels to give pathological behaviour). This approach is suitable since semantic equivalences are used to identify the same behaviour in different abstractions of a system, and clearly two discretisations are two abstractions.

Starting from this point, we define an equivalence relation over the states of the model that relates states that have the same possible reactions. This equivalence relation defines equivalence classes of states with the same behaviour and from this we can use a classical notion of equivalence, bisimulation, to define our semantic equivalence.

This new semantic equivalence, compression bisimulation, has not been chosen randomly but through understanding the differences between discretisations and ensuring that the semantic equivalence has the desirable properties mentioned above. Ensuring a semantic equivalence is an equivalence relation is not hard. Establishing congruence is much harder because stoichiometry coefficients greater than one lead to a complex transition system. Being able to prove congruence played a major role in the selection of compression equivalence. Moreover, since two discretisations of a single species should have the same behaviour under the equivalence, it was necessary to prove this as well.

The first result of this paper shows that in the sequential case, a single species, two discretisations are related by compression bisimulation. The second shows that compression bisimulation is a congruence with respect to the cooperation operator. A corollary of this is that in the general case of a Bio-PEPA system, namely with multiple species, two discretisations are related by compression bisimulation.

The structure of the document is as follows. First, we introduce Bio-PEPA, then we present some basic ideas relating to semantic equivalences and then we present compression bisimulation and show it has the desired properties. We then give an example.

2 Bio-PEPA

This section presents an overview of Bio-PEPA [6]. The main components of a Bio-PEPA system are the sequential components describing the behaviour of each of the species and the model component describing the interactions between the species and initial amounts. Additionally, a context is defined, including functional rates, compartments, parameters. The syntax of the sequential (species) components is defined as

$$S ::= (\alpha, \kappa) \text{ op } S \mid S + S \mid C \qquad \text{op} ::= \downarrow \mid \uparrow \mid \oplus \mid \ominus \mid \odot.$$

In the prefix term (α, κ) op S, α is an action name and can be viewed as the name or label of a reaction, κ is the stoichiometry coefficient of the species and the prefix combinator op represents the role of the element in the reaction. Specifically, $\downarrow$ is 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 $C \stackrel{def}{=} S$. The syntax of model components is defined as

$$P ::= P \underset{\mathcal{L}}{\bowtie} P \mid S(x)$$

The process $P \underset{\mathcal{L}}{\bowtie} Q$ denotes the synchronisation between components P and Q and the set $\mathcal{L}$ specifies those activities on which the components must synchronise. In the model component $S(x)$, the parameter $x \in \mathbb{R}$ represents the concentration or level. We work with a constrained set of Bio-PEPA model components as given by the following definition which specifies a well-formed set of components. We ensure that a species consists of a choice between reactions, and no reaction name is repeated within a species. At the model level, there can only be one species component for each species.

Definition 1. *A Bio-PEPA sequential component C is* well-defined *if it has the form*

$$C \stackrel{def}{=} (\alpha_1, \kappa_1) \, op_1 \, C + \ldots + (\alpha_n, \kappa_n) \, op_n \, C \quad \textit{written as} \quad C \stackrel{def}{=} \sum_{i=1}^{n} (\alpha_i, \kappa_i) \, op_i \, C$$

where $\alpha_i \neq \alpha_j$ for $i \neq j$.

   A Bio-PEPA model component P is well-defined *if it has the form*

$$P \stackrel{def}{=} C_1(x_1) \underset{\mathcal{L}_1}{\bowtie} \ldots \underset{\mathcal{L}_{p-1}}{\bowtie} C_p(x_p),$$

each C_i is a well-defined sequential component, the elements of each $\mathcal{L}_j$ appear in P and if $i \neq j$ then $C_i \neq C_j$.

We define a Bio-PEPA system, consisting of a set of well-defined sequential components, a well-defined model component and context, as follows.

$$
\text{prefixReac} \quad \frac{}{(\alpha,\kappa){\downarrow}S(l) \xrightarrow{(\alpha,[S:\downarrow(l,\kappa)])}_c S(l-\kappa)} \quad \kappa \leq l \leq N_S
$$

$$
\text{prefixProd} \quad \frac{}{(\alpha,\kappa){\uparrow}S(l) \xrightarrow{(\alpha,[S:\uparrow(l,\kappa)])}_c S(l+\kappa)} \quad 0 \leq l \leq N_S - \kappa
$$

$$
\text{prefixMod} \quad \frac{}{(\alpha,\kappa)\ \mathrm{op}\ S(l) \xrightarrow{(\alpha,[S:\mathrm{op}(l,\kappa)])}_c S(l)} \quad \begin{array}{l} 0 < l \leq N_S \ \text{if}\ \mathrm{op} = \oplus \\ 0 \leq l \leq N_S \ \text{if}\ \mathrm{op} \in \{\ominus,\odot\} \end{array}
$$

$$
\text{choice1} \quad \frac{S_1(l) \xrightarrow{(\alpha,w)}_c S_1'(l)}{(S_1 + S_2)(l) \xrightarrow{(\alpha,w)}_c S_1'(l)}
$$

$$
\text{choice2} \quad \frac{S_2(l) \xrightarrow{(\alpha,w)}_c S_2'(l)}{S_1 + S_2(l) \xrightarrow{(\alpha,w)}_c S_2'(l)}
$$

$$
\text{constant} \quad \frac{S(l) \xrightarrow{(\alpha,[S:\ \mathrm{op}\ (l,\kappa)])}_c S'(l)}{C(l) \xrightarrow{(\alpha,[C:\ \mathrm{op}\ (l,\kappa)])}_c S'(l)} \quad C \stackrel{def}{=} S
$$

$$
\text{coop1} \quad \frac{P_1 \xrightarrow{(\alpha,w)}_c P_1'}{P_1 \bowtie_M P_2 \xrightarrow{(\alpha,w)}_c P_1' \bowtie_M P_2} \quad \alpha \notin M
$$

$$
\text{coop2} \quad \frac{P_2 \xrightarrow{(\alpha,w)}_c P_2'}{P_1 \bowtie_M P_2 \xrightarrow{(\alpha,w)}_c P_1 \bowtie_M P_2'} \quad \alpha \notin M
$$

$$
\text{coop3} \quad \frac{P_1 \xrightarrow{(\alpha,w_1)}_c P_1' \quad P_2 \xrightarrow{(\alpha,w_2)}_c P_2'}{P_1 \bowtie_M P_2 \xrightarrow{(\alpha,w_1::w_2)}_c P_1' \bowtie_M P_2'} \quad \alpha \in M
$$

Fig. 1. Operational semantics of Bio-PEPA

Definition 2. *A Bio-PEPA system $\mathcal{P}$ is a 6-tuple $\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, P \rangle$, where $\mathcal{V}$ is the set of compartments, $\mathcal{N}$ is the set of quantities describing each species, $\mathcal{K}$ is the set of parameters, $\mathcal{F}$ is the set of functional rates, $Comp$ is the set of well-defined sequential components and P is a well-defined model component.*

Elements of $\mathcal{N}$ have the form $C : H = h, N = n, M = m, V = v, unit = u$ where C is a species name that is defined in $Comp$, $H = h$ defines the step size, $N = n$ defines the maximum number of levels for C, $M = m$ defines the maximum concentration for C, $V = v$ names the compartment in which C appears and $unit = u$ defines the measurement unit of the concentration.

