Equivalences for a biological process algebra

Vashti Galpin

Laboratory for Foundations of Computer Science, School of Informatics
University of Edinburgh, United Kingdom

Abstract

This paper investigates Bio-PEPA, the stochastic process algebra for biological modelling developed by Ciocchetta and Hillston. It focusses on Bio-PEPA with levels where molecular counts are grouped or concentrations are discretised into a finite number of levels. Basic properties of well-defined Bio-PEPA systems are established after which equivalences used for the stochastic process algebra PEPA are considered for Bio-PEPA, and are shown to be identical for well-defined Bio-PEPA systems. Two new semantic equivalences parameterised by functions, called g-bisimilarity and weak g-bisimilarity are introduced. Different functions lead to different equivalences for Bio-PEPA. Congruence is shown for both forms of g-bisimilarity under certain reasonable conditions on the function and the use of these equivalences are demonstrated with a biologically-motivated example where two similar species are treated as a single species, and modelling of alternative pathways in the MAPK kinase signalling cascade.

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

1. Introduction

Biological modelling has an important role in systems biology since it can be used both to model the behaviour seen in the laboratory and to make predictions of behaviours which can then lead to new experimental hypotheses.

Email address: Vashti.Galpin@ed.ac.uk (Vashti Galpin)

Recently, computer science modelling techniques have been applied to the modelling of biological systems [1]. A specific computer science approach that has been applied is that of process algebra [2, 3, 4, 5, 6]. It is an ongoing experiment to determine the worth of process algebra techniques for systems biology modelling [9]. The two biological modelling examples, one related to grouping of species and the other about alternative signalling pathways, presented here are further evidence that this approach is beneficial.

The process algebra considered in this paper, Bio-PEPA [7], is based on the stochastic process algebra PEPA [8]. Stochastic behaviour allows for quantitative models. These models are important in biology since often in understanding what happens, it is necessary to understand how fast or when it happens. Bio-PEPA is used in a reagent-centric way employing the process-as-species metaphor, whereas many other process algebras take the process-as-molecule approach [9]. Bio-PEPA has been used successfully to model and analyse Goldbeter's model of cyclin oscillation [11, 12], the Repressilator [13], the NF-κB signalling pathway [14], the MAPK model [15], circadian clocks [16, 17] and the gp130/JAK/STAT pathway [18].

A strength of Bio-PEPA is that it supports various methods of modelling analysis including ordinary differential equations, stochastic simulation and continuous time Markov chains (CTMCs). To obtain CTMCs, the molecular counts are stratified or concentrations are discretised, resulting in a finite number of levels and a finite labelled transition system. Thus a biological system described as a Bio-PEPA system with levels can be expressed as a finite CTMC, giving access to the many analysis techniques available for this mathematical structure.

Much of the biological modelling done under the framework of process algebras has been focussed on mapping to stochastic simulation and ordinary differential equations (ODEs) and often the labelled transition system and its properties have been overlooked. This paper is an opportunity to consider an approach to biological modelling that uses a stochastic process algebra as a process algebra, rather than as a language front-end to an existing technique. This is achieved by developing an understanding of the labelled transition systems generated by a model and their relationship to other transition systems by comparing behaviour based on the standard process algebra notion of semantic equivalence.

This raises the question of what constitutes the same behaviour in a biological setting. When modelling, it may be convenient (either for model size or because the underlying details are not known) to group some species together. If one can show that a model without the grouping has the same behaviour as the model with the grouping, then there is an argument for working with the simpler model of the two. Likewise, if there are two different sequences of reactions that produce the same species, then to simplify modelling it can be useful to view them as having similar behaviour. This requires the ability to abstract away from the details of some reactions and through this abstraction, be able to focus on the reactions that are relevant. The examples in this paper will illustrate both of these techniques.

