A semantic equivalence for Bio-PEPA based on discretisation of continuous values

Vashti Galpin[*,a], Jane Hillston[a,b]

[a] *Laboratory for Foundations of Computer Science, School of Informatics*
University of Edinburgh, United Kingdom
[b] *Centre for Systems Biology (CSBE), University of Edinburgh*

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 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 number of levels, by using the idea of equivalence from concurrency and process algebra. We present a novel behavioural semantic equivalence, compression bisimilarity, and investigate when this equates two discretisations of the same model and the circumstances in which this equivalence is a congruence with respect to the synchronisation operator.

Key words: process algebra, biological modelling, discretisation, semantic equivalence, compression bisimulation, congruence, stoichiometry

[*]Corresponding author
Email addresses: Vashti.Galpin@ed.ac.uk (Vashti Galpin), Jane.Hillston@ed.ac.uk (Jane Hillston)

1. Introduction

The use of process algebras for modelling biological systems has become a popular technique [1, 2, 3, 4, 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 describe and analyse 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 intervals or levels. Typically, there is a fixed step size for the whole system and each species has a maximum level 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 enable a new type of analysis based on behaviour that can be used to compare models. Bio-PEPA does not currently consider hybrid approaches combining discrete and continuous representations as some other process algebras do [14, 15] but this is a possible topic for future work.

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. These two approaches

can be understood as two observers choosing different information to observe. In this paper we consider a qualitative equivalence. Qualitative equivalences are useful because they allow us to understand the structure of the model, and as will be shown in this paper, to determine how large a model needs to be to show all behaviour which is possible in the model. 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 makes the analysis of the system easier. In applying the idea of a semantic equivalence in systems biology, similar advantages will be gained in modelling hence the importance and timeliness of this research. Biologically, congruence can be viewed as the situation where two collections of species have the same reactions, regardless of the environment or medium into which they are placed. 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. These approaches are not mutually exclusive and we are able to successfully combine them here.

We have an 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.

The subtlety in finding such an equivalence lies in consideration of the different roles which processes may play within a reaction (e.g. reactant, modifier, product) and the stoichiometries which dictate a degree of involvement for each process. These characteristics of the reactions interact with the finite levels to place constraints on the possible behaviours which a process may exhibit at a particular level.

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, bisimilarity, to define our semantic equivalence. Hence we combine an existing notion of equivalence with a basic idea of having the same reaction capabilities which is the feature that characterises two abstractions of the same model.

This new semantic equivalence, compression bisimilarity, 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 a form of 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 bisimilarity. The second shows that compression bisimilarity is a congruence with respect to the cooperation operator under certain conditions. The third result is that in the general case of a Bio-PEPA system, namely with multiple species, two discretisations are related by compression bisimilarity under the same conditions. This paper is an extension and revision of the paper which appeared in CMSB'09 [16]. In particular we have replaced the previous *current action decomposition property* (CADP) with the more natural *matching derivative* (MD) property and a notion of *compatibility*. Furthermore the class of systems we considered has been generalised by the introduction of an explicit minimum level rather

than the assumption that all species can be exhausted.

Our paper has the following structure. We first present Bio-PEPA with a more general presentation than in previous research, after which we consider the types of transition system based on states of integer vectors that we obtain from Bio-PEPA systems. In this section, we introduce a running example of discretisations. Next we define compression bisimilarity, and first show that two discretisations of the same species are compression bisimilar. This is followed by a general proof of congruence for the synchronisation operator and a result comparing two discretisations involving this same operator, together with examples that illustrate the need for various conditions. We then present a biological example, discuss the application of this research to other biological formalisms, describe related work and finish with suggestions for further research.

2. Bio-PEPA

This section presents an overview of Bio-PEPA [6]. The definitions presented here are slightly more general than the original definitions, in that the number of levels are bounded by a minimum and maximum value rather than ranging from zero to a maximum. We motivate this generalisation later in this section when we define a Bio-PEPA system with levels.

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, and parameters. A species can be viewed as a population of molecules. The species or sequential component can be viewed as a template for the behaviour of members of that population, which may include a choice between taking part in different reactions. Model components then shift the focus to the population level.

The syntax of a sequential component is defined as

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

In the prefix term $(\alpha, \kappa)\ \mathtt{op}\ S$, 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 $\mathtt{op}$ 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. Levels are obtained by using a fixed step size to divide up the the range of concentration into finite number of discrete values. 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

$$C \stackrel{def}{=} (\alpha_1, \kappa_1)\ \mathtt{op}_1\ C + \ldots + (\alpha_q, \kappa_q)\ \mathtt{op}_q\ C \quad \text{written as} \quad C \stackrel{def}{=} \sum_{i=1}^{q} (\alpha_i, \kappa_i)\ \mathtt{op}_i\ C$$

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. This definition is more general that the original Bio-PEPA definition [6] since it includes a minimum concentration and minimum level.

$$\texttt{prefixReac} \quad \frac{}{(\alpha,\kappa)\downarrow S(l) \xrightarrow{(\alpha,[S:\downarrow(l,\kappa)])}_c S(l-\kappa)} \quad N'_S + \kappa \leq l \leq N_S$$

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

$$\texttt{prefixMod} \quad \frac{}{(\alpha,\kappa)\ \mathsf{op}\ S(l) \xrightarrow{(\alpha,[S:\mathsf{op}(l,\kappa)])}_c S(l)} \quad \begin{array}{l} N'_S + \kappa \leq l \leq N_S \text{ if } \mathsf{op} = \oplus \\ N'_S \leq l \leq N_S \text{ if } \mathsf{op} \in \{\ominus, \odot\} \end{array}$$

$$\texttt{choice1} \quad \frac{S_1(l) \xrightarrow{(\alpha,w)}_c S'_1(l')}{(S_1 + S_2)(l) \xrightarrow{(\alpha,w)}_c S'_1(l')}$$

$$\texttt{choice2} \quad \frac{S_2(l) \xrightarrow{(\alpha,w)}_c S'_2(l')}{S_1 + S_2(l) \xrightarrow{(\alpha,w)}_c S'_2(l')}$$

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

$$\texttt{coop1} \quad \frac{P_1 \xrightarrow{(\alpha,w)}_c P'_1}{P_1 \underset{L}{\bowtie} P_2 \xrightarrow{(\alpha,w)}_c P'_1 \underset{L}{\bowtie} P_2} \quad \alpha \notin L$$

$$\texttt{coop2} \quad \frac{P_2 \xrightarrow{(\alpha,w)}_c P'_2}{P_1 \underset{L}{\bowtie} P_2 \xrightarrow{(\alpha,w)}_c P_1 \underset{L}{\bowtie} P'_2} \quad \alpha \notin L$$

$$\texttt{coop3} \quad \frac{P_1 \xrightarrow{(\alpha,w_1)}_c P'_1 \quad P_2 \xrightarrow{(\alpha,w_2)} P'_2}{P_1 \underset{L}{\bowtie} P_2 \xrightarrow{(\alpha,w_1::w_2)}_c P'_1 \underset{L}{\bowtie} P'_2} \quad \alpha \in L$$

Figure 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 over $Comp$.

Elements of $\mathcal{N}$ have the form $C : H = h, N = n, M = m, N' = n', M' = m', V = v, unit = u$ where C is a species name that is defined in $Comp$, $H = h$ defines the step size, $N = n$ defines the maximum level for C, $M = m$ defines the maximum concentration for C, $N' = n'$ defines the minimum level for C, $M' = m'$ defines the

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

$$\texttt{Qual} \qquad \frac{P \xrightarrow{(\alpha,w)}_c P'}{\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, P\rangle \xrightarrow{\alpha} \langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, P'\rangle}$$

Figure 2: Operational semantics of Bio-PEPA (continued)

minimum concentration for C, $V = v$ names the compartment in which C appears and $unit = u$ defines the measurement unit of the concentration.

For details of the other elements of the context and a definition of well-defined Bio-PEPA system, see [6, 8]. The notation $\langle \mathcal{T}, P\rangle$ will be used for $\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, P\rangle$ when the details of the tuple are not relevant. For the rest of this paper, we work with well-defined Bio-PEPA models and systems.

This definition allows for a number of compartments in V but in this paper, we will assume one compartment only, and that there is a single step size that applies to all species. For a presentation of Bio-PEPA with compartments and membranes (together called locations) and the constraints imposed due to conservation of mass on the step size of a location by its size and the sizes of other locations, see [17][1].

The model component is defined in terms of concentrations, but can be expressed in terms of levels which discretise the concentration. We assume that the step size H is the same for all species to ensure conservation of mass.

Before the definition of a Bio-PEPA system with levels, we motivate using a minimum level instead of zero. We seek to develop as general a theory as possible which may be applicable to other models of interacting populations besides those arising in biochemistry. Having an arbitrary minimal level gives us that generality whilst retaining the more intuitive case, with zero as the lowest level, as a special case. Moreover, resticting population levels to a range of interest may help tackle the state space explosion problem by excluding uninteresting states and thus making the model more

[1]Note that in the Bio-PEPA Eclipse Plug-in (`www.biopepa.org`), step size is associated with location, not species [18].

amenable to analysis. Furthermore there may also be biologically based reasons for limiting what part of the state space to explore. For example, consider Michaelis-Menten kinetics where the rates are calculated under the assumption that there is a much higher concentration of substrate than of enzyme. By setting a minimum level for the substrate, it is possible to ensure that the transition system obtained is limited to that part where the assumption holds.

Definition 3. A *Bio-PEPA system with levels* uses a step size H to discretise the concentration into integral levels. The *maximum level* for species C is $N_C = \lceil M_C/H \rceil$ and the *minimum level* for species C is $N'_C = \lceil M'_C/H \rceil$, where M_C is the maximum concentration for C and M'_C is the minimum concentration for C. The *initial level* for species C in $\lceil x_C/H \rceil$ where x_C is the initial concentration for C.

A species at level k is an abstraction of a species having a concentration value somewhere in the interval $((k-1) \times H, k \times H]$. Thus although we take $\lceil M'_C/H \rceil$ as the minimum level, the minimum concentration M'_C is implicitly included. It is immediate that a species C with maximum level N_C and minimum level N'_C has $N_C - N'_C + 1$ levels.

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 level, and N'_S the minimum level for the species S. The side conditions in the first three rules are modified to work with an explicit minimum level. Note that we treat the minimum level as a boundary that cannot be crossed (just as is the case for zero) hence for the activator rule, we require that there be κ more of the species for the activation to be enabled. 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 equivalence in this paper is qualitative and only considers the action α, ignoring 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 `Qual` defines the relation $\rightarrow$ which we will

use in this paper because it focusses on the qualitative behaviour of systems, providing the context but only the reaction name on the transitions.

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

Definition 4. The derivative set ds($\mathcal{P}$) is the smallest set such that $\mathcal{P} \in ds(\mathcal{P})$ and if $\mathcal{P}' \in ds(\mathcal{P})$ and $\mathcal{P}' \xrightarrow{\alpha} \mathcal{P}''$ then $\mathcal{P}'' \in ds(\mathcal{P})$.

As will be discussed in the following section, the derivative set of a Bio-PEPA system with levels is finite. 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[2].

Definition 5. 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 level for species component S.

$$
\begin{aligned}
\mathcal{A}(((\alpha,\kappa) \downarrow S)(l)) &= \{\alpha\} \text{ if } N'_S + \kappa \leq l \leq N_S \text{ otherwise } \emptyset \\
\mathcal{A}(((\alpha,\kappa) \uparrow S)(l)) &= \{\alpha\} \text{ if } N'_S \leq l \leq N_S - \kappa \text{ otherwise } \emptyset \\
\mathcal{A}(((\alpha,\kappa) \oplus S)(l)) &= \{\alpha\} \text{ if } N'_S + \kappa \leq l \leq N_S \text{ otherwise } \emptyset \\
\mathcal{A}(((\alpha,\kappa) \ominus S)(l)) &= \{\alpha\} \\
\mathcal{A}(((\alpha,\kappa) \odot S)(l)) &= \{\alpha\} \\
\mathcal{A}((S_1 + S_2)(l)) &= \mathcal{A}(S_1(l)) \cup \mathcal{A}(S_2(l)) \\
\mathcal{A}(C(l)) &= \mathcal{A}(S(l)) \text{ where } C \stackrel{def}{=} S \\
\mathcal{A}(P_1 \underset{L}{\bowtie} P_2) &= \mathcal{A}(P_1) \setminus L \cup \mathcal{A}(P_2) \setminus L \cup (\mathcal{A}(P_1) \cap \mathcal{A}(P_2) \cap L)
\end{aligned}
$$

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.