The notation $\langle \mathcal{T}, P \rangle$ will be used for $\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, P \rangle$ since the details of the tuple are not relevant here. The model component is defined in terms of concentrations, but can be expressed terms of levels where these are a discretisation of the concentration. In $\mathcal{N}$, for each species, there are three elements H, N and M which represent the step size, the maximum number of levels and the maximum concentration respectively. Their relationship is defined as $N = \lceil M/H \rceil$. A species with a concentration then has an associated level in the range $0, 1, \ldots, N - 1, N$, giving a total of $N + 1$ possible levels. We call the system obtained in this way a *Bio-PEPA system with levels*. We will assume that the step size H is the same for all species to ensure conservation of mass.

Final
$$\dfrac{P \xrightarrow{(\alpha,w)}_c P'}{\langle \mathcal{V},\mathcal{N},\mathcal{K},\mathcal{F},Comp,P\rangle \xrightarrow{(\alpha,r_\alpha[w,\mathcal{N},\mathcal{K}])}_s \langle \mathcal{V},\mathcal{N},\mathcal{K},\mathcal{F},Comp,P'\rangle}$$

Enrich
$$\dfrac{P \xrightarrow{(\alpha,w)}_c P'}{\langle \mathcal{V},\mathcal{N},\mathcal{K},\mathcal{F},Comp,P\rangle \xrightarrow{(\alpha,w)}_{sc} \langle \mathcal{V},\mathcal{N},\mathcal{K},\mathcal{F},Comp,P'\rangle}$$

Fig. 2. Operational semantics of Bio-PEPA (continued)

The operational semantics for Bio-PEPA systems with levels is given in Tables 1 and 2. In the first table, N_S refers to the maximum number of levels for the species S. In the rule `coop3`, $w_1::w_2$ represents list concatenation. For the rule `Final`, $r_\alpha[w,\mathcal{N},\mathcal{K}] = f_\alpha[w,\mathcal{N},\mathcal{K}]/H \in (0,\infty)$ where f_α is the functional rate for the reaction α from $\mathcal{F}$ and H is the step size. We do not discuss this or the string w further as the equivalences in this paper only consider the action α and ignore the rest of the transition label.

The operational semantics creates three different transition systems. The rules for $\rightarrow_c$ define the capability relation. The rule `Final` defines the system/stochastic relation $\rightarrow_s$ which includes the context, and the rate at which the transition takes place appears together with the action. The rule `Enrich` defines the system-capability relation $\rightarrow_{sc}$. This relation is necessary since it contains the context information as well as the detailed information that is captured in the list of strings w.

The following definition describes the derivative set for the relation $\rightarrow_{sc}$. In this paper, we work almost exclusively with this relation since it provides the necessary information about the context.

Definition 3. *The system-capability derivative set ds(P) is the smallest set such that* $\mathcal{P} \in ds(\mathcal{P})$ *and if* $\mathcal{P}' \in ds(\mathcal{P})$ *and* $\mathcal{P}' \xrightarrow{(\alpha,r)}_{sc} \mathcal{P}''$ *then* $\mathcal{P}'' \in ds(\mathcal{P})$.

The next definition captures the reactions that are immediately possible with respect to the operational semantics. This means it takes into account the stoichiometry of a reaction as well as the current level of a species.

Definition 4. *The set of* current actions *enabled in* $\langle \mathcal{T},P\rangle$ *is defined as* $\mathcal{A}(\langle \mathcal{T},P\rangle) = \mathcal{A}(P)$ *where* N_S *is the maximum number of levels for species component S.*

$$\mathcal{A}(((\alpha,\kappa)\downarrow S)(l)) = \{\alpha\} \ \text{if } \kappa \leq l \leq N_S \ \text{otherwise } \emptyset$$
$$\mathcal{A}(((\alpha,\kappa)\uparrow S)(l)) = \{\alpha\} \ \text{if } 0 \leq l \leq N_S - \kappa \ \text{otherwise } \emptyset$$
$$\mathcal{A}(((\alpha,\kappa)\oplus S)(l)) = \{\alpha\} \ \text{if } 0 < l \leq N_S \ \text{otherwise } \emptyset$$
$$\mathcal{A}(((\alpha,\kappa)\ominus S)(l)) = \{\alpha\} \ \text{if } 0 \leq l \leq N_S \ \text{otherwise } \emptyset$$
$$\mathcal{A}(((\alpha,\kappa)\odot S)(l)) = \{\alpha\} \ \text{if } 0 \leq l \leq N_S \ \text{otherwise } \emptyset$$
$$\mathcal{A}((S_1 + S_2)(l)) = \mathcal{A}(S_1(l)) \cup \mathcal{A}(S_2(l))$$
$$\mathcal{A}(C(l)) = \mathcal{A}(S(l)) \ \text{where } C \stackrel{\text{def}}{=} S$$
$$\mathcal{A}(P_1 \bowtie_L P_2) = \mathcal{A}(P_1) \setminus L \cup \mathcal{A}(P_2) \setminus L \cup (\mathcal{A}(P_1) \cap \mathcal{A}(P_2) \cap L)$$

The stoichiometry plays a role in defining the set of current actions. A species definition specifies a set of actions (reactions), but the current action set may be a subset if the current level is insufficient to satisfy the constraints imposed by the stoichiometry.

Since we are working with Bio-PEPA systems that vary only in step size and maximum numbers of levels, we require notation to capture this. Given a Bio-PEPA system $\mathcal{P}$ we can define a system where the lowest maximum number of levels for any species is n. As mentioned previously, we assume that H is identical for all components.

Definition 5. *Let* $\mathcal{P} = \langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, P \rangle$ *be a Bio-PEPA system with well-defined P parameterised by concentration. For $n \in \mathbb{N}$, the* Bio-PEPA system with levels $\mathcal{P}^n$ *is defined as* $\mathcal{P}^n = \langle \mathcal{V}, \mathcal{N}', \mathcal{K}, \mathcal{F}, Comp, P' \rangle$ *where*

1. $\gamma = (1/n) \cdot \min\{m \mid C : H = h, N = n', M = m, V = v, unit = u \in \mathcal{N}\}$
2. $C : H = h, N = n', M = m, V = v, unit = u \in \mathcal{N} \quad \Rightarrow$
$$C : H = \gamma, N = \lceil m/\gamma \rceil, M = m, V = v, unit = u \in \mathcal{N}'$$
3. $P \stackrel{\text{def}}{=} C_1(x_1) \underset{L_1}{\bowtie} \dots \underset{L_{p-1}}{\bowtie} C_p(x_p) \Rightarrow P' \stackrel{\text{def}}{=} C_1(\lceil x_1/\gamma \rceil) \underset{L_1}{\bowtie} \dots \underset{L_{p-1}}{\bowtie} C_p(\lceil x_p/\gamma \rceil)$

$\mathcal{N}$ contains information about each species. The definition above identifies the species with the smallest concentration, determines the new step size that will ensure n levels for that species and then adjusts the other species to use the same step size (to conserve mass) hence modifying $\mathcal{N}$. Since $\mathcal{P}$ is a Bio-PEPA system with species components parameterised by concentration and we wish to work with a system with levels, the initial concentrations in the third point are transformed to initial levels. We will use the notation $\mathcal{P}^n = \langle \mathcal{T}^n, P \rangle$ to indicate that the lowest number of maximum levels for any species is n and refer to $\mathcal{P}^n$ as a discretisation of $\mathcal{P}$.

3 Semantic Equivalences

In process algebras, a semantic equivalence defines what it means for two models to have the same behaviour. A classical notion of equivalence is that of bisimulation [14].