To motivate this direction of research, there are at least three (overlapping) approaches to developing equivalences for a biological process algebra:

1. consider equivalences that are useful in Computer Science;
2. investigate how abstraction is used by biologists and develop equivalences based on this; and
3. find different abstractions of a biological model and develop an equivalence that identifies the behaviour found in the two different abstractions.

The second approach is work-in-progress, and currently focusses on abstracting from fast reactions in the style of the Quasi-Steady-State Assumption [20]. The third approach has been taken in the development of compression bisimilarity [21, 22] where two discretisations (both with a sufficient number of levels) of the same Bio-PEPA model are viewed as embodying the same behaviour. Both of these approaches overlap with the first approach since bisimulation-style equivalences are used.

This paper also provides an opportunity to consider Bio-PEPA as a process algebra with levels, rather than specifically as a modelling language for biological examples. Well-defined Bio-PEPA systems have a restricted form and it is important to understand the impact on the labelled transition systems that are obtained.

The structure of this article is as follows: first Bio-PEPA is introduced, and illustrated with a running example. After this presentation, some basic properties of the labelled transition systems obtained from well-defined Bio-PEPA models are proved and motivated. Next, the equivalences defined for PEPA are considered for Bio-PEPA, and it is proved using the properties from the previous section, that some of these equivalences coincide. A new equivalence is defined for Bio-PEPA and it is shown how other equivalences can be expressed in the new equivalence. The conditions under which congruence is obtained for the new equivalence are described and proved. Additionally a weak variant of the equivalence is defined and investigated. Then two examples will be given, after which related work and conclusions will be presented.

2. Bio-PEPA

This section presents Bio-PEPA [7]. The main components of a Bio-PEPA system are the sequential or species components describing the behaviour of each of the chemical species and the model component which combines the species components and hence models the interactions between the species. Additionally, a context is defined to store information such as functional rates, compartments and parameters. The syntax of the sequential/species components is defined by

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

In the prefix term $(\alpha, \kappa)\ \mathsf{op}\ S$, α is an action name from a set of action names $\mathcal{A}$ and can be viewed as the name or label of a reaction, κ is the stoichiometric coefficient[1] of the species and the prefix combinator op represents the role of the element in the reaction. If a species is a reactant in the reaction then $\downarrow$ is used, if a product then $\uparrow$, if an activator then $\oplus$, if an inhibitor then $\ominus$, and $\odot$ is used for a generic modifier. The operator + expresses the choice between two sequential components. Constants, which are species names, are defined to be a specific sequential component using the notation

[1] The stoichiometry/stoichiometric coefficient of a species with respect to a specific reaction is the relative quantity of that species involved in the reaction compared to other species in the reaction. At the molecule level, it decsribes the exact number of molecules. In the reaction $A + 3B \rightarrow 2C + D$, three times as much B as A is consumed to produce as much D as A and twice as much C as A.

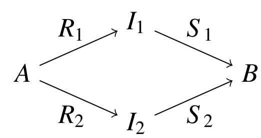

Figure 1: Example reactions

$C \stackrel{def}{=} S$ where C is the species name and S is the sequential component. The set of sequential components is $\mathbf{S}$.

A running example is now presented to illustrate the concepts presented here. This example will also be used later in the paper to demonstrate the new equivalence that will be defined.

Example 1. Consider the reactions defined in Figure 1. There are four species A, I_1, I_2, B, and the reaction rates are given as R_1, R_2, S_1 and S_2. In Bio-PEPA, the species are defined as follows.

$$
\begin{aligned}
A &\stackrel{def}{=} (\alpha_1, 1) \downarrow A + (\alpha_2, 1) \downarrow A \\
I_1 &\stackrel{def}{=} (\alpha_1, 1) \uparrow I_1 + (\beta_1, 1) \downarrow I_1 \\
I_2 &\stackrel{def}{=} (\alpha_2, 1) \uparrow I_2 + (\beta_2, 1) \downarrow I_2 \\
B &\stackrel{def}{=} (\beta_1, 1) \uparrow B + (\beta_2, 1) \uparrow B
\end{aligned}
$$

Here, reactions have been given names, so that the reactions in which A becomes I_i is called α_i and the reactions in which I_i becomes B are called β_i.