Proposition 1. *For a Bio-PEPA model P, $\alpha \in \mathcal{A}(P)$ if and only if $P \xrightarrow{(\alpha,w)} P'$.*

[2]Note that this definition is somewhat different to that in [6].

PROOF. Straightforward from the definition of the set of current actions and the operational semantics.

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

Definition 6. 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 $\lambda \in \mathbb{N}$, the *Bio-PEPA system with levels* $\mathcal{P}^\lambda$ is defined as $\mathcal{P}^\lambda = \langle \mathcal{V}, \mathcal{N}', \mathcal{K}, \mathcal{F}, Comp, P' \rangle$ where

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

$\mathcal{N}$ contains information about each species. The definition above identifies the species with the smallest concentration range, determines the new step size that will ensure λ 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 $\mathcal{P}^\lambda = \langle \mathcal{T}^\lambda, P \rangle$ to indicate that the lowest number of levels for any species is λ and refer to $\mathcal{P}^\lambda$ as a discretisation of $\mathcal{P}$. Note that we only decorate $\mathcal{T}$ and not P since information about levels and step size are contained in $\mathcal{T}$ whereas the definition of P is independent of this information.

3. Transition systems with levels

We want to characterise and investigate the transition systems over which compression bisimilarity will be defined.

Let $\mathcal{P} = \langle \mathcal{T}, P \rangle$ be a well-defined Bio-PEPA system and consider the labelled transition system over $\rightarrow$ generated for $\mathcal{P}$ using the rules from Figure 1 and the rule `Qual` from Figure 2. Each state in this transition system is a Bio-PEPA system and has the form $\langle \mathcal{T}, P' \rangle$.

Moreover, P' only differs from P in the level of some species components. Therefore a state in the labelled transition system can be represented uniquely by a vector of levels, one for each species, for example $(x_1, \ldots, x_p)$[3].

Before a definition of these transition systems, we need some notation. Since the operational semantics are defined in terms of minimum and maximum levels, $\underline{x}$ and $\overline{x}$ are used to indicate these values respectively. To describe the vector representation of the labelled transition system obtained from a Bio-PEPA model P where the starting levels are $x_1, \ldots, x_p$, the minimum levels are $\underline{x}_1, \ldots, \underline{x}_p$ and the maximum levels are $\overline{x}_1, \ldots, \overline{x}_p$, all with respect to the species $C_1, \ldots, C_p$, we use the notation $P(x_1, \ldots, x_p\, [\underline{x}_1, \ldots, \underline{x}_p; \overline{x}_1, \ldots, \overline{x}_p])$ and require $\underline{x}_i \leq x_i \leq \overline{x}_i$ for all $1 \leq i \leq p$. Where P is clear, we may omit it.

If all the minimums are zero, namely $\underline{x}_i = 0$ for $1 \leq i \leq p$, then we write $P(x_1, \ldots, x_n\, [\overline{x}_1, \ldots, \overline{x}_n])$ or $(x_1, \ldots, x_n\, [\overline{x}_1, \ldots, \overline{x}_n])$. A vector $(y_1, \ldots, y_p)$ is then a state of the labelled transition system $P(x_1, \ldots, x_p\, [\underline{x}_1, \ldots, \underline{x}_p; \overline{x}_1, \ldots, \overline{x}_p])$ where for all $1 \leq i \leq p$, $\underline{x}_i \leq y_i \leq \overline{x}_i$.

We call a labelled transition system whose states are elements of $\mathbb{N}^p$ for a fixed $p \geq 1$ a transition system with levels. This is analogous to the definition of continuous time Markov chain with levels [19]. A state represents the current levels of the p species in the system and a transition represents a reaction involving some or all of these species, with the state after the transition representing the changed levels of the species as a result of the reaction.

[3] In contrast to Definition 6, x (together with y and z) will now be used to represent an integral level.

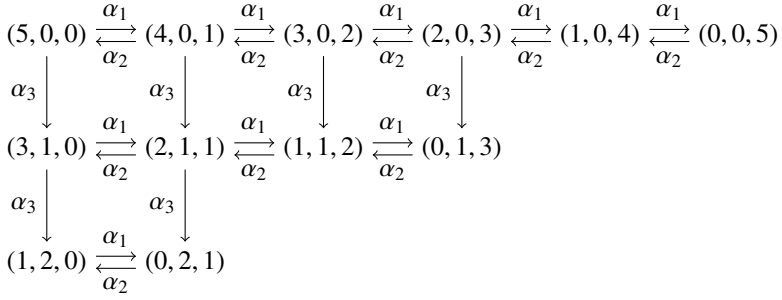

Figure 3: $P(5,0,0\,[5,5,5])$.

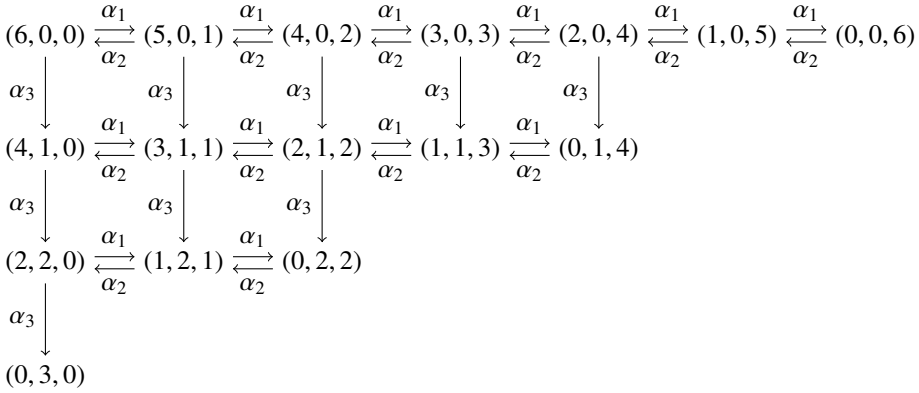

Figure 4: $P(6,0,0\,[6,6,6])$.

As an example, consider the following Bio-PEPA model

$$
\begin{aligned}
A &\stackrel{def}{=} (\alpha_1, 1) \downarrow A + (\alpha_2, 1) \uparrow A + (\alpha_3, 2) \downarrow A \\
B &\stackrel{def}{=} (\alpha_3, 1) \uparrow B \\
C &\stackrel{def}{=} (\alpha_1, 1) \uparrow C + (\alpha_2, 1) \downarrow C \\
P &\stackrel{def}{=} A(\ell_A) \underset{*}{\bowtie} (B(\ell_B) \underset{*}{\bowtie} C(\ell_C))
\end{aligned}
$$

The transition system $P(5,0,0\,[5,5,5])$ is presented in Figure 3 and $P(6,0,0\,[6,6,6])$ in Figure 4.

We wish to understand how these transition systems vary when the initial levels or

the bounds change. It is possible to have a Bio-PEPA system $\mathcal{P}$ with p species $C_1, \ldots C_p$ such that $\mathcal{P}(x_1, \ldots, x_p\,[\underline{x}_1, \ldots, \underline{x}_p; \overline{x}_1, \ldots, \overline{x}_p])$ and $\mathcal{P}(z_1, \ldots, z_p\,[\underline{z}_1, \ldots, \underline{z}_p; \overline{z}_1, \ldots, \overline{z}_p])$ are identical for $x_j = z_j$ for all $1 \leq i \leq p$ but $\underline{x}_j \neq \underline{z}_j$ and $\overline{x}_j \neq \overline{z}_j$ for all $1 \leq i \leq p$. This captures the idea that some of the minimums and maximums are not in effect constraining the shape of the transition system. For the above example, $P(5, 0, 0\,[5, 5, 5])$ and $P(5, 0, 0\,[6, 6, 6])$ are isomorphic. We now elucidate on this theme.

We start by considering the shape of transition systems with levels in general. Then we consider properties of individual species and finally interactions between species. Using the notation we have defined, our aim is to consider bounds on levels in an essentially syntactic manner where we only consider information obtained directly from minimum or maximum levels, starting levels and stoichiometry. Finally, we consider bounds in a more semantic fashion by considering the dynamics of systems with respect to interaction of species. We consider how to make the range of levels as small as possible but still large enough to demonstrate behaviour that is not pathological.

Before beginning this work, it is interesting to consider levels of generality. On the one hand, we have a Bio-PEPA system $\langle \mathcal{T}, P \rangle$ where $\mathcal{T}$ contains information about minimum, maximum and starting levels and which determines the exact behaviour of the system for each species and hence is least general and most specific of those that we consider here. We will demonstrate a canonical form of a Bio-PEPA system using the roles of species – this can be viewed as more general because many Bio-PEPA systems have the same canonical form. On the other hand, we have a Bio-PEPA model P which has no information about levels, so this can be viewed as multiple transition systems for varying level values and as most general in this context. More interestingly, at least theoretically, is a Bio-PEPA model with only starting levels specified. Similarly to the most general model, this can be viewed as multiple transition systems, each with different minimum and maximum levels. These systems are of interest as they allow us to understand which minimum and maximum levels will give a sufficiently large system that can demonstrate all possible behaviours of the system.

In the sequel, we focus on equality between transition systems (in terms of isomorphism). This allows us to say when one Bio-PEPA system is the same as another, in terms of states, transition structure and transition labels. First, we observe that tran-

sitions systems with levels obtained from a Bio-PEPA system are finite state. This is because a species has a maximum and minimum level, and hence a finite number of levels. Moreover, there are a finite number of species in a Bio-PEPA systems, hence there are only a finite number of states.

For the rest of this paper, the term "transition system" will imply "transition system with levels obtained from a Bio-PEPA system or model". We require a definition of isomorphism between transition systems to capture the idea that two transition systems have the same structure. This definition is standard. Note that the transition label must remain unchanged by the function. The transition systems in Figures 3 and 4 are not isomorphic.

Definition 7. Given T_1 and T_2 transition systems, an *isomorphism* $f : T_1 \to T_2$ is a bijective function between states and between transitions such that for $s, s' \in T_1$,

$$f(s \xrightarrow{\alpha} s') = f(s) \xrightarrow{\alpha} f(s')$$

3.1. Constraints on levels for individual species

Now we consider the roles that individual species play in a Bio-PEPA system. The following mutually exclusive classification captures how the level of a species may change over the transition system.

Definition 8. A species C can be categorised in the following way.

- C is *non-decreasing* if the prefix $\downarrow$ and the prefix $\oplus$ do not occur in the definition of C.
- C is *non-increasing* if the prefix $\uparrow$ does not occur in the definition of C.
- C is *static* if it is both non-decreasing and non-increasing.
- C is *dynamic* if neither non-decreasing nor non-increasing.

We can then use these classifications to determine some bounds on levels.

Definition 9. The transition system $T = P(x_1, \ldots, x_p\, [\underline{x}_1, \ldots, \underline{x}_p; \overline{x}_1, \ldots, \overline{x}_p])$ is *canonical* if for $1 \leq i \leq p$

$$\begin{aligned} \underline{x}_i &= x_i \quad \text{if } C_i \text{ non-decreasing} \\ \overline{x}_i &= x_i \quad \text{if } C_i \text{ non-increasing} \end{aligned}$$

Definition 10. Given a transition system $T = P(x_1, \ldots, x_p\, [\underline{x}_1, \ldots, \underline{x}_p; \overline{x}_1, \ldots, \overline{x}_p])$ then its *canonical form* is $P(x_1, \ldots, x_p\, [\underline{z}_1, \ldots, \underline{z}_p; \overline{z}_1, \ldots, \overline{z}_p])$ where

$$\underline{z}_i = \begin{cases} x_i & C_i \text{ non-decreasing} \\ \underline{x}_i & \text{otherwise} \end{cases} \qquad \overline{z}_i = \begin{cases} x_i & C_i \text{ non-increasing} \\ \overline{x}_i & \text{otherwise} \end{cases}$$

We then get the following result.