Definition 6. *A binary relation $\mathcal{R}$ is a* bisimulation *if for any $(P, Q) \in \mathcal{R}$ and for any θ whenever*

1. $P \xrightarrow{\theta} P'$, *there exists Q' such that $Q \xrightarrow{\theta} Q'$ and $(P', Q') \in \mathcal{R}$, and*
2. $Q \xrightarrow{\theta} Q'$, *there exists P' such that $P \xrightarrow{\theta} P'$ and $(P', Q') \in \mathcal{R}$*

P and Q are bisimilar, *$P \sim Q$ if $(P, Q) \in \mathcal{R}$ for some bisimulation $\mathcal{R}$.*

This leads to the definition $\sim = \bigcup \{\mathcal{R} \mid \mathcal{R} \text{ a bisimulation}\}$ and one can show that $\sim$ is the largest bisimulation. Moreover, it can also be shown that bisimulation is an equivalence relation therefore it is reflexive, symmetric and transitive. Bisimulation is a fine-grained notion of behaviour and equates far fewer models than language/trace equivalence, for example. It requires that related models can match each other's transitions and that the resultant models also have this property. Consider the labelled transition systems in Figure 3. They generate the same strings/traces but they are not bisimilar because we cannot find anything to match with Q_1. Q_2 is not suitable since it only has a b transition and Q_2' is not suitable since it only has a c transition.

As mentioned in the introduction, we also wish that our new semantic equivalence be a congruence with respect to the language we use.
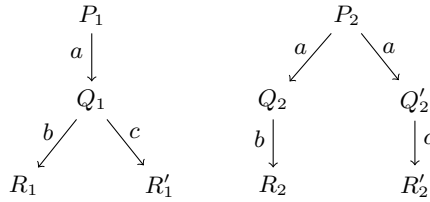
Fig. 3. Example of transition systems that are not bisimilar

$$\langle \mathcal{T}^2, A(2)\rangle \xrightarrow{(\alpha, v_1)} \langle \mathcal{T}^2, A(1)\rangle \xrightarrow{(\alpha, v_0)} \langle \mathcal{T}^2, A(0)\rangle$$

$$\langle \mathcal{T}^3, A(3)\rangle \xrightarrow{(\alpha, u_2)} \langle \mathcal{T}^3, A(2)\rangle \xrightarrow{(\alpha, u_1)} \langle \mathcal{T}^3, A(1)\rangle \xrightarrow{(\alpha, u_0)} \langle \mathcal{T}^3, A(0)\rangle$$

Fig. 4. Example of discretisations that are not bisimilar

3.1 Compression Bisimulation

We now define the new equivalence. As noted in the introduction, our approach here is to consider the systems we want to be equivalent and to work from there. We want our equivalence to be a congruence and we also want to equate discretisations with sufficiently large numbers of levels because this is our starting point. However, having said that, we are still interested in an equivalence that is similar to classical equivalences such as bisimulation. Note that we cannot use bisimulation directly here. This can be shown by the species component $A \stackrel{def}{=} (\alpha, 1){\uparrow}A$. Figure 4 gives the transition system for two different discretisation, one where the maximum level is 2 and the other where it is 3. We can relate $\langle \mathcal{T}^2, A(i)\rangle$ and $\langle \mathcal{T}^3, A(i)\rangle$ for $0 \leq i \leq 2$ but we cannot relate $\langle \mathcal{T}^3, A(3)\rangle$ to any of $\langle \mathcal{T}^2, A(i)\rangle$.

However, although we cannot use bisimulation directly, we are able to use it indirectly over equivalence classes and achieve the goals of congruence and equating discretisations. We now present definitions that allow us to achieve that.

We first need to define the equivalence relation that will define the relevant equivalence classes. The current level together with the stoichiometry associated with a reaction determine which reactions can occur, therefore we are interested in grouping together those states of the Bio-PEPA system for which the same reactions can take place. The collection of enabled reactions becomes our underlying notion of behaviour. This captures the similarities that we see between different discretisations. From a biological point of view, this is sensible because it is an observational notion of equivalence.

In light of this, we can define an equivalence relation over Bio-PEPA systems that depends on $\mathcal{A}$ which defines the actions that are currently enabled. Two processes are related if their current action sets are the same.

Definition 7. *The* current action relation $\mathcal{H}$ *over well-defined Bio-PEPA systems is defined as* $\mathcal{H} = \{(\mathcal{P}_1, \mathcal{P}_2) \mid \mathcal{A}(\mathcal{P}_1) = \mathcal{A}(\mathcal{P}_2)\}$.

Proposition 1. $\mathcal{H}$ *is an equivalence relation.*

Because $\mathcal{H}$ is an equivalence relation, it defines equivalence classes of Bio-PEPA systems which have the same current actions. For a set of Bio-PEPA systems $\mathcal{X}$, the equivalence classes of $\mathcal{X}$ with respect to $\mathcal{H}$ is denoted $\mathcal{X}/\mathcal{H}$. $\mathcal{A}$ can be extended to the equivalence classes in the obvious manner. Hence, for $P \in H$ an equivalence class, $\mathcal{A}(H) = \mathcal{A}(\mathcal{P})$.

We are interested in considering the equivalence classes over the derivative set of a given Bio-PEPA system $\mathcal{P}$ because we want to consider the overall behaviour of individual Bio-PEPA systems and we define $\mathcal{P}_{\mathcal{H}} = ds(\mathcal{P})/\mathcal{H}$. Since we want to define a bisimulation-style equivalence we need to define transitions between equivalence classes. The basic idea is that if there is a transition between individual members of two equivalence classes then there is a transition between those equivalence classes.

Definition 8. *For $H, H' \in \mathcal{P}_{\mathcal{H}}$, $H \overset{\alpha}{\hookrightarrow} H'$ if there exists $\mathcal{P} \in H$ and $\mathcal{P}' \in H'$ such that $\mathcal{P} \xrightarrow{(\alpha,w)}_{sc} \mathcal{P}'$.*

We can then finalise the definition for our new equivalence as follows. We use Definition 6 for the definition of $\sim$ but substitute $\overset{\alpha}{\hookrightarrow}$ for all instances of $\overset{\theta}{\rightarrow}$ and, moreover the relation $\sim$ is defined between equivalence classes.

Definition 9. *$\mathcal{P}$ and $\mathcal{Q}$ are compression bisimilar, $\mathcal{P} \simeq \mathcal{Q}$, if $\mathcal{P}_{\mathcal{H}} \sim \mathcal{Q}_{\mathcal{H}}$.*

3.2 Equivalence and Congruence Results

This section details results about the new equivalence. First we need show that it is an equivalence relation.

Proposition 2. *$\mathcal{P} \simeq \mathcal{Q}$ is an equivalence relation.*

Proof. This is straightforward because $\sim$ is an equivalence relation. $\square$

Next, we consider the sequential case of two discretisations and show that they are equated by the new equivalence. The sequential case consists of considering a single species and two discretisations. The first theorem of the paper shows that given a single species component and two discretisations, then the two discretisations are compression bisimilar because their induced equivalence classes are bisimilar (in fact, they are isomorphic). First, some notation and various lemmas are required. The complexity of these results is due to the fact that stoichiometry can be larger than one. For a sequential Bio-PEPA component $C \overset{def}{=} \sum_{i=1}^{m}(\alpha_i, \kappa_i) \, \mathrm{op}_i \, C$, let

$$T_\uparrow = \{\kappa_i \mid (\alpha_i, \kappa_i) \uparrow C \text{ appears in the definition of } C\} \qquad t_\uparrow = |T_\uparrow|$$
$$T_\downarrow = \{\kappa_i \mid (\alpha_i, \kappa_i) \downarrow C \text{ appears in the definition of } C\} \cup$$
$$\qquad \{1 \mid (\alpha_i, \kappa_i) \oplus C \text{ appears in the definition of } C\} \qquad t_\downarrow = |T_\downarrow|$$
$$\mathcal{A}_C = \{\alpha_i \mid (\alpha_i, \kappa_i) \, \mathrm{op}_i \, C \text{ appears in the definition of } C\}$$
$$k_m = \max\{k_\downarrow, k_\uparrow, 1\}, \ k_\downarrow = \max(T_\downarrow), \ k_\uparrow = \max(T_\uparrow), \text{ hence } k_m \geq k_\downarrow, k_m \geq k_\uparrow.$$