The syntax of model components is given by the grammar

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

The process $P \underset{\mathcal{L}}{\bowtie} Q$ denotes the synchronisation between components P and Q and the set $\mathcal{L} \subseteq \mathcal{A}$ specifies those activities on which the components must synchronise. In the model component $S(x)$, the parameter $x \in \mathbb{R}$ represents the molecular count or concentration. The conversion of amounts to discrete levels will be described later. The set of all Bio-PEPA model components is $\mathbf{P}$.

Example 2. The model component for the system described in Figure 1 is

$$M \overset{def}{=} A(n) \underset{\{\alpha_1,\alpha_2\}}{\bowtie} ((I_1(0) \underset{\emptyset}{\bowtie} I_2(0)) \underset{\{\beta_1,\beta_2\}}{\bowtie} B(0))$$

where the initial quantity of A is n, and all other species are absent.

To obtain well-behaved systems, a constrained set of Bio-PEPA model components called well-defined is considered. This ensures that a species is defined as a choice between reactions, and that no reaction name is repeated for a species. At the model level, there can only be one species component for each species. The components defined in Examples 1 and 2 are well-defined.

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

$$C \overset{def}{=} (\alpha_1, \kappa_1)\ \mathsf{op}_1\ C + \ldots + (\alpha_n, \kappa_n)\ \mathsf{op}_n\ C \quad \text{written as} \quad C \overset{def}{=} \sum_{i=1}^{n} (\alpha_i, \kappa_i)\ \mathsf{op}_i\ C$$

where $\alpha_i \neq \alpha_j$ for $i \neq j$. A model component P is *well-defined* if it has the form

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

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 $C_i \neq C_j$.

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

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. The Bio-PEPA model of $\mathcal{P}$ is denoted by $\pi(\mathcal{P}) = P$.

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

The *context* of a Bio-PEPA system is the collection of definition sets $\mathcal{V}$, $\mathcal{N}$, $\mathcal{K}$, $\mathcal{F}$ and *Comp* and these are denoted by $\mathcal{T}$, giving the notation $\langle \mathcal{T}, P \rangle$.

The set of well-defined Bio-PEPA systems is $\tilde{\mathbf{P}}$. The set of well-defined contexts is defined by $\tilde{\mathbf{T}}$.

For details of the other elements of the context and the definition of a well-defined Bio-PEPA context and system, see [7, 11].

This definition allows for many compartments in V but in this paper, the assumptions are only one compartment and a single step size for all species. This ensures that mass is conserved. 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 [23][2].

The model component is typically defined in terms of molecular counts. These can be converted to molar concentrations by dividing by the Avogadro constant and the volume. Depending on the kinetic law, reaction rates may need to be scaled. For a more detailed explanation, see [7, 11]. Both molecular counts and concentrations can be expressed discretised and expressed as levels by definition of a step size. The following definition is based on concentration.

Definition 3. Given a Bio-PEPA system $\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, P \rangle$ with Bio-PEPA model $P \stackrel{def}{=} C_1(x_1) \underset{\mathcal{L}_1}{\bowtie} \ldots \underset{\mathcal{L}_{p-1}}{\bowtie} C_p(x_p)$ where each x_j is a concentration, then the *Bio-PEPA system with levels* is the Bio-PEPA system $\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, P' \rangle$ with Bio-PEPA model $P' \stackrel{def}{=} C_1(l_1) \underset{\mathcal{L}_1}{\bowtie} \ldots \underset{\mathcal{L}_{p-1}}{\bowtie} C_p(l_p)$ where for each species C_j and associated level l_j

- $C_j : H = h, N = n, M = m, V = v, unit = u$ appears in $\mathcal{N}$.
- the number of levels for the species is defined in terms of the maximum concentration and step size, namely $N = \lceil M/H \rceil$.
- C_j has levels, $0, \ldots, N$, giving $N + 1$ levels.
- $l_j = \lceil x_j/H \rceil$.