Proposition 2. *A transition system is isomorphic to its canonical form.*

PROOF. The stoichiometry of a species is fixed by its definition, hence its classification cannot change. Consider a species whose bounds vary between its original form and canonical form, and assume it is non-decreasing so that its minimum level is set to its initial level in the canonical form. It is not possible for an interaction with another species to cause its level to decrease. A similar argument can be made for non-increasing species. The maximum and minimum levels for a dynamic species remain unchanged in the canonical form. □

Note that interaction with other species may reduce the range of levels for a species but interaction cannot increase this range. For example, species A may be non-decreasing and have available to it a range of levels from its initial level to $\overline{x}_A$. However, when put in cooperation with species C and sharing a reaction in which A is the product and C is the reactant, it may be the case that A never reaches its maximum level $\overline{x}_A$ because there is insufficient C as a result of C reaching its minimum level. The next section considers the dynamics of systems with more than one species.

3.2. Constraints on levels for multiple species

An important concept that will be used in the rest of the paper relates to whether a transition system has sufficient size in each dimension, where a dimension represents

a species, and size is the number of levels for that species. In the previous section, we looked at reducing minimum and maximum levels. In this section, we investigate making them large enough.

A starting point for a definition of full behaviour is that there must exist at least one state from which every reaction is possible but as we will see later, we want a broader definition that asserts that for any given reaction, it must be possible to go from one state in which every reaction is possible to another state in which every reaction is possible.

To see why more reactions become possible as the number of levels increases, consider a single species with an odd number of levels and a central starting level. With three levels, reactions for which the species has stoichiometry 1 (as a product or a reactant) become possible, then at five levels any reactions with stoichiometry 2 are possible. Since the stoichiometry must be finite, there comes a point where all reactions, no matter what their stoichiometry are possible. After this, as levels are increased, all that changes is that the number of states in which all reactions are possible increases. When there are multiple species interacting, similar things happen as the number of levels increase but are constrained by the interaction of species.

For convenience, we make the assumption that our Bio-PEPA models are defined in such a way that for large enough systems, there are states where all reactions are possible[4]. Note that this is always the case for individual species.

Definition 11. A well-defined Bio-PEPA system with reactions $R = \{\alpha_1, \ldots, \alpha_r\}$ displays *full behaviour* if for all $\alpha_i \in R$, there exist states s_1 and s_2 with $\mathcal{A}(s_1) = \mathcal{A}(s_2) = R$ such that $s_1 \xrightarrow{\alpha_i} s_2$.

Later when we consider equivalence classes of systems, we will demonstrate the number of levels a species must have to show full behaviour and justify this definition.

Since we want to compare different transition systems, and since these are defined by their initial values, we also need to consider starting behaviour. Clearly, we can

[4]For models where there are no such states then it may be possible to generalise to maximal sets of reactions.

identify the reactions that are possible in a starting state, and we must ensure that we do not compare transition systems whose starting states have different sets of potential reactions. If we want to ensure that a starting state is a state from which all reactions are possible, then we need to ensure that for $P(x_1, \ldots, x_p\,[\underline{x}_1, \ldots, \underline{x}_p; \overline{x}_1, \ldots, \overline{x}_p])$ and for each species C_i, that $\underline{x}_i \leq x_i - k_\downarrow$ and $x_i + k_\uparrow \leq \overline{x}_i$ where $k_\downarrow$ is the maximum reactant stoichiometry for C_i and $k_\uparrow$ is the maximum product stoichiometry for C_i.

We can use the actual transition system generated to determine the level bounds. This ensures that the bounds are the smallest necessary for full behaviour.

Definition 12. A well-defined Bio-PEPA system $\langle \mathcal{T}, P \rangle$ is *compact* if for all $1 \leq i \leq p$, $\lambda_i = \overline{x_i} - \underline{x_i} + 1$ are the smallest values such that $P(x_1, \ldots, x_p\,[\underline{x}_1, \ldots, \underline{x}_p; \overline{x}_1, \ldots, \overline{x}_p])$ shows full behaviour.

We can also obtain results about shifting levels, both initial, and minimum and maximum.

Proposition 3. *Given a transition system* $T = P(x_1, \ldots, x_p\,[\underline{x}_1, \ldots, \underline{x}_p; \overline{x}_1, \ldots, \overline{x}_p])$ *and* $(m_1, \ldots, m_p) \in \mathbb{Z}^p$ *such that* $m_i \geq -x_i$ *for all* $1 \leq i \leq p$ *then* $P(x_1 + m_1, \ldots, x_p + m_p\,[\underline{x}_1 + m_1, \ldots, \underline{x}_p + m_p; \overline{x}_1 + m_1, \ldots, \underline{x}_n + m_p])$ *is isomorphic to* T.

Notice that we cannot scale systems and hope to get similarly shaped systems. For example, consider $P(5, 0, 0\,[5, 5, 5]$ and $P(10, 0, 0\,[10, 10, 10])$. This latter transition system, like $P(6, 0, 0\,[6, 6, 6])$ in Figure 4 has many more states than $P(5, 0, 0\,[5, 5, 5])$, hence isomorphism is not possible. Also it has similar structure in the lower left hand corner to $P(6, 0, 0\,[6, 6, 6])$ which differs from the structure of $P(5, 0, 0\,[5, 5, 5])$.

Now that we have completed the characterisation and investigation of these transition systems generated by Bio-PEPA systems, we can proceed with defining an appropriate equivalence for them.

4. 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 bisimilarity [20]. It is defined over a collection of processes or models, $\mathcal{M}$.

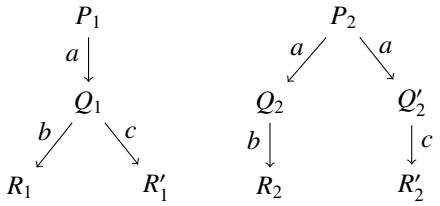

Figure 5: Example of transition systems that are not bisimilar

Definition 13. A binary relation $\mathcal{R}$ over $\mathcal{M}$ 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 bisimilarity 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 [21]. 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 5. 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.

4.1. 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

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

$$\langle \mathcal{T}^3, A(7)\rangle \xrightarrow{(\alpha, u_2)} \langle \mathcal{T}^3, A(6)\rangle \xrightarrow{(\alpha, u_1)} \langle \mathcal{T}^3, A(5)\rangle \xrightarrow{(\alpha, u_0)} \langle \mathcal{T}^3, A(4)\rangle$$

Figure 6: Example of discretisations that are not bisimilar

equivalences such as bisimilarity. Note that we cannot use bisimilarity directly here. This can be shown by the species component $A \stackrel{def}{=} (\alpha, 1)\downarrow A$. Figure 6 gives the transition system for two different discretisations, one where the minimum level is 0 and the maximum level is 2 and the other where the minimum level is 4 and the maximum level is 7. The first discretisation has 3 levels and the second, 4 levels. We can relate $\langle \mathcal{T}^2, A(0)\rangle$ with $\langle \mathcal{T}^3, A(4)\rangle$, $\langle \mathcal{T}^2, A(1)\rangle$ with $\langle \mathcal{T}^3, A(5)\rangle$ and $\langle \mathcal{T}^2, A(2)\rangle$ with $\langle \mathcal{T}^3, A(6)\rangle$, but we cannot relate $\langle \mathcal{T}^3, A(7)\rangle$ to any of $\langle \mathcal{T}^2, A(i)\rangle$. Trace equivalence cannot be used either as $\langle \mathcal{T}^3, A(7)\rangle$ has a longer trace than any of $\langle \mathcal{T}^2, A(i)\rangle$.

However, although we cannot use bisimilarity 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. Unfortunately, it is necessary to use the term *equivalence* in two different ways. Here we are considering an equivalence relation that will divide our states into different classes based on their potential behaviour, namely their outgoing transitions. We will then define a *semantic equivalence* based on bisimilarity that will associate classes with the same behaviour where this definition of behaviour considers both the transition and the resultant state.

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. Although the definition is motivated by our understanding of the transitions that are possible, it is also sensible in biological terms since 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 14. 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 4. *$\mathcal{H}$ 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 $\mathcal{P} \in H$ 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 15. 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 $\mathcal{P} \xrightarrow{\alpha} \mathcal{P}'$.

We can then finalise the definition for our new equivalence as follows. We use Definition 13 for the definition of $\sim$ but substitute $\overset{\alpha}{\hookrightarrow}$ for all instances of $\xrightarrow{\theta}$ and, moreover the relation $\sim$ is defined between equivalence classes.

Definition 16. A binary relation $\mathcal{R}$ is a *bisimulation* if for any $(H_1, H_2) \in \mathcal{R}$ and for any α whenever

1. $H_1 \overset{\alpha}{\hookrightarrow} H_1'$, there exists H_2' such that $H_2 \overset{\alpha}{\hookrightarrow} H_2'$ and $(H_1', H_2') \in \mathcal{R}$, and
2. $H_2 \overset{\alpha}{\hookrightarrow} H_1$, there exists H_1' such that $H_1 \overset{\alpha}{\hookrightarrow} H_2'$ and $(H_1', H_2') \in \mathcal{R}$

H_1 and H_2 are *bisimilar*, $H_1 \sim H_2$ if $(H_1, H_2) \in \mathcal{R}$ for some bisimulation $\mathcal{R}$.

Definition 17. $\mathcal{P}$ and $\mathcal{Q}$ are *compression bisimilar*, $\mathcal{P} \backsimeq \mathcal{Q}$, if $\mathcal{P}_{\mathcal{H}} \sim \mathcal{Q}_{\mathcal{H}}$.

We need show that it is an equivalence relation.

Proposition 5. *$\mathcal{P} \backsimeq \mathcal{Q}$ is an equivalence relation.*

PROOF. This is straightforward because $\sim$ is an equivalence relation. □

We now consider various results for the equivalence.

4.2. Equivalence of sequential systems

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. Before proceeding, we need to define some values of interest that will be used in these proofs.

Definition 18. For a sequential Bio-PEPA component $C \stackrel{def}{=} \sum_{i=1}^{q} (\alpha_i, \kappa_i)\ \mathsf{op}_i\ C$, let

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

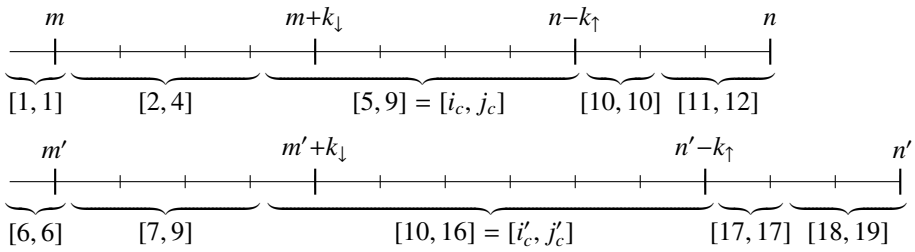

Figure 7: The equivalence classes of two discretisations of a species component

Additionally, we will use the following notation for the rest of this section: m, m' are minimum levels, n, n' are maximum levels and $\lambda = n - m + 1, \lambda' = n' - m' + 1$ are total number of levels.

The diagram in Figure 7 illustrates the equivalence classes for two discretisations of $C \stackrel{def}{=} (\alpha, 2) \uparrow C + (\beta, 3) \uparrow C + (\gamma, 4) \downarrow C + (\delta, 1) \oplus C$. The top discretisation has $m = 1$, $n = 12$ and hence $\lambda = 12$. The bottom discretisation has $m' = 6$, $n' = 19$ and $\lambda' = 14$. It also demonstrates how the various stoichiometry coefficients result in different equivalence classes. In the top diagram in Figure 7, there are five equivalence classes with the leftmost consisting of the level at which only α and β are possible, the next class where δ, α and β are possible, after which we find the central class where all reactions are possible, followed by the single level where γ, δ and α can happen. The rightmost class consists of the levels which only allow γ and δ. Considering the lower diagram, a similar pattern can be seen and this pattern is the intuition behind the first theorem.

The lemmas that follow prove the properties of the equivalence classes as shown in this diagram, and the diagram will be used as a running example to illustrate the concepts. 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 $C \stackrel{\text{def}}{=} \sum_{i=1}^{q} (\alpha_i, \kappa_i)\ \mathsf{op}_i\ C$ and the Bio-PEPA system $\mathcal{S}^\lambda = \langle \mathcal{T}^\lambda, C \rangle$, the equivalence classes of $\mathcal{S}^\lambda_{\mathcal{H}}$ form a strict order.*

PROOF. The set $ds(\mathcal{S}^\lambda)$ contains elements of the form $\langle \mathcal{T}^\lambda, C(l)\rangle$, where l ranges over $m, \ldots, n$. First, we need to show that each equivalence class is a subsequence of $m, \ldots, n$. Let $i \leq j$ be two values in an equivalence class, then we need to show for any k such that $i < k < j$, k is in the same class. Note that for level i and level j, the same actions are possible since they are in the same equivalence class. By inspection of the side conditions of the prefix rules, it is clear that the same actions are possible at level k, hence k is in the same equivalence class. 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. □

For convenience, we identify the equivalence class that consists of the levels from which all actions are possible. Examples of this class can be seen in Figure 7.