The diagram in Figure 5 illustrates the equivalence classes for two discretisations of $C \overset{def}{=} (\alpha, 2) \uparrow C + (\beta, 3) \uparrow C + (\gamma, 4) \downarrow C + (\delta, 1) \oplus C$ with $n = 11$ and $n' = 13$. It

$$0 \qquad\qquad k_\downarrow \qquad\qquad n{-}k_\uparrow \qquad\qquad n$$

$$[0,0] \qquad [1,3] \qquad [4,8]=[i_c,j_c] \qquad [9,9] \qquad [10,11]$$

$$0 \qquad\qquad k_\downarrow \qquad\qquad n'{-}k_\uparrow \qquad\qquad n'$$

$$[0,0] \qquad [1,3] \qquad [4,10]=[i'_c,j'_c] \qquad [11,11] \quad [12,13]$$
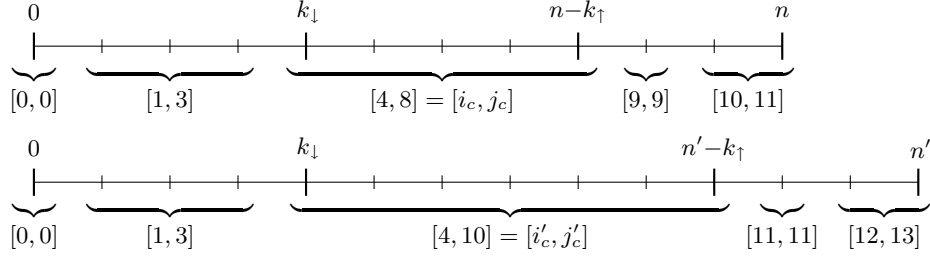
Fig. 5. The equivalence classes of two discretisations of a species component

also demonstrates how the various stoichiometry coefficients result in different equivalent classes. The lemmas that follow prove the properties of the equivalence classes as shown in this diagram.

The next lemma establishes that equivalence classes can be ordered which makes them easier to manipulate in later lemmas. The following lemma builds on this and shows that there are a fixed number of equivalence classes if there are sufficient levels and it contributes to the definition of the isomorphism in the first theorem.

Lemma 1. *For a sequential Bio-PEPA component $C \stackrel{\text{def}}{=} \sum_{i=1}^{m} (\alpha_i, \kappa_i)\ op_i\ C$ and the Bio-PEPA system $\mathcal{S}^n = \langle \mathcal{T}^n, C \rangle$, the equivalence classes of $\mathcal{S}^n_{\mathcal{H}}$ form a strict order.*

Proof. The set $ds(\mathcal{S}^n)$ contains elements of the form $\langle \mathcal{T}^n, C(l) \rangle$, where l ranges over $0, \ldots, n$. Each equivalence class is a subsequence of $0, \ldots, n$ because of the side conditions of the prefix rules. Therefore a class can be described by its smallest and largest elements $[i, j]$ for $i \leq j$. These intervals do not overlap because the equivalence classes form a partition. Hence for any two equivalence classes $[i, j]$ and $[i', j']$, either $j < i'$ or $j' < i$, and this property defines a strict order over the equivalence classes. $\qquad\square$

Lemma 2. *For a sequential Bio-PEPA component $C \stackrel{\text{def}}{=} \sum_{i=1}^{m} (\alpha_i, \kappa_i)\ op_i\ C$ and the Bio-PEPA system $\mathcal{S}^n = \langle \mathcal{T}^n, C \rangle$, if $n \geq k_\uparrow + k_\downarrow + 1$, then $\mathcal{S}^n_{\mathcal{H}}$ has $t_\uparrow + t_\downarrow + 1$ equivalence classes.*

Proof. By Lemma 1, a sequence of equivalence classes $[i_1, j_1], [i_2, j_2], \ldots, [i_t, j_t]$ partitioning $\mathcal{S}^n_{\mathcal{H}}$ exist. We show that there is an equivalence class $[i_c, j_c]$ with $\mathcal{A}([i_c, j_c]) = \mathcal{A}_C$, $i_c = k_\downarrow$ and $j_c = n - k_\uparrow$. This is well-defined since $n \geq k_\downarrow + k_\uparrow + 1$. Consider $l \in [i_c, j_c]$. Any production prefix $(\alpha, \kappa)\uparrow C$ is enabled since $0 \leq i_c \leq l \leq n - k_\uparrow \leq n - \kappa$. Any reactant prefix $(\alpha, \kappa)\downarrow C$ is enabled because $\kappa \leq k_\downarrow \leq l \leq n - k_\uparrow \leq n$. Prefixes containing $\ominus$ or $\odot$ can always generate transitions. Since $l \geq 1$, any prefix of the form $(\alpha, \kappa) \oplus C$ is enabled. Hence $\mathcal{A}([i_c, j_c]) = \mathcal{A}_C$ and the class cannot be larger.

Next we consider the equivalence classes that come before $[i_c, j_c]$. Order the elements of $T_\downarrow$ from smallest to largest, $\tau_1, \tau_2, \ldots, \tau_{t_\downarrow - 1}, \tau_{t_\downarrow}$ where $\tau_{t_\downarrow} = k_\downarrow$ then we have that $[i_1, j_1], [i_2, j_2], \ldots, [i_{c-1}, j_{c-1}]$ is $[0, \tau_1 - 1], [\tau_1, \tau_2 - 1], \ldots, [\tau_{t_\downarrow - 1}, \tau_{t_\downarrow} - 1]$ which gives $t_\downarrow$ equivalence classes.

Similarly, we have $t_\uparrow$ equivalence classes corresponding to the elements of $T_\uparrow$. This means that there are $t_\downarrow + t_\uparrow + 1$ equivalence classes in total. $\qquad\square$

Corollary 1. *Let $C \stackrel{\text{def}}{=} \sum_{i=1}^{m}(\alpha_i, \kappa_i) \; op_i \; C$ be a sequential Bio-PEPA component which has stoichiometry coefficient 1 for all reactant prefixes and product prefixes then any discretisation with $n \geq 3$ has three equivalence classes.*

The next lemma establishes a lower bound on the size of the equivalence class that is capable of performing all actions, namely the "central" class $[i_c, j_c]$ with $\mathcal{A}([i_c, j_c]) = \mathcal{A}_c$. This class is the only one that differs in cardinality for different discretisations, and it grows in size as the number of levels are increased. The other classes do not differ between different discretisations because they are defined by the same stoichiometry coefficients. The value k_m is used since knowing that k_m is a bound on the size of $[i_c, j_c]$ is important for a later lemma.

Lemma 3. *Let $C \stackrel{\text{def}}{=} \sum_{i=1}^{m}(\alpha_i, \kappa_i) \; op_i \; C$ be a sequential Bio-PEPA component and let $\mathcal{S}^n = \langle \mathcal{T}^n, C \rangle$ for $n \geq k_\uparrow + k_m + k_\downarrow$. Then $[i_c, j_c]$, the equivalence class of $\mathcal{S}^n / \mathcal{H}$ with $\mathcal{A}[i_c, j_c] = \mathcal{A}_C$, has cardinality greater than k_m.*

Proof. It is the case that $i_c = k_\downarrow$ and $j_c = n - k_\uparrow$. The cardinality of $[i_c, j_c]$ is $j_c - i_c + 1 = n - k_\uparrow - k_\downarrow + 1 \geq k_\uparrow + k_m + k_\downarrow - k_\uparrow - k_\downarrow + 1 = k_m + 1 > k_m$. □

This implies that the cardinality/width of $[i_c, j_c]$ is greater than both $k_\downarrow$ and $k_\uparrow$. The next lemma relates the equivalence classes obtained for two different values of n by expressing the classes for the larger value in terms of the intervals of the smaller value.