[2]In the Bio-PEPA Eclipse Plug-in (`www.biopepa.org`), step size is associated with location, not species [24].

As mentioned above, the step size H is assumed to be the same for all species to ensure conservation of mass. A well-defined context for the running example is now given.

Example 3. A well-defined context for the Bio-PEPA model M is

$$
\begin{array}{rcl}
\mathcal{T} & = & \mathcal{V}, \mathcal{K}, \mathcal{F}, \mathcal{N}, Comp \qquad \text{where} \qquad \mathcal{V} = \{v\} \qquad \mathcal{K} = \emptyset \\
\mathcal{F} & = & \{f_{\alpha_1} = fMA(r_1),\ f_{\alpha_2} = fMA(r_2),\ f_{\beta_1} = fMA(s_1),\ f_{\beta_2} = fMA(s_2)\,\} \\
\mathcal{N} & = & \{A : H{=}h, N{=}n, V{=}v;\ I_1 : H{=}h, N{=}n, V{=}v;\ I_2 : H{=}h, N{=}n, V{=}v; \\
 & & \ B : H{=}h, N{=}n, V{=}v;\ \} \\
Comp & = & \{ \text{ model definitions for } A, I_1, I_2, B, \text{ from Example 1 } \}.
\end{array}
$$

Every species has the same step size h and also the same maximum level n. The set of constants $\mathcal{K}$ is empty since the constants r_i and s_i, $i = 1, 2$ have not been defined. Since all reactions are mass action and hence the amount of product is dependent on the amount of reactant, the following definition for the mass action function can be used[3].

$$
fMA(c)[w, \mathcal{N}, \mathcal{K}] = c \prod \{\, lh \mid C{:}{\downarrow}(l, 1) \text{ appears in } w \,\}
$$

Here, l is the level of species C and w is the list obtained from the transition which describes the behaviour of the system.

To describe the behaviour of a Bio-PEPA system, semantics must be defined in terms of the operators of the algebra. The operational semantics for Bio-PEPA systems with levels is given in Figure 2 where N_S is the the maximum number of levels for the species S. These operational semantics define three distinct labelled transition systems. The first, the capability relation, has labels that capture the information about the species that take part in a reaction. The label consists of a reaction name α and a list w that records all the species that took part in the reaction, their current amount and their stoichiometry in the reaction.

Definition 4. Given a Bio-PEPA system, the *capability relation* is

$$
\rightarrow_c \subseteq \mathbf{P} \times \Theta \times \mathbf{P}
$$

[3]This definition only applies in the case that there is at least one reactant and no modifiers, and the stoichiometry of all reactants is 1 but it is sufficient for the example.

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

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

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

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

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

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

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

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

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

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

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

Figure 2: Operational semantics of Bio-PEPA

where $\mathbf{P}$ is the set of Bio-PEPA model components and where $\theta \in \Theta$ is defined by $\theta = (\alpha, w)$ with $\alpha \in \mathcal{A}$, the set of action types and the list w defined by

$$w ::= [\, S : \mathsf{op}\,(l, \kappa)\,] \mid w :: w$$

with $S \in \mathbf{S}$, $l \in \mathbb{N}$, $l \geq 0$ the level, and $\kappa \in \mathbb{N}, \kappa \geq 1$ the stoichiometry coefficient. It is the smallest relation defined by the first nine rules (rules starting with lower case letters) in Figure 2, and an element of the transition system is written $P \xrightarrow{(\alpha,w)}_c P'$.

Note that although w is defined as a list, the order of elements is not important and hence it can be considered as a multiset when convenient.