Definition 19. Given a sequential Bio-PEPA component $C \stackrel{def}{=} \sum_{i=1}^{q} (\alpha_i, \kappa_i)\ \mathsf{op}_i\ C$ and the Bio-PEPA system $\mathcal{S}^\lambda = \langle \mathcal{T}^\lambda, C\rangle$, if $\mathcal{S}^\lambda_{\mathcal{H}}$ has an equivalence class H such that $\mathcal{A}(H) = \mathcal{A}_C$, then H is the *central equivalence class* and denoted $[i_c, j_c]$.

As will be shown by the construction in Lemma 2, this class is unique. We can use the stoichiometric coefficients to characterise the number of equivalence classes, assuming sufficient levels are used. The equivalence classes below the central class are determined by stoichiometric coefficients that appear in reaction terms and activator terms, since the availability of a species determines whether a reaction can occur. In Figure 7, in the upper diagram, the term $(\delta, 1) \oplus C$ determines the end of the first equivalence class and the start of the second one and the coefficient 4 in the term $(\gamma, 4) \downarrow C$ determines the end of the second class and the start of the central class. Similarly for classes above the central class, the product terms play the same role since the maximum level constrains how much can be produced in a reaction. The coefficient 2 in $(\alpha, 2) \uparrow C$ determines the last class, and the coefficient 3 in $(\beta, 3) \uparrow C$ determines the boundary between the central class and next class. Considering this latter class, note that $n - 2$ is the level at which α reactions become possible (and then remain

possible at every smaller level) whereas $n-3$ is the level at which β reactions become possible (and remain possible at every smaller level), so the class $[n-3+1, n-2]$ exactly covers those levels (only one in this case) where α reactions (and γ and δ reactions) are possible but β reactions are not. The following lemma formalises these concepts and gives a fixed number of equivalence classes if there are sufficient levels.

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

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

Next we consider the equivalence classes that come before $[i_c, j_c]$. We can 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 the sequence of equivalence classes $[i_1, j_1], [i_2, j_2], \ldots, [i_{c-1}, j_{c-1}]$ is exactly the sequence $[m, m + \tau_1 - 1], [m + \tau_1, m + \tau_2 - 1], \ldots, [m + \tau_{t_\downarrow - 1}, m + \tau_{t_\downarrow} - 1]$ which gives $t_\downarrow$ classes.

Likewise $[i_{c+1}, j_{c+1}], \ldots, [i_{t-1}, j_{t-1}], [i_t, j_t]$ is the sequence of equivalence classes $[n - \tau'_{t_\uparrow} + 1, n - \tau'_{t_\uparrow - 1}], \ldots, [n - \tau'_2 + 1, n - \tau'_1], [n - \tau'_1 + 1, n]$ where $\tau'_1, \tau'_2, \ldots, \tau'_{t_\uparrow}$ are the ordered elements of $T_\uparrow$ with $\tau'_{t_\uparrow} = k_\uparrow$. This gives another $t_\uparrow$ equivalence classes. This means that there are $t_\downarrow + t_\uparrow + 1$ equivalence classes in total. □

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

The next lemma establishes a lower bound on the size of the central equivalence class that is capable of performing all actions. 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 as demonstrated in Lemma 2, and illustrated in Figure 7. The value k_m, the maximum stoichiometry of any reaction that involves a reactant, activator or product, 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* $C \stackrel{\text{def}}{=} \sum_{i=1}^{q} (\alpha_i, \kappa_i)\ \mathsf{op}_i\ C$ *be a sequential Bio-PEPA component and let* $\mathcal{S}^\lambda = \langle \mathcal{T}^\lambda, C \rangle$ *for* $\lambda \geq k_\uparrow + k_m + k_\downarrow + 1$. *Then* $[i_c, j_c]$ *the central equivalence class of* $\mathcal{S}^\lambda_{\mathcal{H}}$ *has cardinality greater than* k_m.

PROOF. Note that $i_c = m + k_\downarrow$ and $j_c = n - k_\uparrow$. The cardinality of $[i_c, j_c]$ is $j_c - i_c + 1 = n - k_\uparrow - m - k_\downarrow + 1 = \lambda - k_\uparrow - k_\downarrow \geq k_\uparrow + k_m + k_\downarrow + 1 - k_\uparrow - k_\downarrow = k_m + 1 > k_m$. □

This implies that the cardinality of $[i_c, j_c]$ is greater than both $k_\downarrow$ and $k_\uparrow$ and from this we obtain the following corollary.

Corollary 2. *Let* $C \stackrel{\text{def}}{=} \sum_{i=1}^{q} (\alpha_i, \kappa_i)\ \mathsf{op}_i\ C$ *be a sequential Bio-PEPA component and let* $\mathcal{S}^\lambda = \langle \mathcal{T}^\lambda, C \rangle$ *for* $\lambda \geq k_\uparrow + k_m + k_\downarrow + 1$. *Then* $\mathcal{S}^\lambda$ *demonstrates full behaviour.*

PROOF. We need to show that for every reaction α, there exist $i, j \in [i_c, j_c]$ such that $\langle \mathcal{T}^\lambda, C(i) \rangle \xrightarrow{\alpha} \langle \mathcal{T}^\lambda, C(j) \rangle$. This is clearly true since for any reaction where C has the role of a product, choose i_c then $i_c + \kappa < j_c$ and for any reaction where C has the role of a reactant or an activator, choose j_c then $i_c < j_c + \kappa$ since $\kappa \leq k_m$ which is less than the cardinality of $[i_c, j_c]$. □

This shows that when λ is sufficiently large, then for every reaction, there is a transition from the central equivalence class to itself. Once λ is this large, all transitions that can take place between equivalence classes for a species are enabled, hence our definition of full behaviour.

The next lemma relates the equivalence classes obtained for two different values of m and n by expressing the classes of the discretisation with the larger n value in terms

of the other classes. Considering the two discretisations in Figure 7, it can be seen that $m' = m + 5$ and $n' = n + 7$ and that the boundaries of the classes as defined by the stoichiometric coefficients retain these offsets. If $m' < m$ then a negative offset would be necessary but this is not a problem whenever the central class is large enough.

Lemma 4. *Let $C \stackrel{\text{def}}{=} \sum_{i=1}^{q} (\alpha_i, \kappa_i)\, \mathsf{op}_i\, C$ be a sequential Bio-PEPA component and let $\mathcal{S} = \langle \mathcal{T}, C \rangle$. Let $\lambda, \lambda' \geq k_\uparrow + k_m + k_\downarrow + 1$ and $n' \geq n$ where $m' = m + d_1$ for $d_1 \in \mathbb{Z}$, $d_1 \geq -m$ and $n' = n + d_2$ for $d_2 \in \mathbb{N}$, $d_2 \geq 1$. Then the equivalence classes of $\mathcal{S}^{\lambda}_{\mathcal{H}}$ are described by the ordered intervals $[m, j_1], \ldots, [i_c, j_c], \ldots, [i_t, n]$ and the equivalence classes of $\mathcal{S}^{\lambda'}_{\mathcal{H}}$ are described by the ordered intervals $[m + d_1, j_1 + d_1], \ldots, [i_{c-1} + d_1, j_{c-1} + d_1], [i_c + d_1, j_c + d_2], [i_{c+1} + d_2, j_{c+1} + d_2], \ldots, [i_t + d_2, n + d_2]$ where $[i_c, j_c]$ and $[i_c + d_1, j_c + d_2]$ are the central equivalence classes.*

Proof. The elements of $\mathcal{S}^{\lambda}_{\mathcal{H}}$ are $[m, j_1], \ldots, [i_c, j_c], \ldots, [i_t, n]$ and those of $\mathcal{S}^{\lambda'}_{\mathcal{H}}$ are $[m', j'_1], \ldots, [i'_c, j'_c], \ldots, [i'_t, n']$. Since $m' = m + d_1$ and equivalence class boundaries are defined by stoichiometric coefficients, then for the first $t_\downarrow$ equivalence classes of $\mathcal{S}^{\lambda}_{\mathcal{H}}$, we know that $j'_l = j_l + d_1$ and $i'_l = i_l + d_1$ for $1 \leq l \leq t_\downarrow$. Similarly since $n' = n + d_2$, for the last $t_\uparrow$ classes we can show that $j'_l = j_l + d_2$ and $i'_l = i_l + d_2$ for $t - t_\uparrow \leq l \leq t$.

Finally, we need to consider the central class. Using the offset argument then we should have $[i'_c, j'_c] = [i_c + d_1, j_c + d_2] = [m' + k_\downarrow, n' + k_\uparrow]$. To check this, consider $i_c + d_1 = m + k_\downarrow + d_1 = m' + k_\downarrow$. Similarly, $j_c + d_2 = n' + k_\uparrow$. □

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 $C \stackrel{\text{def}}{=} \sum_{i=1}^{q} (\alpha_i, \kappa_i)\, \mathsf{op}_i\, C$ 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 $\mathcal{S}^{\lambda}_{\mathcal{H}}$ and $E'_1, \ldots, E'_t$ be the ordered equivalence classes of $\mathcal{S}^{\lambda'}_{\mathcal{H}}$. If $\lambda, \lambda' \geq k_\downarrow + k_m + k_\uparrow + 1$ then $E_{p_1} \stackrel{\alpha}{\hookrightarrow} E_{p_2}$ if and only if $E'_{p_1} \stackrel{\alpha}{\hookrightarrow} E'_{p_2}$.*

Proof. Let $E_p = [i_p, j_p]$ and $E'_p = [i'_p, j'_p]$ for all $1 \leq p \leq t$. For each $E_{p_1} \stackrel{\alpha}{\hookrightarrow} E_{p_2}$, there exists a transition $\langle \mathcal{T}^{\lambda}, C(l_1) \rangle \stackrel{\alpha}{\longrightarrow} \langle \mathcal{T}^{\lambda}, C(l_2) \rangle$ with $l_1 \in E_{p_1}$ and $l_2 \in E_{p_2}$ and

where $l_2 = l_1 + \nu$ for $\nu \in \mathbb{Z}$. Note that ν is determined by κ and type of reaction prefix. We need to find $\langle \mathcal{T}^{\lambda'}, C(l'_1)\rangle \xrightarrow{\alpha} \langle \mathcal{T}^{\lambda}, C(l'_2)\rangle$ with $l'_1 \in E'_{p_1}$ and $l'_2 \in E'_{p_2}$ and hence $E'_{P_1} \xhookrightarrow{\alpha} E'_{p_2}$. Without loss of generality, assume $n' > n$ with $n' = n + d_2$ for $d_2 \in \mathbb{N}$, $d_2 \geq 1$ and let $m' = m + d_1$ for $d_1 \in \mathbb{Z}$ and $d_1 \geq -m$. We can then use Lemma 4.

If we call the equivalence classes below the central class "lower" and those above "upper", we can identify 9 cases of transitions for consideration. First, note that transitions are not possible from lower classes to upper classes or from upper classes to lower because the central class is larger than any stoichiometric coefficient.

Next, we consider transitions from lower to lower with a change from level $l_1 \in [i_{p_1}, j_{p_1}]$ to $l_2 \in [i_{p_2}, j_{p_2}]$ in $\mathcal{S}^{\lambda}$. Let $l'_1 = l_1 + d_1$ and $l'_2 = l_2 + d_1$. By the lemma, $l'_1 \in [i_{p_1} + d_1, j_{p_1} + d_1] = [i'_{p_1}, j'_{p_1}]$, $l'_2 \in [i_{p_2} + d_1, j_{p_2} + d_1] = [i'_{p_2}, j'_{p_2}]$ and we have a matching transition in $\mathcal{S}^{\lambda'}$. This also applies in the case of a transition from a lower class to the central class. A similar argument can be used for transitions from upper to upper, or upper to central using $l'_1 = l_1 + d_2$ and $l'_2 = l_2 + d_2$.

For a transition from the central class to itself, it is always possible to find a matching transition because of the size of the central class in both $\mathcal{S}^{\lambda}_{\mathcal{H}}$ and $\mathcal{S}^{\lambda'}_{\mathcal{H}}$.