Lemma 4. *Let $C \stackrel{\text{def}}{=} \sum_{i=1}^{m}(\alpha_i, \kappa_i) \; op_i \; C$ be a sequential Bio-PEPA component and let $\mathcal{S} = \langle \mathcal{T}, C \rangle$. Let $n' = n + d$. Then the equivalence classes of $\mathcal{S}^n_\mathcal{H}$ are described by the ordered intervals $[0, j_1], \ldots, [i_c, j_c], \ldots, [i_{t-1}, j_{t-1}], [i_t, n]$ and the equivalence classes of $\mathcal{S}^{n'}_\mathcal{H}$ are described by the ordered intervals $[0, j_1], \ldots, [i_c, j_c + d], \ldots, [i_{t-1} + d, j_{t-1} + d], [i_t + d, n + d]$ where $[i_c, j_c]$ and $[i_c, j_c + d]$ are the equivalence classes in which all actions of C are possible.*

Proof. The elements of $\mathcal{S}^n_\mathcal{H}$ are $[0, j_1], \ldots, [i_c, j_c], \ldots, [i_{t-1}, j_{t-1}], [i_t, n]$ and those of $\mathcal{S}^{n'}_\mathcal{H}$ are $[0, j'_1], \ldots, [i'_c, j'_c], \ldots, [i'_{t-1}, j'_{t-1}], [i'_t, n']$. The first $t_\downarrow$ equivalence classes are the same in both cases because they are defined by the same stoichiometric coefficients so $\mathcal{S}^n_\mathcal{H}$ is $[0, j_1], \ldots, [i_c, j'_c], \ldots, [i'_{t-1}, j'_{t-1}], [i'_t, n']$. Since $j_c = n - k_\uparrow$ and $j'_c = n' - k_\uparrow$, $j'_c = j_c + d$ and this offset is the same throughout the remaining equivalence classes which are defined by the same stoichiometric coefficients hence $\mathcal{S}^n_\mathcal{H}$ can be written $[0, j_1], \ldots, [i_c, j_c + d], \ldots, [i_{t-1} + d, j_{t-1} + d], [i_t + d, n + d]$. □

Finally the most important lemma shows that the same transitions occur between equivalence classes if the numbers of levels are large enough. This contributes to the isomorphism defined in the theorem about sequential Bio-PEPA systems.

Lemma 5. *Let $C \stackrel{\text{def}}{=} \sum_{i=1}^{m}(\alpha_i, \kappa_i) \; op_i \; C$ be a sequential Bio-PEPA component and let $\mathcal{S} = \langle \mathcal{T}, C \rangle$. Let $E_1, \ldots, E_t$ be the ordered equivalence classes of $\mathcal{S}^n_\mathcal{H}$ and $E'_1, \ldots, E'_t$ be the ordered equivalence classes of $\mathcal{S}^{n'}_\mathcal{H}$. If $n \geq k_\downarrow + k_m + k_\uparrow$ then $E_p \stackrel{\alpha}{\longleftrightarrow} E_q$ if and only if $E'_p \stackrel{\alpha}{\longleftrightarrow} E'_q$.*

Proof. Let $n' = n + d$. Lemma 4 gives notation for the two sequences of equivalence classes and let $[i_c, j_c]$ and $[i'_c, j'_c]$ be the classes whose current action set is $\mathcal{A}_C$. Consider $[i_p, j_p] \overset{\alpha}{\hookrightarrow} [i_q, j_q]$. Since $\mathcal{A}([i_p, j_p]) = \mathcal{A}([i'_p, j'_p])$, $[i'_p, j'_p] \overset{\alpha}{\hookrightarrow} [i', j']$. We need to show that $[i', j'] = [i'_q, j'_q]$. There are two transitions $\langle \mathcal{T}^n, C(l) \rangle \xrightarrow{(\alpha, w)} \langle \mathcal{T}^n, C(l_1) \rangle$ with $l \in [i_p, j_p]$ and $l_1 \in [i_q, j_p]$ and $\langle \mathcal{T}^{n'}, C(l') \rangle \xrightarrow{(\alpha, w')} \langle \mathcal{T}^{n'}, C(l'_1) \rangle$ with $l' \in [i'_p, j'_p]$ and $l'_1 \in [i', j']$. So we need to show that $l'_1 \in [i'_q, j'_q]$. We also need to consider the other direction. If $[i'_p, j'_p] \overset{\alpha}{\hookrightarrow} [i'_q, j'_q]$ then $[i_p, j_p] \overset{\alpha}{\hookrightarrow} [i, j]$ and we need to show that $l \in [i_q, j_q]$.

Prefixes (α, κ) op C **where** op $\in \{\oplus, \ominus, \odot\}$**:** A transition does not result in a change of level therefore $p = q$ and $[i', j'] = [i'_p, j'_p]$. Similarly for the other direction.

We now consider the relative positions of p, q and c. The cases ($p < c$ and $q > c$) and ($q < c$ and $p > c$) cannot occur due to Lemma 3 as no transition can change the value of l or l' sufficiently.

Prefixes of the form $(\alpha, \kappa)\uparrow C$**:** Here the level l changes to $l + \kappa$ and $p \leq q$.

If $q < c$ then $l \in [i_p, j_p]$ and $l + \kappa \in [i_q, j_q]$ hence choosing $l' = l$ leads to $l' \in [i_p, j_p]$ and $l' + \kappa \in [i_q, j_q]$ as required. The other direction is similar.

For $q = c$ and $p < c$, let $l' = l$ then $l' + \kappa \in [i_c, j_c + d]$ since $l' + \kappa \in [i_c, j_c]$. For the other direction, $l' \in [i_p, j_p]$ and $l' + \kappa \in [i_c, j_c + d]$. We need to show that $l + \kappa \in [i_c, j_c]$. The largest value l' can take is $i_c - 1$. The width of $[i_c, j_c]$ is $j_c - i_c + 1 > k_m$ by Lemma 3. Therefore $l + \kappa = l' + \kappa \leq i_c - 1 + \kappa \leq i_c - 1 + k_m < i_c - 1 + j_c - i_c + 1 \leq j_c$.

If $p > c$, then $l \in [i_p, j_p]$ and $l + \kappa \in [i_q, j_q]$, Let $l' = l + d$ then $l' \in [i_p + d, j_p + d]$ and $l' + \kappa \in [i_q + d, j_q + d] = [i'_q, j'_q]$ as required. For the other direction, let $l = l' - d$.

If $p = c$ and $q = c$ let $l' = l$ then $l + \kappa \in [i_c, j_c]$ implies $l' + \kappa \in [i_c, j_c + d]$. For the other direction, we need to show that $l, l + \kappa \in [i_c, j_c]$. Choose $l = i_c$ and note that $l + \kappa = i_c + \kappa \leq i_c + k_m < j_c + 1$ by Lemma 3 therefore $l + \kappa \leq j_c$.