Example 4. For the example, the following transition is obtained

$$A(n) \underset{\{\alpha_1,\alpha_2\}}{\bowtie} ((I_1(0) \underset{\emptyset}{\bowtie} I_2(0)) \underset{\{\beta_1,\beta_2\}}{\bowtie} B(0)) \xrightarrow{(\alpha_1, A:\downarrow(n,1)::I_1:\uparrow(0,1))}_c$$
$$A(n{-}1) \underset{\{\alpha_1,\alpha_2\}}{\bowtie} ((I_1(1) \underset{\emptyset}{\bowtie} I_2(0)) \underset{\{\beta_1,\beta_2\}}{\bowtie} B(0)).$$

Since this is a well-defined Bio-PEPA model, it is only the levels of each species that change, and vector notation can be used instead to describe models. Hence, the above transition can be written as

$$(n, 0, 0, 0) \quad \xrightarrow{(\alpha_1, A:\downarrow(n,1)::I_1:\uparrow(0,1))}_c \quad (n{-}1, 1, 0, 0)$$

where the vectors represent the quantities of each species in the order that they appear in the composition. Note that transitions are only possible if there are sufficient quantities of the reactants available. For example, $(n{-}3, 0, 3, 0)$ does not have a β_1 transition since there is no I_1.

The second transition system, whose transitions are inferred from those of the capability relation, are labelled with the reaction name and a value which describes the exponential distribution from which the rate of reaction is drawn.

Definition 5. Given a Bio-PEPA system, the *stochastic relation* is

$$\rightarrow_s \subseteq \tilde{\mathbf{P}} \times \Gamma \times \tilde{\mathbf{P}}$$

where $\tilde{\mathbf{P}}$ is the set of well-defined Bio-PEPA systems and where $\gamma \in \Gamma$ has the form (α, r) with $\alpha \in \mathcal{A}$ the set of action types and $r \in \mathbb{R}$ with $r > 0$. It is the smallest relation defined by the rule `Final` in Figure 2 and an element of the transition system is written $\mathcal{P} \xrightarrow{(\alpha,r)}_s \mathcal{P}'$. In 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.

This transition system includes the context and the rate at which the reaction takes place. This rate is functional in the sense that it is specified as a function which takes various arguments both from the context (such as step size and constants) and from the string w which contains information about the current amount of each species involved in the reaction and their stoichiometries for the reaction. The current concentration for a species can be calculated as $l \times H$ where l is the level of the species and H is the step size. For full details of rate derivation, refer to [7, 13].

Example 5. Considering the transition from the previous example,

$$(n, 0, 0, 0) \quad \xrightarrow{(\alpha_1, A{:}\downarrow(n,1){::}I_1{:}\uparrow(0,1))}_c \quad (n{-}1, 1, 0, 0)$$

and applying the rule `Final`, the following is obtained

$$\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, (n, 0, 0, 0)\rangle \xrightarrow{(\alpha_1, fMA(r_1)[A{:}\downarrow(n,1){::}I_1{:}\uparrow(0,1)), \mathcal{N}, \mathcal{K}]/h)}_s$$
$$\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, (n{-}1, 1, 0, 0)\rangle.$$

Using the definition of *fMA*, this gives the transition

$$\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, (n, 0, 0, 0)\rangle \quad \xrightarrow{(\alpha_1, r_1 n)}_s \quad \langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, (n{-}1, 1, 0, 0)\rangle.$$

The last transition system to be defined is the system-capability relation. Note that $\rightarrow_c$ includes the list w which contains information about the reaction that has occurred and the species involved in it and their levels, whereas $\rightarrow_s$ contains the context information but no longer has the information from w. The system-capability relation ensures that all information is available in a single relation.