For a transition from the central class to a lower class, we have $l_1 \in [i_c, j_c]$ and $l_2 \in [i_p, j_p]$. In fact $l_1 \in [i_c, i_c + k_{\downarrow} - 1]$ otherwise the transition could not reach a lower class. Let $l'_1 = l_1 + d_1$ and $l'_2 = l_2 + d_1$. Then by the lemma, $l'_1 \in [i_c + d_1, i_c + k_{\downarrow} - 1 + d_1] = [i'_c, i'_c + k_{\downarrow} - 1] \subseteq [i'_c, j'_c]$, $l'_2 \in [i_p + d_1, j_p + d_1] = [i'_p, j'_p]$. Hence there is a matching transition in $\mathcal{S}^{\lambda'}$.

For a transition from the central class to an upper class, we have $l_1 \in [i_c, j_c]$ and $l_2 \in [i_p, j_p]$. Like before $l_1 \in [j_c - k_{\uparrow} + 1, j_c]$ so that the transition can reach an upper class. Let $l'_1 = l_1 + d_2$ and $l'_2 = l_2 + d_2$. Then by the lemma, $l'_1 \in [j_c - k_{\uparrow} + 1 + d_2, j_c + d_2] = [j'_c - k_{\uparrow} + 1, j'_c] \subseteq [i'_c, j'_c]$, $l'_2 \in [i_p + d_2, j_p + d_2] = [i'_p, j'_p]$, leading to a matching transition in $\mathcal{S}^{\lambda'}$.

Similarly, it is possible to show that every transition between classes in $\mathcal{S}^{\lambda'}_{\mathcal{H}}$ is matched by one in $\mathcal{S}^{\lambda}_{\mathcal{H}}$. □

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.

The following theorem shows that for large enough value of λ, 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 $C \stackrel{\text{def}}{=} \sum_{i=1}^{q} (\alpha_i, \kappa_i)\ \mathsf{op}_i\ C$ then $\mathcal{S}^{\lambda} \backsimeq \mathcal{S}^{\lambda'}$ for $\lambda, \lambda' \geq k_{\downarrow} + k_m + k_{\uparrow} + 1$.*

PROOF. Without loss of generality, assume that λ is the maximum level for species C in $\mathcal{S}^{\lambda}$ and λ' is the maximum level for species C in $\mathcal{S}^{\lambda'}$.

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

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

As mentioned previously, a sufficient number of levels is crucial for this result. If we have fewer levels, then it is not possible to prove Lemma 5. For example, if the central class is too small for one of the discretisations, it may be possible for there to be a transition from an equivalence class below the central class to one above, or with fewer levels, some behaviour may not be displayed because certain reactions are excluded because their stoichiometric coefficients are too large with respect to the number of levels. In these cases, the transition systems over the equivalence class will certainly differ from the transition system with sufficiently many levels. Hence, by ensuring both systems are large enough, we can show that they have same structure of transitions.

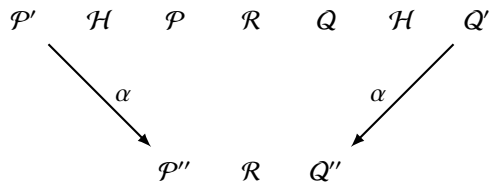

Figure 8: Action equivalence

4.3. Equivalence of parallel systems

We next consider a conditional congruence result for the synchronisation operator. In this theorem, the notation $[\mathcal{P}]$ 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 present a different definition for compression bisimulation, then some properties describing the actions that are possible in a synchronisation and a few lemmas.

Compression bisimulation is defined in terms of bisimilarity of equivalence classes but it can also be defined directly as follows.

Definition 20. A relation $\mathcal{R}$ over Bio-PEPA systems is an *action equivalence* if for $(\mathcal{P}, \mathcal{Q}) \in \mathcal{R}$,

1. for all $\mathcal{P}'$ such that $(\mathcal{P}', \mathcal{P}) \in \mathcal{H}$ and $\mathcal{P}' \xrightarrow{\alpha} \mathcal{P}''$ there exist $\mathcal{Q}'$ and $\mathcal{Q}''$ with $(\mathcal{Q}', \mathcal{Q}) \in \mathcal{H}, \mathcal{Q}' \xrightarrow{\alpha} \mathcal{Q}''$ and $(\mathcal{P}'', \mathcal{Q}'') \in \mathcal{R}$,
2. for all $\mathcal{Q}'$ such that $(\mathcal{Q}', \mathcal{Q}) \in \mathcal{H}$ and $\mathcal{Q}' \xrightarrow{\alpha} \mathcal{Q}''$ there exist $\mathcal{P}'$ and $\mathcal{P}''$ with $(\mathcal{P}', \mathcal{P}) \in \mathcal{H}, \mathcal{P}' \xrightarrow{\alpha} \mathcal{P}''$ and $(\mathcal{P}'', \mathcal{Q}'') \in \mathcal{R}$.

Let $\mathcal{P} \leftrightarrows \mathcal{Q}$ whenever $(\mathcal{P}, \mathcal{Q}) \in \mathcal{R}$ for $\mathcal{R}$ an action equivalence.

Figure 8 illustrates the structure of the relation. Using this relation, we are able to ensure that we have a correct congruence result. The key is to ensure that the $\mathcal{P}'$ and $\mathcal{Q}'$ are the correct ones. We discuss this in more detail below. First we give some results about this equivalence and show that it is the same as compression bisimulation.

Proposition 6.

1. $\leftrightarrows$ *is the largest action equivalence.*
2. $\leftrightarrows$ *is an equivalence relation.*

PROOF. Straightforward. □

Proposition 7. $\mathcal{P} \leftrightarrows \mathcal{Q} \;\Leftrightarrow\; \mathcal{P} \backsimeq \mathcal{Q}$.

PROOF. ($\Rightarrow$) Let $\mathcal{R} = \{([\mathcal{P}], [\mathcal{Q}]) \mid \mathcal{P} \leftrightarrows \mathcal{Q}\}$. This is a bisimulation. ($\Leftarrow$) Let $\mathcal{R} = \{(\mathcal{P}, \mathcal{Q}) \mid \mathcal{P} \backsimeq \mathcal{Q}\}$. This is an action equivalence. □

We first define a condition that ensures that basic matching is always possible.

Definition 21. Two well-defined Bio-PEPA systems, $\langle \mathcal{T}_1, P_1 \underset{L}{\bowtie} Q_1 \rangle$, $\langle \mathcal{T}_2, P_2 \underset{L}{\bowtie} Q_2 \rangle$, have the *matching derivative (MD) property* if there exists a total relation $\mathcal{M}$ with $\mathcal{M} \subseteq ds(\langle \mathcal{T}_1, P_1 \underset{L}{\bowtie} Q_1 \rangle) \times ds(\langle \mathcal{T}_2, P_2 \underset{L}{\bowtie} Q_2 \rangle)$ such that

1. if $(\langle \mathcal{T}_1, P_1' \underset{L}{\bowtie} Q_1' \rangle, \langle \mathcal{T}_2, P_2' \underset{L}{\bowtie} Q_2' \rangle) \in \mathcal{M}$ then $\langle \mathcal{T}_1, P_1' \rangle \leftrightarrows \langle \mathcal{T}_2, P_2' \rangle$, $\langle \mathcal{T}_1, Q_1' \rangle \leftrightarrows \langle \mathcal{T}_2, Q_2' \rangle$,
2. $(\langle \mathcal{T}_1, P_1 \underset{L}{\bowtie} Q_1 \rangle, \langle \mathcal{T}_2, P_2 \underset{L}{\bowtie} Q_2 \rangle) \in \mathcal{M}$.

The term "total" means that each element from one side of the relation is matched with at least one element from the other side of the relation. We also need something stronger than this for congruence. As mentioned above, we must ensure that we work with the correct elements of the equivalence class that is performing the transition. Otherwise it is easy to construct a proof that appears correct but is not because although a matching transition appears to have been found, it is actually not in the transition system under consideration. To this end, we define a notion of compatibility that ensures the transitions under consideration are from a specific transition system. It ensures that when we are considering action equivalent systems, that the systems from which the transitions come, when combined give systems that are in the transition system of interest. To achieve this end, an additional condition is imposed on the matching of transitions. Because we are considering two pairs of systems and their synchronisations, this definition cannot be reduced to a definition of action equivalence with additional constraints.

Definition 22. Given well-defined Bio-PEPA systems, $\langle \mathcal{T}_1, P_1 \underset{L}{\bowtie} Q_1 \rangle, \langle \mathcal{T}_2, P_2 \underset{L}{\bowtie} Q_2 \rangle$, with the MD property based on relation $\mathcal{M}$ then they have *compatibility* if for all $(\langle \mathcal{T}_1, P_1' \underset{L}{\bowtie} Q_1' \rangle, \langle \mathcal{T}_2, P_2' \underset{L}{\bowtie} Q_2' \rangle) \in \mathcal{M}$ and for all α, whenever

1. $(\langle \mathcal{T}_1, P_3' \rangle, \langle \mathcal{T}_1, P_1' \rangle) \in \mathcal{H}$, and $(\langle \mathcal{T}_2, P_4' \rangle, \langle \mathcal{T}_2, P_2' \rangle) \in \mathcal{H}$, with $\langle \mathcal{T}_1, P_3' \rangle \xrightarrow{\alpha} \langle \mathcal{T}_1, P_3'' \rangle$ and $\langle \mathcal{T}_2, P_4' \rangle \xrightarrow{\alpha} \langle \mathcal{T}_2, P_4'' \rangle$ and $\langle \mathcal{T}_1, P_3'' \rangle \leftrightarroweq \langle \mathcal{T}_2, P_4'' \rangle$, and
2. $(\langle \mathcal{T}_1, Q_3' \rangle, \langle \mathcal{T}_1, Q_1' \rangle) \in \mathcal{H}$, and $(\langle \mathcal{T}_2, Q_4' \rangle, \langle \mathcal{T}_2, Q_2' \rangle) \in \mathcal{H}$, with $\langle \mathcal{T}_1, Q_3' \rangle \xrightarrow{\alpha} \langle \mathcal{T}_1, Q_3'' \rangle$ and $\langle \mathcal{T}_2, Q_4' \rangle \xrightarrow{\alpha} \langle \mathcal{T}_2, Q_4'' \rangle$ and $\langle \mathcal{T}_1, Q_3'' \rangle \leftrightarroweq \langle \mathcal{T}_2, Q_4'' \rangle$,

then $\langle \mathcal{T}_1, P_3' \underset{L}{\bowtie} Q_3' \rangle \in ds(\langle \mathcal{T}_1, P_1 \underset{L}{\bowtie} Q_1 \rangle)$ and $\langle \mathcal{T}_2, P_4' \underset{L}{\bowtie} Q_4' \rangle \in ds(\langle \mathcal{T}_2, P_2 \underset{L}{\bowtie} Q_2 \rangle)$.

We next prove two lemmas that are necessary for the theorem. The first lemma illustrates that when given two subcomponents with the same actions, then two models built out of these two subcomponents using the cooperation operator, retain the property of having the same actions (but not necessarily the same actions as the subcomponents). We use this lemma in the second lemma which demonstrates that any two systems that have the MD property also have the properties of having the same actions. Both of these lemmas permit reasoning about the actions available to cooperations.