If $p = c$ and $q > c$ then $l \in [i_c, j_c]$ and $l + \kappa \in [i_q, j_q]$. To match this, choose $l' = l + d$. For the other direction, $l' \in [i_c, j_c + d]$ and $l' + \kappa \in [i_q + d, j_q + d]$. Let $l = l' - d$, then $l \in [i_c - d, j_c]$ and $l + \kappa \in [i_q, j_q]$. We need to show $l \geq i_c$. The smallest value that $l' + \kappa$ can take is $j_c + d + 1$ which implies $l + d + \kappa = l' + \kappa \geq j_c + d + 1$. So $l \geq j_c - \kappa + 1 \geq j_c - k_m + 1 > i_c$ by Lemma 3.

Prefixes of the form $(\alpha, \kappa)\downarrow C$**:** The level changes from l to $l - \kappa$ and $p \geq q$. These are similar to the previous case. □

The following theorem shows that for large enough value of n, two discretisations of a sequential Bio-PEPA system are compression bisimilar.

Theorem 1. *Let* $\mathcal{S} = \langle \mathcal{T}, C \rangle$ *be a well-defined Bio-PEPA system with the single species component* $C \overset{\text{def}}{=} \sum_{i=1}^{m} (\alpha_i, \kappa_i) \text{op}_i C$ *then* $\mathcal{S}^n \simeq \mathcal{S}^{n'}$ *for* $n, n' \geq k_\downarrow + k_m + k_\uparrow$.

Proof. Without loss of generality, assume that n is the maximum number of levels for species C in $\mathcal{S}^n$ and n' is the maximum number of levels for species C in $\mathcal{S}^{n'}$.

We will show that $\mathcal{S}^n_{\mathcal{H}}$ is isomorphic to $\mathcal{S}^{n'}_{\mathcal{H}}$ hence $\mathcal{S}^n_{\mathcal{H}} \sim \mathcal{S}^{n'}_{\mathcal{H}}$ and therefore $\mathcal{S}^n \simeq \mathcal{S}^{n'}$. Let $f : \mathcal{S}^n_{\mathcal{H}} \rightarrow \mathcal{S}^{n'}_{\mathcal{H}}$ be defined as $f(B) = D$ if $\mathcal{A}(B) = \mathcal{A}(D)$. This function is well-defined by Lemma 2 since $\mathcal{S}^n_{\mathcal{H}}$ and $\mathcal{S}^{n'}_{\mathcal{H}}$ have the same number of equivalence classes and hence for $B, B' \in \mathcal{S}^n_{\mathcal{H}}$, $f(B) = f(B')$ implies $B = B'$ and for any $D \in \mathcal{S}^{n'}_{\mathcal{H}}$, there exists $B \in \mathcal{S}^n_{\mathcal{H}}$ such that $f(B) = D$.

Additionally, define $f(B \xrightarrow{\alpha} B') = f(B) \xrightarrow{\alpha} f(B')$. This is a homomorphism because it preserves transitions. By Lemma 5, for any $D \xrightarrow{\alpha} D'$, there exist $B, B' \in \mathcal{S}_{\mathcal{H}}^n$ such that $f(B) \xrightarrow{\alpha} f(B')$. Hence f is a isomorphism. $\qquad\square$

Classically in congruence proofs, there would be a proof for each operator, hence there would be one for each of the prefix operators and then one for the choice operator. We do not need to show that the new semantic equivalence is a congruence with respect to the prefix operators and the choice operator since we work specifically with well-defined model components which give a constrained syntax that restricts how the prefix operators and the choice operator can be used.

We next consider a congruence result for the synchronisation operator. In this theorem, the notation $[\mathcal{P}]_{\mathcal{H}}$ refers to the equivalence class generated by $\mathcal{H}$, the current action relation, that contains the Bio-PEPA system $\mathcal{P}$. From this it is possible to obtain the result about compression bisimilarity between two different discretisations of a model component. First, we need a lemma and then a property describing the actions that are possible in a synchronisation. The property captures the idea that the actions in L are those that are shared by both components in the synchronisation.

Lemma 6. *Equality with respect to $\mathcal{A}$ is preserved by cooperation. In other words,*

$$\mathcal{A}(\langle \mathcal{T}, P_1 \rangle) = \mathcal{A}(\langle \mathcal{T}, P_2 \rangle) \quad \Rightarrow \quad \begin{cases} \mathcal{A}(\langle \mathcal{T}, P_1 \underset{L}{\bowtie} Q \rangle) = \mathcal{A}(\langle \mathcal{T}, P_2 \underset{L}{\bowtie} Q \rangle) & and \\ \mathcal{A}(\langle \mathcal{T}, Q \underset{L}{\bowtie} P_1 \rangle) = \mathcal{A}(\langle \mathcal{T}, Q \underset{L}{\bowtie} P_2 \rangle) \end{cases}$$

Definition 10. *Given a well-defined Bio-PEPA system of the form $\langle \mathcal{T}, P \underset{L}{\bowtie} Q \rangle$, it has the* current action decomposition property *if $\mathcal{A}(\langle \mathcal{T}, P_1 \underset{L}{\bowtie} Q_1 \rangle) = \mathcal{A}(\langle \mathcal{T}, P_2 \underset{L}{\bowtie} Q_2 \rangle)$ implies $\mathcal{A}(\langle \mathcal{T}, P_1 \rangle) = \mathcal{A}(\langle \mathcal{T}, P_2 \rangle)$ and $\mathcal{A}(\langle \mathcal{T}, Q_1 \rangle) = \mathcal{A}(\langle \mathcal{T}, Q_2 \rangle)$ for all systems $\langle \mathcal{T}, P_1 \underset{L}{\bowtie} Q_1 \rangle, \langle \mathcal{T}, P_2 \underset{L}{\bowtie} Q_2 \rangle \in ds(\langle \mathcal{T}, P \underset{L}{\bowtie} Q \rangle)$.*
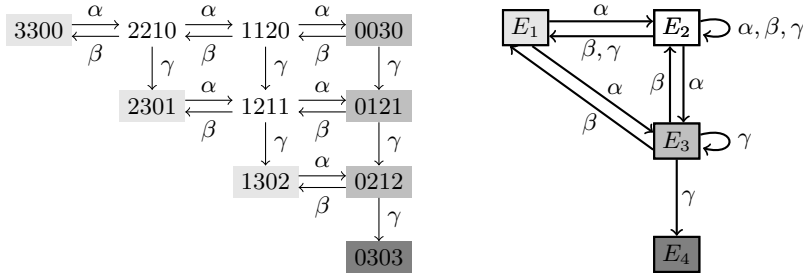
Theorem 2. *Let $\langle \mathcal{T}_1, P_1 \rangle$, $\langle \mathcal{T}_2, P_2 \rangle$, $\langle \mathcal{T}_1, Q_1 \rangle$ and $\langle \mathcal{T}_2, Q_2 \rangle$ be well-defined Bio-PEPA systems such that $\langle \mathcal{T}_1, P_1 \underset{L}{\bowtie} Q_1 \rangle$ and $\langle \mathcal{T}_2, P_2 \underset{L}{\bowtie} Q_2 \rangle$ both have the current action decomposition property. If $\langle \mathcal{T}_1, P_1 \rangle \simeq \langle \mathcal{T}_2, P_2 \rangle$ and $\langle \mathcal{T}_1, Q_1 \rangle \simeq \langle \mathcal{T}_2, Q_2 \rangle$ then it is the case that $\langle \mathcal{T}_1, P_1 \underset{L}{\bowtie} Q_1 \rangle \simeq \langle \mathcal{T}_2, P_2 \underset{L}{\bowtie} Q_2 \rangle$.*

Proof. Let $\mathcal{R} = \big\{ \big([\langle \mathcal{T}_1, P_1' \underset{L}{\bowtie} Q_1' \rangle]_{\mathcal{H}}, [\langle \mathcal{T}_2, P_2' \underset{L}{\bowtie} Q_2' \rangle]_{\mathcal{H}} \big) \mid [\langle \mathcal{T}_1, P_1' \rangle]_{\mathcal{H}} \sim [\langle \mathcal{T}_2, P_2' \rangle]_{\mathcal{H}}$ and $[\langle \mathcal{T}_1, Q_1' \rangle]_{\mathcal{H}} \sim [\langle \mathcal{T}_2, Q_2' \rangle]_{\mathcal{H}} \big\}$. We want to show that $\mathcal{R}$ is a bisimulation. This proof follows the standard technique for this case but requires a few additional steps. We only consider the case of $\alpha \in L$. The other two cases are simpler.

Consider a transition from $[\langle \mathcal{T}_1, P_1' \underset{L}{\bowtie} Q_1' \rangle]_{\mathcal{H}}$. This may be generated by a transition from another system in this class, say $\langle \mathcal{T}_1, P_3' \underset{L}{\bowtie} Q_3' \rangle \xrightarrow{(\alpha, w)}_{sc} \langle \mathcal{T}_1, P_3'' \underset{L}{\bowtie} Q_3'' \rangle$. By shorter inferences and by the definition of transitions between equivalence classes we can infer $[\langle \mathcal{T}_1, P_3' \rangle]_{\mathcal{H}} \xrightarrow{\alpha} [\langle \mathcal{T}_1, P_3'' \rangle]_{\mathcal{H}}$ and $[\langle \mathcal{T}_1, Q_3' \rangle]_{\mathcal{H}} \xrightarrow{\alpha} [\langle \mathcal{T}_1, Q_3'' \rangle]_{\mathcal{H}}$.

From the current action decomposition property, we know that these transitions are $[\langle \mathcal{T}_1, P_1' \rangle]_{\mathcal{H}} \xrightarrow{\alpha} [\langle \mathcal{T}_1, P_3'' \rangle]_{\mathcal{H}}$ and $[\langle \mathcal{T}_1, Q_1' \rangle]_{\mathcal{H}} \xrightarrow{\alpha} [\langle \mathcal{T}_1, Q_3'' \rangle]_{\mathcal{H}}$.

By the definition of $\mathcal{R}$, there exist P_2'' and Q_2'' such that $[\langle \mathcal{T}_2, P_2' \rangle]_{\mathcal{H}} \xrightarrow{\alpha} [\langle \mathcal{T}_2, P_2'' \rangle]_{\mathcal{H}}$ and $[\langle \mathcal{T}_2, Q_2' \rangle]_{\mathcal{H}} \xrightarrow{\alpha} [\langle \mathcal{T}_2, Q_2'' \rangle]_{\mathcal{H}}$ such that $[\langle \mathcal{T}_1, P_3'' \rangle]_{\mathcal{H}} \sim [\langle \mathcal{T}_2, P_2'' \rangle]_{\mathcal{H}}$ and $[\langle \mathcal{T}_1, Q_3'' \rangle]_{\mathcal{H}} \sim [\langle \mathcal{T}_2, Q_2'' \rangle]_{\mathcal{H}}$.

Fig. 6. Transition system for MC when $n = 3$ (left) and associated equivalence classes (right).

We can then infer $\langle \mathcal{T}_2, P_4' \rangle \xrightarrow{(\alpha,w_2)}_{sc} \langle \mathcal{T}_2, P_4'' \rangle$ with $\mathcal{A}(\langle \mathcal{T}_2, P_4' \rangle) = \mathcal{A}(\langle \mathcal{T}_2, P_2' \rangle)$ and $\mathcal{A}(\langle \mathcal{T}_2, P_4'' \rangle) = \mathcal{A}(\langle \mathcal{T}_2, P_2'' \rangle)$ and $\langle \mathcal{T}_2, Q_2' \rangle \xrightarrow{(\alpha,v_2)}_{sc} \langle \mathcal{T}_2, Q_2'' \rangle$ with $\mathcal{A}(\langle \mathcal{T}_2, Q_4' \rangle) = \mathcal{A}(\langle \mathcal{T}_2, Q_2' \rangle)$ and $\mathcal{A}(\langle \mathcal{T}_2, Q_4'' \rangle) = \mathcal{A}(\langle \mathcal{T}_2, Q_2'' \rangle)$.

By Lemma 6 and the above equalities, $\mathcal{A}(\langle \mathcal{T}_2, P_4' \bowtie_L Q_4' \rangle) = \mathcal{A}(\langle \mathcal{T}_2, P_2' \bowtie_L Q_2' \rangle)$ and $\mathcal{A}(\langle \mathcal{T}_2, P_4'' \bowtie_L Q_4'' \rangle) = \mathcal{A}(\langle \mathcal{T}_2, P_2'' \bowtie_L Q_2'' \rangle)$.

We can infer $\langle \mathcal{T}_2, P_4' \bowtie_L Q_4' \rangle \xrightarrow{(\alpha,w)}_{sc} \langle \mathcal{T}_2, P_4'' \bowtie_L Q_4'' \rangle$ which leads to the matching transition $[\langle \mathcal{T}_2, P_2' \bowtie_L Q_2' \rangle]_{\mathcal{H}} \xrightarrow{\alpha} [\langle \mathcal{T}_2, P_2'' \bowtie_L Q_2'' \rangle]_{\mathcal{H}}$ and by definition we know that $\left([\langle \mathcal{T}_1, P_3'' \bowtie_L Q_3'' \rangle]_{\mathcal{H}}, [\langle \mathcal{T}_2, P_2'' \bowtie_L Q_2'' \rangle]_{\mathcal{H}}\right) \in \mathcal{R}$. $\qquad\square$

Corollary 2. *Let $\mathcal{P} = \langle \mathcal{T}, P \rangle$ be a well-defined Bio-PEPA system then $\mathcal{P}^n \simeq \mathcal{P}^{n'}$ for $n, n' \geq k_\downarrow + k_m + k_\uparrow$ if the current action decomposition property applies to pairs of subcomponents of $\mathcal{P}^n$ and $\mathcal{P}^{n'}$.*

Proof. Since $\mathcal{P}$ is well-defined, $P \stackrel{def}{=} C_1(l_1) \bowtie_{\mathcal{L}_1} \ldots \bowtie_{\mathcal{L}_{m-1}} C_m(l_m)$ with each C_i a sequential component. By Theorem 1 and $n, n' > k_\downarrow + k_m + k_\uparrow$, $\langle \mathcal{T}^n, C_i \rangle \simeq \langle \mathcal{T}^{n'}, C_i \rangle$ for all i. By repeated applications of Theorem 2, $\langle \mathcal{T}^n, P \rangle \simeq \langle \mathcal{T}^{n'}, P' \rangle$. $\qquad\square$

4 Example

We give an example of discretisations and the associated equivalence classes. Consider the substrate-enzyme-product reactions $S + E \leftrightarrows SE \to P + E$ which can be expressed in Bio-PEPA as

$$S \stackrel{def}{=} (\alpha, 1) \downarrow S + (\beta, 1) \uparrow S \qquad E \stackrel{def}{=} (\alpha, 1) \downarrow E + (\beta, 1) \uparrow E + (\gamma, 1) \uparrow E$$

$$SE \stackrel{def}{=} (\alpha, 1) \uparrow SE + (\beta, 1) \downarrow SE + (\gamma, 1) \downarrow SE \qquad P \stackrel{def}{=} (\gamma, 1) \uparrow P$$

$$MC \stackrel{def}{=} S(x) \bowtie_{\{\alpha,\beta\}} E(x) \bowtie_{\{\alpha,\beta,\gamma\}} SE(0) \bowtie_{\{\gamma\}} P(0)$$