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

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

Lemma 7. *Given well-defined Bio-PEPA systems, $\langle \mathcal{T}_1, P_1 \underset{L}{\bowtie} Q_1 \rangle, \langle \mathcal{T}_2, P_2 \underset{L}{\bowtie} Q_2 \rangle$, with the MD property then for all $(\langle \mathcal{T}_1, P_1' \underset{L}{\bowtie} Q_1' \rangle, \langle \mathcal{T}_2, P_2' \underset{L}{\bowtie} Q_2' \rangle) \in \mathcal{M}$,*

$$\mathcal{A}(\langle \mathcal{T}_1, P_1' \underset{L}{\bowtie} Q_1' \rangle) = \mathcal{A}(\langle \mathcal{T}_2, P_2' \underset{L}{\bowtie} Q_2' \rangle).$$

PROOF. If $(\langle \mathcal{T}_1, P_1' \underset{L}{\bowtie} Q_1' \rangle, \langle \mathcal{T}_2, P_2' \underset{L}{\bowtie} Q_2' \rangle) \in \mathcal{M}$ then we know $\langle \mathcal{T}_1, P_1' \rangle \leftrightarroweq \langle \mathcal{T}_2, P_2' \rangle$ and $\langle \mathcal{T}_1, Q_1' \rangle \leftrightarroweq \langle \mathcal{T}_2, Q_2' \rangle$. Hence we have $[\langle \mathcal{T}_1, P_1' \rangle] \sim [\langle \mathcal{T}_2, P_2' \rangle]$ and $[\langle \mathcal{T}_1, Q_1' \rangle] \sim [\langle \mathcal{T}_2, Q_2' \rangle]$. Since bisimilarity requires matching on transitions and we have equivalence classes over $\mathcal{H}$, we have $\mathcal{A}(\langle \mathcal{T}_1, P_1' \rangle) = \mathcal{A}(\langle \mathcal{T}_2, P_2' \rangle)$ and $\mathcal{A}(\langle \mathcal{T}_1, Q_1' \rangle) = \mathcal{A}(\langle \mathcal{T}_2, Q_2' \rangle)$. By two applications of Lemma 6, we obtain $\mathcal{A}(\langle \mathcal{T}_1, P_1' \underset{L}{\bowtie} Q_1' \rangle) = \mathcal{A}(\langle \mathcal{T}_2, P_2' \underset{L}{\bowtie} Q_2' \rangle)$. □

The following theorem is a congruence result. By considering systems with the MD property, we know that $\langle \mathcal{T}_1, P_1' \rangle \nsim \langle \mathcal{T}_2, P_2' \rangle$ and $\langle \mathcal{T}_1, Q_1' \rangle \nsim \langle \mathcal{T}_2, Q_2' \rangle$ for all systems of interest. From additional conditions given by the MD property and compatibility, it is then possible to show that cooperations are compression bisimilar. The proof proceeds as is standard for congruence proofs relating to parallel operators. The relation that we show to be a compression bisimulation is exactly the relation that demonstrates the MD property, and as is standard, all possible transitions must be shown to have matching transitions with the targets of these transitions appearing as a pair in the relation. Due to the effects of stoichiometry, this requires careful reasoning and the additional conditions of the MD property and compatibility.

Theorem 2. *Let $\langle \mathcal{T}_1, P_1 \underset{L}{\bowtie} Q_1 \rangle$, $\langle \mathcal{T}_2, P_2 \underset{L}{\bowtie} Q_2 \rangle$ be two well-defined Bio-PEPA systems with the MD property and compatibility then $\langle \mathcal{T}_1, P_1 \underset{L}{\bowtie} Q_1 \rangle \simeq \langle \mathcal{T}_2, P_2 \underset{L}{\bowtie} Q_2 \rangle$.*

PROOF. Since the two systems have the MD property, there is a relation $\mathcal{M}$ with the appropriate definition. We show that $\mathcal{M}$ is an action equivalence hence we can conclude that $\langle \mathcal{T}_1, P_1 \underset{L}{\bowtie} Q_1 \rangle \simeq \langle \mathcal{T}_2, P_2 \underset{L}{\bowtie} Q_2 \rangle$ since $\mathcal{M}$ contains this pair. We only consider the case of $\alpha \in L$. The other two cases are similar but simpler.

Let $(\langle \mathcal{T}_1, P_1 \underset{L}{\bowtie} Q_1 \rangle, \langle \mathcal{T}_2, P_2 \underset{L}{\bowtie} Q_2 \rangle) \in \mathcal{M}$. We consider an arbitrary transition $\langle \mathcal{T}_1, P_3' \underset{L}{\bowtie} Q_3' \rangle \xrightarrow{\alpha} \langle \mathcal{T}_1, P_3'' \underset{L}{\bowtie} Q_3'' \rangle$ for $(\langle \mathcal{T}_1, P_1 \underset{L}{\bowtie} Q_1 \rangle, \langle \mathcal{T}_1, P_3 \underset{L}{\bowtie} Q_3 \rangle) \in \mathcal{H}$.

By shorter inferences and then applying `Qual`, we have $\langle \mathcal{T}_1, P_3' \rangle \xrightarrow{\alpha} \langle \mathcal{T}_1, P_3'' \rangle$ and $\langle \mathcal{T}_1, Q_3' \rangle \xrightarrow{\alpha} \langle \mathcal{T}_1, Q_3'' \rangle$.

Since $\mathcal{M}$ is total, there exists $\langle \mathcal{T}_2, P_4' \underset{L}{\bowtie} Q_4' \rangle$ such that $\langle \mathcal{T}_1, P_3' \rangle \nsim \langle \mathcal{T}_2, P_4' \rangle$ and $\langle \mathcal{T}_1, Q_3' \rangle \nsim \langle \mathcal{T}_2, Q_4' \rangle$ and these are compatible for $P_1 \underset{L}{\bowtie} Q_1$ and $P_2 \underset{L}{\bowtie} Q_2$.

Since $\langle \mathcal{T}_1, P_3' \rangle \nsim \langle \mathcal{T}_2, P_4' \rangle$, there exists $\langle \mathcal{T}_2, P_6' \rangle$ such that $\langle \mathcal{T}_2, P_6' \rangle \xrightarrow{\alpha} \langle \mathcal{T}_2, P_6'' \rangle$ with $(\langle \mathcal{T}_2, P_6 \rangle, \langle \mathcal{T}_2, P_4 \rangle) \in \mathcal{H}$ and $\langle \mathcal{T}_1, P_3'' \rangle \nsim \langle \mathcal{T}_2, P_6'' \rangle$. Similarly $\langle \mathcal{T}_1, Q_3' \rangle \nsim \langle \mathcal{T}_2, Q_4' \rangle$, and there exists $\langle \mathcal{T}_2, Q_6' \rangle$ such that $\langle \mathcal{T}_2, Q_6' \rangle \xrightarrow{\alpha} \langle \mathcal{T}_2, Q_6'' \rangle$ with $(\langle \mathcal{T}_2, Q_6 \rangle, \langle \mathcal{T}_2, Q_4 \rangle) \in \mathcal{H}$ and $\langle \mathcal{T}_1, Q_3'' \rangle \nsim \langle \mathcal{T}_2, Q_6'' \rangle$.

¿From these transitions, by shorter inferences followed by application of the `coop3` rule and the `Qual` rule, we obtain the transition $\langle \mathcal{T}_2, P_6' \underset{L}{\bowtie} Q_6' \rangle \xrightarrow{\alpha} \langle \mathcal{T}_2, P_6'' \underset{L}{\bowtie} Q_6'' \rangle$. Moreover by compatibility, these are valid derivatives in $ds(\langle \mathcal{T}_2, P_2 \underset{L}{\bowtie} Q_2 \rangle)$ and hence in the transition system under consideration.

To complete the proof we must show that $(\langle \mathcal{T}_2, P'_6 \underset{L}{\bowtie} Q'_6\rangle, \langle \mathcal{T}_2, P'_2 \underset{L}{\bowtie} Q'_2\rangle) \in \mathcal{H}$ to prove this is a matching α-transition and that $(\langle \mathcal{T}_1, P''_3 \underset{L}{\bowtie} Q''_3\rangle, \langle \mathcal{T}_2, P''_6 \underset{L}{\bowtie} Q''_6\rangle) \in \mathcal{M}$.

We start with the latter. From above, we know that $\langle \mathcal{T}_1, P''_3\rangle \stackrel{\cdot}{\sim} \langle \mathcal{T}_2, P''_6\rangle$ and $\langle \mathcal{T}_1, Q''_3\rangle \stackrel{\cdot}{\sim} \langle \mathcal{T}_2, Q''_6\rangle$ hence the pair is in $\mathcal{M}$.

For the former point, consider the following. Since $\mathcal{A}(\langle \mathcal{T}_2, P'_6\rangle) = \mathcal{A}(\langle \mathcal{T}_2, P'_4\rangle)$ and $\mathcal{A}(\langle \mathcal{T}_2, Q'_6\rangle) = \mathcal{A}(\langle \mathcal{T}_2, Q'_4\rangle)$ then by two applications of Lemma 6,

$$\begin{aligned}
\mathcal{A}(\langle \mathcal{T}_2, P'_6 \underset{L}{\bowtie} Q'_6\rangle) &= \mathcal{A}(\langle \mathcal{T}_2, P'_4 \underset{L}{\bowtie} Q'_4\rangle) \\
&= \mathcal{A}(\langle \mathcal{T}_1, P'_3 \underset{L}{\bowtie} Q'_3\rangle) \quad \text{(paired by } \mathcal{M}\text{, so Lemma 7 applies)} \\
&= \mathcal{A}(\langle \mathcal{T}_1, P'_1 \underset{L}{\bowtie} Q'_1\rangle) \quad (\langle \mathcal{T}_1, P'_3 \underset{L}{\bowtie} Q'_3\rangle \in [\langle \mathcal{T}_1, P'_1 \underset{L}{\bowtie} Q'_1\rangle]) \\
&= \mathcal{A}(\langle \mathcal{T}_2, P'_2 \underset{L}{\bowtie} Q'_2\rangle) \quad \text{(paired by } \mathcal{M}\text{, so Lemma 7 applies)}
\end{aligned}$$

Therefore $(\langle \mathcal{T}_2, P'_6 \underset{L}{\bowtie} Q'_6\rangle, \langle \mathcal{T}_2, P'_2 \underset{L}{\bowtie} Q'_2\rangle) \in \mathcal{H}$ as required. □

As mentioned above, this is described as a conditional congruence result. This is because we cannot just use the fact that $\langle \mathcal{T}_1, P_1\rangle \stackrel{\cdot}{\sim} \langle \mathcal{T}_2, P_2\rangle$ and $\langle \mathcal{T}_1, Q_1\rangle \stackrel{\cdot}{\sim} \langle \mathcal{T}_2, Q_2\rangle$ to obtain the result. This fact is implied directly by the existence of $\mathcal{M}$ but additional relationships are required. Examples to illustrate this are presented after the next theorem.

We now prove a theorem about two discretisations of the same system. As before we need to ensure that our transition systems are sufficiently large. We need some definitions first.

Definition 23. Given a well-defined Bio-PEPA system $\mathcal{P} = \langle \mathcal{T}, P\rangle$ with $P \stackrel{def}{=} C_1(l_1) \underset{\mathcal{L}_1}{\bowtie} \ldots \underset{\mathcal{L}_{p-1}}{\bowtie} C_p(l_p)$. Define

$$\begin{aligned}
K_\uparrow &= \max(\{k_\uparrow \mid C \text{ appears in } P\}) \\
K_\downarrow &= \max(\{k_\downarrow \mid C \text{ appears in } P\}) \\
K_m &= \max(\{K_\downarrow, K_\uparrow\}) \\
\text{levels}_{\mathcal{P}}(C_i) &= \{y_i \mid (\ldots, y_i, \ldots) \text{ a state in the transition system of } \mathcal{P}\} \\
\min_{\mathcal{P}}(C_i) &= \min(\text{levels}_{\mathcal{P}}(C_i)) \\
\max_{\mathcal{P}}(C_i) &= \max(\text{levels}_{\mathcal{P}}(C_i)) \\
\text{size}_{\mathcal{P}}(C_i) &= \max_{\mathcal{P}}(C_i) - \min_{\mathcal{P}}(C_i) + 1 \\
\mu_{\mathcal{P}}(C_i) &= N_{C_i} - N'_{C_i} + 1
\end{aligned}$$

The first three values capture the largest and smallest stoichiometric coefficients in a Bio-PEPA model and their maximum. The next four consider the actual levels for a particular species within a given transition system. The final value captures the range of levels of a species as defined by its maximum and minimum number of levels. We make the following observation that relates the number of levels that appear in a transition system for a species to its range of levels.

For a well-defined Bio-PEPA system $\mathcal{P} = \langle \mathcal{T}, P \rangle$ with $P \stackrel{def}{=} C_1(l_1) \underset{\mathcal{L}_1}{\bowtie} \ldots \underset{\mathcal{L}_{p-1}}{\bowtie} C_p(l_p)$, we have that for all $1 \leq i \leq p$, $\mu_{\mathcal{P}}(C_i) \geq \text{size}_{\mathcal{P}}(C_i)$ since it is not possible for a species to go beyond its maximum and minimum levels.