Figure 6 gives the transition system for MC when the maximum number of levels for all species is three as well as that for the equivalence classes. In the transition system for MC each state is a Bio-PEPA system and is indicated by its vector representation which describes the level of each species in that system using the vector (S, E, SE, P) but we omit the brackets and commas in the diagram. Additionally, we have omitted the strings w from the diagram for reasons of space. The shading shows the different equivalence
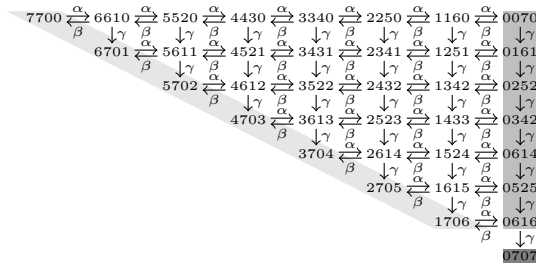
$$7700 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 6610 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 5520 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 4430 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 3340 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 2250 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 1160 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 0070$$

$$\downarrow\gamma \quad \downarrow\gamma \quad \downarrow\gamma \quad \downarrow\gamma \quad \downarrow\gamma \quad \downarrow\gamma \quad \downarrow\gamma$$

$$6701 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 5611 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 4521 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 3431 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 2341 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 1251 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 0161$$

$$\downarrow\gamma \quad \downarrow\gamma \quad \downarrow\gamma \quad \downarrow\gamma \quad \downarrow\gamma \quad \downarrow\gamma$$

$$5702 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 4612 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 3522 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 2432 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 1342 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 0252$$

$$\downarrow\gamma \quad \downarrow\gamma \quad \downarrow\gamma \quad \downarrow\gamma \quad \downarrow\gamma$$

$$4703 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 3613 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 2523 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 1433 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 0342$$

$$\downarrow\gamma \quad \downarrow\gamma \quad \downarrow\gamma \quad \downarrow\gamma$$

$$3704 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 2614 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 1524 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 0614$$

$$\downarrow\gamma \quad \downarrow\gamma \quad \downarrow\gamma$$

$$2705 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 1615 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 0525$$

$$\downarrow\gamma \quad \downarrow\gamma$$

$$1706 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} 0616$$

$$\downarrow\gamma$$

$$0707$$

Fig. 7. The transition system for MC when $n = 7$

classes in both diagrams. Figure 7 gives the system when the maximum number of levels is seven and the equivalence classes are also shown. This demonstrates how the two discretisations are related by the equivalence classes given in Figure 6.

5 Related work

The use of process algebras for modelling systems biology has multiplied rapidly since the first paper advocated the use of the π-calculus [15]. Approaches include the κ-calculus [2], stochastic π-calculus [3, 1], Beta-binders [4] and Bio-Ambients [5]. Most of these approaches use stochastic simulation as their analysis tool, and few approaches have considered the use of semantic equivalences. Weak bisimulation is shown to be a congruence for the bio-κ-calculus as is a context bisimulation which allows for the modelling of cell interaction [16]. Observational equivalence has been used to show that CCS specifications of elements of lactose operon regulation have the same behaviour as more detailed models [17]. In an example of biological modelling using hybrid systems, bisimulation is used to quotient the state space with respect to a subset of variables as a technique for state space reduction [18]. Bisimulation has also been used in the comparison of ambient-style models and membrane-style models [19] and the comparison of a term-rewriting calculus and a simple brane calculus [20].

6 Conclusions and Further Research

This paper has presented a new semantic equivalence for Bio-PEPA called compression bisimulation and shown that it is a congruence and it identifies different discretisations of the same system.

A first step for further work is to find a syntactic characterisation of the Bio-PEPA systems that exhibit the current action decomposition property so that those systems with this property can be easily identified. Since reactions have unique names in Bio-PEPA, an extension that would be useful is to allow for a relation over names to relate different reactions in different models. We would also like to extend the equivalence to be quantitative and take into account reaction rates but it is not immediately obvious how to do this. Finally, we wish to apply the equivalence to various biological models and to other formalisms using discretisation.

References

1. Blossey, R., Cardelli, L., Phillips, A.: A compositional approach to the stochastic dynamics of gene networks. In Priami, C., Cardelli, L., Emmott, S., eds.: Transactions on Computational Systems Biology IV. LNCS 3939, Springer (2006) 99–122
2. Danos, V., Laneve, C.: Formal molecular biology. Theoretical Computer Science **325** (2004) 69–110
3. Priami, C., Regev, A., Shapiro, E., Silverman, W.: Application of a stochastic name-passing calculus to representation and simulation of molecular processes. Information Processing Letters **80** (2001) 25–31
4. Priami, C., Quaglia, P.: Beta binders for biological interactions. In Danos, V., Schächter, V., eds.: Computational Methods in Systems Biology, International Conference (CMSB 2004), *Paris, France*. LNCS 3082, Springer (2004) 20–33
5. Regev, A., Panina, E., Silverman, W., Cardelli, L., Shapiro, E.: BioAmbients: an abstraction for biological compartments. Theoretical Computer Science **325** (2004) 141–167
6. Ciocchetta, F., Hillston, J.: Bio-PEPA: a framework for the modelling and analysis of biological systems. Theoretical Computer Science, in press.
7. Hillston, J.: A compositional approach to performance modelling. CUP (1996)
8. Ciocchetta, F., Hillston, J.: Calculi for biological systems. In: Formal Methods for Computational Systems Biology, SFM08. LNCS 5016 (2008) 265–312
9. Ciocchetta, F.: Bio-PEPA with SBML-like events. In Back, R.J., Petre, I., eds.: Proceedings of the Workshop on Computational Models for Cell Processes. (2008) 11–22
10. Ciocchetta, F., Hillston, J.: Bio-PEPA: a framework for the modelling and analysis of biological systems. Technical Report EDI-INF-RR-1231, School of Informatics, University of Edinburgh (2008)
11. Ellavarason, K.: An automatic mapping from the Systems Biology Markup Language to the Bio-PEPA process algebra. Master's thesis, University of Trento (2008)
12. Akman, O.E., Ciocchetta, F., Degasperi, A., Guerriero, M.L.: Modelling biological clocks with Bio-PEPA: stochasticity and robustness for the Neurospora Crassa circadian network, to appear in Proceedings of CMSB 2009.
13. Guerriero, M.L.: Qualitative and quantitative analysis of a Bio-PEPA model of the gp130/JAK/STAT signalling pathway, to appear in TCSB.
14. Milner, R.: Communication and concurrency. Prentice Hall (1989)
15. Regev, A., Shapiro, E.: Cellular abstractions: Cells as computation. Nature **419** (2002) 343
16. Laneve, C., Tarissan, F.: A simple calculus for proteins and cells. Theoretical Computer Science **404** (2008) 127–141
17. Pinto, M.C., Foss, L., Mombach, J.C.M., Ribeiro, L.: Modelling, property verification and behavioural equivalence of lactose operon regulation. Computers in Biology and Medicine **37** (2007) 134–148
18. Antoniotti, M., Piazza, C., Policriti, A., Simeoni, M., Mishra, B.: Taming the complexity of biochemical models through bisimulation and collapsing: theory and practice. Theoretical Computer Science **325** (2004) 45–67
19. Ciobanu, G., Aman, B.: On the relationship between membranes and ambients. BioSystems (91) (2008) 515–530
20. Barbuti, R., Maggiolo-Schettini, A., Milazzo, P., Troina, A.: Bisimulation in calculi modelling membranes. Formal Aspects of Computing **20** (2008) 351–377