Theorem 3. *Let $\mathcal{P} = \langle \mathcal{T}, P \rangle$ be a well-defined Bio-PEPA system with the definition $P \stackrel{\text{def}}{=} C_1(l_1) \underset{\mathcal{L}_1}{\bowtie} \ldots \underset{\mathcal{L}_{p-1}}{\bowtie} C_p(l_p)$ and let $\text{size}_{\mathcal{P}}(C_i) \geq K_\downarrow + K_m + K_\uparrow + 1$ for all $1 \leq i \leq p$. If the MD property and compatibility apply to pairs of subcomponents of $\mathcal{P}^\lambda$ and $\mathcal{P}^{\lambda'}$ then $\mathcal{P}^\lambda \backsimeq \mathcal{P}^{\lambda'}$.*

PROOF. Since $\mathcal{P}^\lambda$ is defined as the system where the smallest number of levels of the species in $\mathcal{P}$ is λ, there exists C_j such that $\lambda = \mu_{\mathcal{P}^\lambda}(C_j)$, therefore since a species cannot go outside its minimum and maximum levels, $\lambda \geq \max_{\mathcal{P}}(C_j) - \min_{\mathcal{P}}(C_j) + 1 \geq K_\downarrow + K_m + K_\uparrow + 1$. A similar argument can be made to show $\lambda' \geq K_\downarrow + K_m + K_\uparrow + 1$

Therefore, by Theorem 1, $\langle \mathcal{T}^\lambda, C_i \rangle \backsimeq \langle \mathcal{T}^{\lambda'}, C_i \rangle$ for all i since $\lambda \geq K_\downarrow + K_m + K_\uparrow + 1 \geq k_\downarrow + k_m + k_\downarrow + 1$. By repeated applications of Theorem 2, $\langle \mathcal{T}^\lambda, P \rangle \backsimeq \langle \mathcal{T}^{\lambda'}, P' \rangle$. □

The conditions on this theorem are slightly more general than is necessary, since if we consider a specific bracketing of $C_1(l_1) \underset{\mathcal{L}_1}{\bowtie} \ldots \underset{\mathcal{L}_{p-1}}{\bowtie} C_p(l_p)$, it is possible to be more precise about which subcomponents of P^λ and $P^{\lambda'}$ are to be paired. For the bracketing

$$C_1(l_1) \underset{\mathcal{L}_1}{\bowtie} \Big(C_2(l_2) \underset{\mathcal{L}_2}{\bowtie} \Big(\ldots \underset{\mathcal{L}_{p-3}}{\bowtie} \Big(C_{p-2}(l_{p-2}) \underset{\mathcal{L}_{p-2}}{\bowtie} \Big(C_{p-1}(l_{p-1}) \underset{\mathcal{L}_{p-1}}{\bowtie} C_p(l_p) \Big) \Big) \ldots \Big) \Big)$$

the MD property and compatibility must apply first to $C_{p-1}(l_{p-1})$ and $C_p(l_p)$ in the two different discretisations, and then to $C_{p-2}(l_{p-2})$ and $C_{p-1}(l_{p-1}) \underset{\mathcal{L}_{p-1}}{\bowtie} C_p(l_p)$, and all the way up to $C_1(l_1)$ and the rest of the cooperation. Obviously different bracketings would have different requirements. The theorem condition requires that the MD property and compatibility holds of every possible way to split P into subcomponents, and hence is general enough to cover all bracketings of the expression.

The difference between the theorem for congruence presented above and that of the synchronisation operator in PEPA [7] is that in the latter, we know that the synchronisation reduces the transitions in the identical way for models being constructed. Here, we have a more complex interaction resulting in different transitions for each model.

To see this, consider the equivalence class $E = \{(5,0,0),(3,1,0)\}$ from the transition system $P(5,0,0\,[5,5,5])$ illustrated in Figure 3 and the equivalence class $F = \{(6,0,0),(4,1,0),(2,2,0)\}$ from $P(6,0,0\,[6,6,6])$ illustrated in Figure 4. The actions that are possible from E and F are $\{\alpha_1,\alpha_2\}$. Note that we require the number of levels to be greater than $k_\uparrow + k_m + k_\downarrow = 1+2+2 = 5$ hence with minimum level 0 and maximum level 5 giving us 6 levels, we have sufficient levels in both transition systems.

These are not compression bisimilar because the α_3-transition from E to $\{(1,2,0)\}$ has no equivalent in $P(6,0,0\,[6,6,6])$. Clearly $A(5\,[5])$ and $A(6\,[6])$ are compression bisimilar, but if we use this fact without compatibility then we can incorrectly infer the transition $(2,1,0) \xrightarrow{\alpha_3} (0,2,0)$ which is not in $P(5,0,0\,[5,5,5])$, although $\mathcal{A}((2,1,0)) = \mathcal{A}((3,1,0))$ and $\mathcal{A}((0,2,0)) = \mathcal{A}((0,3,0))$.

In fact if we try to construct a relation $\mathcal{M}$ to show that the two systems have the MD property, we cannot because there is no state to pair $(1,2,0)$ with and no state to pair $(0,3,0)$. In the case of $(1,2,0)$, we need to find a state (x_1,x_2,x_3) such that $A(1\,[5])$ is equivalent to $A(x_1\,[6])$ and $(B \underset{*}{\bowtie} C)(2,0\,[5,5])$ is equivalent to $(B \underset{*}{\bowtie} C)(x_2,x_3\,[5,5])$. The only option for x_1 is 1 since that captures when only the reaction α_1 is possible. Additionally x_3 must be zero since no α_2 is possible. On inspection of $P(6,0,0\,[6,6,6])$ there are no states meeting these criteria. The equivalence classes for $P(5,0,0\,[5,5,5])$ and $P(6,0,0\,[6,6,6])$ are given in Figures 9 and 10.

On the other hand if we consider the transition system $P(7,0,0\,[7,7,7])$, we can construct the relation $\mathcal{M}$ and show compatibility. This leads to the following hypothesis that we wish to explore as further work.

Hypothesis 1. *Define T to be the least common multiple of all the stoichiometric coefficients in a well-defined Bio-PEPA system $\langle \mathcal{T}, P\rangle$. If $\lambda' = \lambda + cT$, $c \in \mathbb{N}$ and $\text{size}_{\mathcal{P}}(C) \geq K_\uparrow + K_m + K_\downarrow + 1$ for all sequential components C in $\mathcal{P}$ then $\langle \mathcal{T}^\lambda, P\rangle \backsimeq \langle \mathcal{T}^{\lambda'}, P\rangle$.*

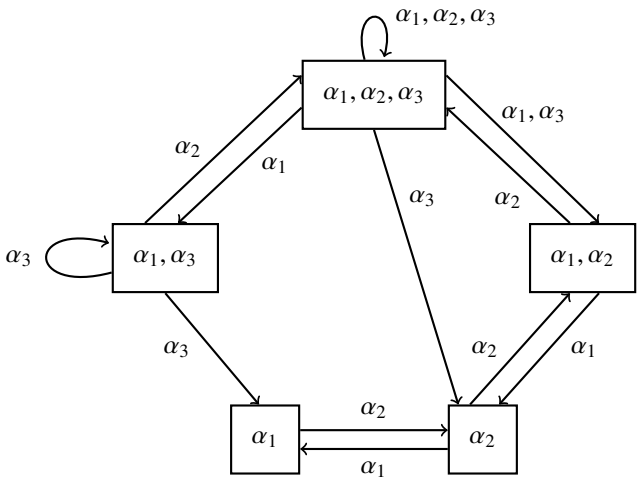

Figure 9: Equivalence classes and transitions for $P(5, 0, 0\,[5, 5, 5])$.

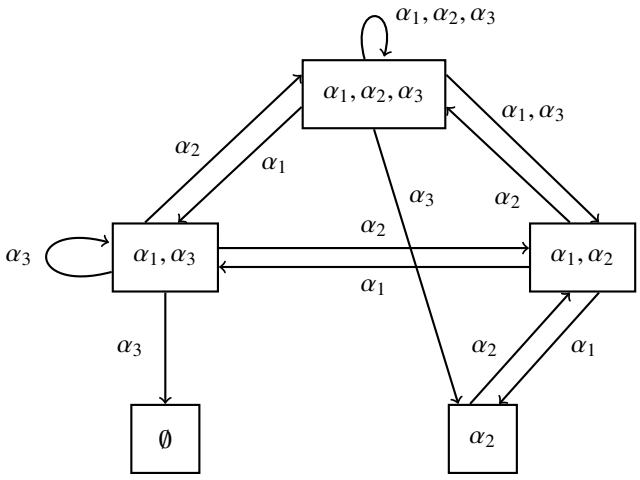

Figure 10: Equivalence classes and transitions $P(6, 0, 0\,[6, 6, 6])$.

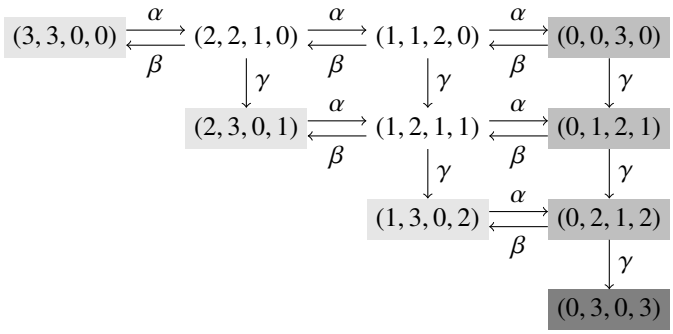

Figure 11: Transition system for *Sys* when $n = 3$

5. Substrate/enzyme example

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

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

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

$$Sys \stackrel{def}{=} S(x) \underset{\{\alpha,\beta\}}{\bowtie} E(x) \underset{\{\alpha,\beta,\gamma\}}{\bowtie} SE(0) \underset{\{\gamma\}}{\bowtie} P(0)$$

Figure 11 gives the transition system for *Sys* when the maximum levels for all species is three, and its equivalence classes are shown in Figure 12. In the transition system, 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). Figure 13 gives the system when the maximum levels is seven. This demonstrates how the two discretisations are related by the equivalence classes given in Figure 12. The shading shows the different equivalence classes in both diagrams.

6. Application to other formalisms for systems biology

We have presented compression bisimilarity and our congruence result in the context of the process algebra Bio-PEPA. However, this style of equivalence has potential

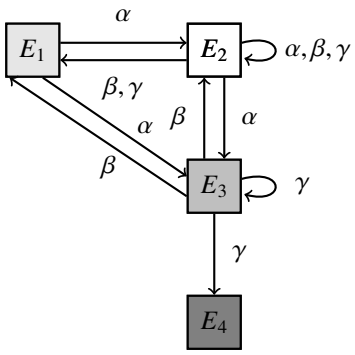

Figure 12: Equivalence classes for substrate/enzyme example

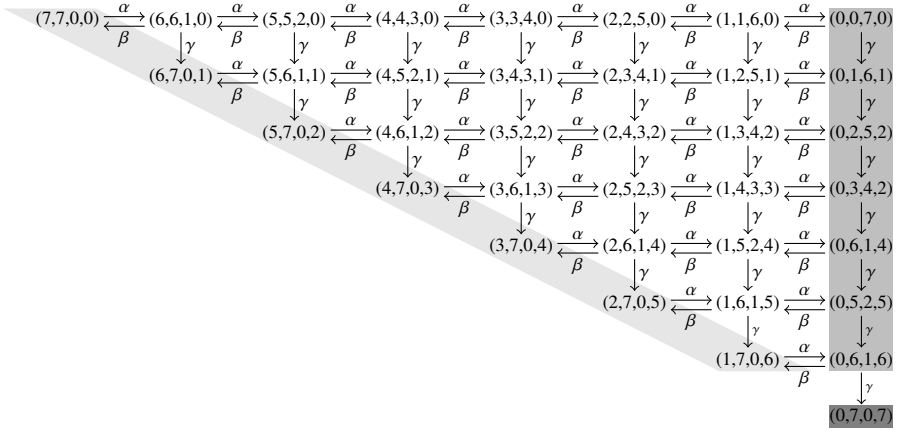

Figure 13: The transition system for *Sys* when $n = 7$

application within other formalisms for systems biology (and within population modelling more generally). In recent years a plethora of description techniques for biochemical systems have been proposed inspired by computer science formalisms. To the best of our knowledge, the only other that explicitly supports discrete models based on levels of concentrations is the Petri net-based modelling framework of Heiner *et al.* [22]. In this approach, places represent species and a token is taken to represent a level of concentration. Within a Petri net, stoichiometry is readily modelled by the multiplicity of arcs, thus this framework similarly supports reactions with stoichiometric coefficients greater than one. Consequently our results should have direct applicability

in this framework also. The rule based description language, Biocham [23], supports multiple interpretations, one of which is boolean models where each species is represented as present or absent, which can be regarded as an extreme abstraction onto just two levels. However, note that this case falls outside the scope of our results since the number of levels will not be sufficient to exhibit all behaviours (cf. Corollary 2). At the other extreme, we can consider the many process algebras which use the abstraction "process-as-molecule". In this case, due to the form of synchronisation used by these formalisms, most cannot support general stoichiometric coefficients [24]. Nevertheless by regarding discretisation levels which are constituted by a single molecule, these process algebras can also give rise to transition systems with levels. Thus compression bisimilarity can equally be applied to such molecular models.

7. 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 [25]. 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.

Laneve and Tarissan [26] define the `bio`-κ-calculus combining ideas from the κ-calculus [2] and brane-based formalisms [5, 27]. They define an operational semantics in which labels on transitions are either protein names decorated with information about binding and rule used, or τ which represents a reaction. Weak bisimulation is shown to be a congruence for the `bio`-κ-calculus with respect to the group operator with creates a solution, and the membrane operator both with respect to the membrane species and the cell species. They extend the calculus to allow for cell splitting and merging and define a context bisimulation that takes into account the transitions representing this interaction and ensuring additional relationships between structures. Stoichiometry is not considered as complexation is only permitted between two proteins.

Semantic equivalence has been used in the comparison of ambient-style models and membrane-style models [28] where a contextual bisimulation similar to that of Laneve

and Tarissan [26] is defined. It is preserved by translation from membrane systems to ambients but not *vice versa* due to differences in contexts and translated contexts.

In the comparison of a term-rewriting calculus, the Calculus of Looping Systems (CLS) [29] and a simple brane calculus PEP [27], labelled transitions semantics are defined as well as semantic equivalences based on strong and weak bisimulation. Neither of these are congruences with respect to CLS. The two forms of bisimulation are also defined for the labelled transitions generated by PEP systems and are both shown to be congruences with respect to these systems. PEP systems can be encoded in CLS. This encoding does not preserve strong bisimulation but does preserve weak bisimulation.

Observational equivalence has been used to show that CCS specifications of elements of lactose operon regulation have the same behaviour as more detailed models [30]. Other related work on semantic equivalences considers a bisimulation parameterised by a function over the context of Bio-PEPA systems [31].

Finally, in an example of biological modelling using hybrid systems, bisimulation over hybrid automata is used to quotient the state space with respect to a subset of variables as a technique for state space reduction [32].

8. Discussion and Further Research

This paper has presented a new semantic equivalence for Bio-PEPA called compression bisimilarity and shown when it is a congruence and when it identifies different discretisations of the same system. It is based on the idea that different discretisations of a system show the same behaviour (within limits) and it is the first equivalence to consider the type of structure that discretisations demonstrate. Furthermore, it is able to account for the structuring of the transition system which is obtained from stoichiometric coefficients. In Section 3, we defined the notion of full behaviour for a transition system with levels, and in Corollary 2, we used the stoichiometry of the reactions a species takes part in to determine full behaviour in the setting of a transition system with levels. Therefore we have captured how to obtain full behaviour within a discretised system.

A biological interpretation of our results can also be considered. As mentioned in

the introduction, congruence can be viewed as the ability to substitute one collection of molecules with another but observe the same behaviour in terms of reactions. It is more difficult to map the conditions that are required for congruence or Hypothesis 1 to biology. These both relate to differences in behaviour that can occur because of the discretisation, and illustrate a lack of monotonicity in behaviour. For example, consider Figures 3 and 4. For odd values, the transition systems obtained are compression bisimilar. Likewise, for even numbers, the transition systems are compression bisimilar. Hence discretisation can lead to slightly differing behaviours, particularly at the edges of transition systems. Since we could choose a discretisation with one molecule per level, logically such discrepancies of behaviour are possible in biological systems with different numbers of molecules. But the granularity of observation is not generally fine enough at present to make such distinctions of behaviour. This raises the question of whether there is a less strict equivalence that ignores these minor differences in transition systems, and this will be explored in further work.

This work opens many avenues of further research. An obvious step is to investigate Hypothesis 1. A possible way to characterise the differences between discretisations with slightly different structures is to consider the dimensions of the central equivalence class, and how they increase as the number of levels increase. We also wish to consider larger biological examples in the future, and these would assist in exploring the hypothesis.

A question of interest is how to ensure that a transition system with levels demonstrates the same behaviour and has the same properties as a continuous model of the system. A Bio-PEPA model with levels can be mapped to a continuous time Markov chain and at the limit, using Kurtz's theorem this CTMC and the ODEs obtained from the Bio-PEPA model have the same behaviour [19]. Moreover, a distance measure has been defined that allows for a empirical methodology to establish the correct step size for obtaining good agreement between the CTMC with levels and the ODEs [19]. Developing an analytic methodology is further research and the full behaviour results for discretised systems may be of use.

Another challenging direction is to extend compression bisimilarity to a quantitative equivalence that takes into account reaction rates. Since rates will change as

number of levels increase, it is not immediately obvious how this can be done. For example, even the intuitive shift result presented in Proposition 3 will not hold in the quantitative setting.

Finally, we wish to extend this equivalence so that it can be used in contexts where we do not have the same reactions names. When dealing with discretisations, we are guaranteed the same names, but for arbitrary systems this is not the case. Hence a relation over reaction names may be necessary. With this extension, we can then investigate applying the equivalence to various biological models.

Acknowledgements: Vashti Galpin is supported by EPSRC Grants EP/E031439/1 and EP/C54370X/01. Jane Hillston is supported by EPSRC ARF EP/C543696/01. The Centre for Systems Biology at Edinburgh is a Centre for Integrative Systems Biology (CISB) funded by the BBSRC and EPSRC in 2006. We thank Maria Luisa Guerriero for her comments and Monika Heiner and Gianfranco Balbo for fruitful discussions.

References

[1] R. Blossey, L. Cardelli, A. Phillips, A compositional approach to the stochastic dynamics of gene networks, in: C. Priami, L. Cardelli, S. Emmott (Eds.), Transactions on Computational Systems Biology IV, LNCS 3939, Springer, 2006, pp. 99–122.

[2] V. Danos, C. Laneve, Formal molecular biology, Theoretical Computer Science 325 (2004) 69–110.

[3] C. Priami, A. Regev, E. Shapiro, W. Silverman, Application of a stochastic name-passing calculus to representation and simulation of molecular processes, Information Processing Letters 80 (2001) 25–31.

[4] C. Priami, P. Quaglia, Beta binders for biological interactions, in: V. Danos, V. Schächter (Eds.), Computational Methods in Systems Biology, International Conference (CMSB 2004), Paris, France, LNCS 3082, Springer, 2005, pp. 20–33.

[5] A. Regev, E. Panina, W. Silverman, L. Cardelli, E. Shapiro, BioAmbients: an abstraction for biological compartments, Theoretical Computer Science 325 (2004) 141–167.

[6] F. Ciocchetta, J. Hillston, Bio-PEPA: a framework for the modelling and analysis of biological systems, Theoretical Computer Science 410 (2009) 3065-3084.

[7] J. Hillston, A compositional approach to performance modelling, CUP, 1996.

[8] F. Ciocchetta, J. Hillston, Calculi for biological systems, in: M. Bernardo, P. Degano, G. Zavattaro (Eds.), Formal Methods for Computational Systems Biology (SFM08), LNCS 5016, 2008, pp. 265–312.

[9] F. Ciocchetta, Bio-PEPA with SBML-like events, in: R.-J. Back, I. Petre (Eds.), Proceedings of the Workshop on Computational Models for Cell Processes, 2008, pp. 11–22.

[10] F. Ciocchetta, J. Hillston, Bio-PEPA: a framework for the modelling and analysis of biological systems, Tech. Rep. EDI-INF-RR-1231, School of Informatics, University of Edinburgh (2008).

[11] K. Ellavarason, An automatic mapping from the Systems Biology Markup Language to the Bio-PEPA process algebra, Master's thesis, University of Trento (2008).

[12] O. E. Akman, F. Ciocchetta, A. Degasperi, M. L. Guerriero, Modelling biological clocks with Bio-PEPA: stochasticity and robustness for the Neurospora Crassa circadian network,, in: P. Degano, R. Gorrieri (Eds.), Computational Methods in Systems Biology, International Conference (CMSB 2009), Bologna, Italy, LNCS 5688, Springer, 2009, pp. 52–67.

[13] M. L. Guerriero, Qualitative and quantitative analysis of a Bio-PEPA model of the gp130/JAK/STAT signalling pathway,, in: C. Priami, R.-J. Back, I. Petre (Eds.), Transactions on Computational Systems Biology XI, LNCS 5750, Springer, 2009, pp. 90–115.

[14] L. Bortolussi, A. Policriti, Hybrid dynamics of stochastic programs, Theoretical Computer Science 411 (2010) 2052–2077.

[15] V. Galpin, J. Hillston, L. Bortolussi, HYPE applied to the modelling of hybrid biological systems, Electronic Notes in Theoretical Computer Science 218 (2008) 33–51.

[16] V. Galpin, J. Hillston, Equivalence and discretisation in Bio-PEPA, in: P. Degano, R. Gorrieri (Eds.), Computational Methods in Systems Biology, International Conference (CMSB 2009), Bologna, Italy, LNCS 5688, Springer, 2009, pp. 189–204.

[17] F. Ciocchetta, M. L. Guerriero, Modelling biological compartments in Bio-PEPA, Electronic Notes in Theoretical Computer Science 227 (2009) 77–95.

[18] A. Duguid, An overview of the Bio-PEPA Eclipse Plug-in, in: Eighth workshop on Process Algebra and Stochastically Time Activities (PASTA), Edinburgh, 2009, pp. 121–132.

[19] F. Ciocchetta, A. Degasperi, J. Hillston, M. Calder, Some investigations concerning the CTMC and the ODE model derived from Bio-PEPA, Electronic Notes in Theoretical Computer Science 229 (2009) 145–163.

[20] R. Milner, Communication and concurrency, Prentice Hall, 1989.

[21] R. J. van Glabbeek, The linear time-branching time spectrum (extended abstract), in: J. C. M. Baeten, J. W. Klop (Eds.), CONCUR '90, LNCS 458, Springer, 1990, pp. 278–297.

[22] M. Heiner, D. Gilbert, R. Donaldson, Petri nets for systems and synthetic biology, in: M. Bernardo, P. Degano, G. Zavattaro (Eds.), Formal Methods for Computational Systems Biology (SFM08), LNCS 5016, Springer, 2008, pp. 215–264.

[23] F. Fages, S. Soliman, Formal cell biology in Biocham, in: M. Bernardo, P. Degano, G. Zavattaro (Eds.), Formal Methods for Computational Systems Biology (SFM08), LNCS 5016, Springer, 2008, pp. 54–80.

[24] M. Calder, J. Hillston, Process algebra modelling styles for biomolecular processes, in: C. Priami, R.-J. Back, I. Petre (Eds.), Transactions on Computational Systems Biology XI, LNCS 5750, Springer, 2009, pp. 1–25.

[25] A. Regev, E. Shapiro, Cellular abstractions: Cells as computation, Nature 419 (2002) 343.

[26] C. Laneve, F. Tarissan, A simple calculus for proteins and cells, Theoretical Computer Science 404 (2008) 127–141.

[27] L. Cardelli, Brane calculi, in: V. Danos, V. Schächter (Eds.), Computational Methods in Systems Biology, International Conference (CMSB 2004), Paris, France, LNCS 3082, Springer, 2005, pp. 257–278.

[28] G. Ciobanu, B. Aman, On the relationship between membranes and ambients, BioSystems (91) (2008) 515–530.

[29] R. Barbuti, A. Maggiolo-Schettini, P. Milazzo, A. Troina, Bisimulation in calculi modelling membranes, Formal Aspects of Computing 20 (2008) 351–377.

[30] M. C. Pinto, L. Foss, J. C. M. Mombach, L. Ribeiro, Modelling, property verification and behavioural equivalence of lactose operon regulation, Computers in Biology and Medicine 37 (2007) 134–148.

[31] V. Galpin, Equivalences for a biological process algebra, in preparation.

[32] M. Antoniotti, C. Piazza, A. Policriti, M. Simeoni, B. Mishra, Taming the complexity of biochemical models through bisimulation and collapsing: theory and practice, Theoretical Computer Science 325 (2004) 45–67.