Definition 6. Given a Bio-PEPA system, the *system-capability relation* is

$$\rightarrow_{sc} \subseteq \tilde{\mathbf{P}} \times \Theta \times \tilde{\mathbf{P}}.$$

It is the smallest relation defined by the rule `Enrich` in Figure 2 and an element of the transition system is written $\mathcal{P} \xrightarrow{(\alpha,r)}_{sc} \mathcal{P}'$.

Example 6. By applying the rule `Enrich` to the transition from Example 4, the following transition is derived

$$\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, (n, 0, 0, 0)\rangle \xrightarrow{(\alpha_1, A:\downarrow(n,1)::I_1:\uparrow(0,1))}_{sc} \langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, (n-1, 1, 0, 0)\rangle.$$

The following two definitions describe the derivative set and the derivative graph for any of the relations. The derivative set is the set of all Bio-PEPA systems that appear in a transition system and the derivative graph is a graph over this set with an edge for every transition in the transition system.

Definition 7. The *derivative set* $ds_x(E)$ (or $ds(E)$ if not ambiguous) is the smallest set such that $E \in ds_x(E)$ and if $E' \in ds(E)$ and $E' \xrightarrow{(\alpha,r)}_x E''$ then $E'' \in ds_x(E)$.

Definition 8. The *derivative graph* $\mathcal{D}_x(E)$ (or $\mathcal{D}(E)$ if not ambiguous) is the labelled directed graph whose nodes are $ds_x(E)$ and whose set of edges are the elements of $\rightarrow_x$.

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[4].

[4]Note that this definition is slightly different to that in [7] since it takes into account the stoichiometry and current level.

Definition 9. The set of *current actions* enabled in $\langle \mathcal{T}, P\rangle$ is defined as $\mathcal{A}(\langle \mathcal{T}, P\rangle) = \mathcal{A}(P)$ where N_S is the maximum number of levels for species component S.

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

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 proper subset of this if the current level is insufficient to satisfy the constraints imposed by the stoichiometry.

Example 7.

$$\mathcal{A}(A(n) \bowtie_{\{\alpha_1,\alpha_2\}} ((I_1(0) \bowtie_{\emptyset} I_2(0)) \bowtie_{\{\beta_1,\beta_2\}} B(0)) = \{\alpha_1, \alpha_2\}$$

Finally, definitions are required to reason about the contents of the structure (α, w) that appears on the $\rightarrow_c$ and $\rightarrow_{sc}$ transitions.

Definition 10. For $\theta \in \Theta$ with $\theta = (\alpha, w)$, $\mathrm{action}(\theta) = \alpha$, $\mathrm{list}(\theta) = w$ and

$$
\begin{aligned}
\mathrm{reacts}(\theta) &= \{S \mid S\!\downarrow\!(l,\kappa) \text{ appears in } w\} & \#\mathrm{reacts}(\theta) &= |\,\mathrm{reacts}(\theta)\,|\\
\mathrm{prods}(\theta) &= \{S \mid S\!\uparrow\!(l,\kappa) \text{ appears in } w\} & \#\mathrm{prods}(\theta) &= |\,\mathrm{prods}(\theta)\,|\\
\mathrm{mods}(\theta) &= \{S \mid S \odot (l,\kappa) \text{ appears in } w\} & \#\mathrm{mods}(\theta) &= |\,\mathrm{mods}(\theta)\,|\\
\mathrm{enzs}(\theta) &= \{S \mid S \oplus (l,\kappa) \text{ appears in } w\} & \#\mathrm{enzs}(\theta) &= |\,\mathrm{enzs}(\theta)\,|\\
\mathrm{inhibs}(\theta) &= \{S \mid S \ominus (l,\kappa) \text{ appears in } w\} & \#\mathrm{inhibs}(\theta) &= |\,\mathrm{inhibs}(\theta)\,|\\
\mathrm{totMods}(\theta) &= \mathrm{mods}(\theta) \cup \mathrm{enzs}(\theta) \cup \mathrm{inhibs}(\theta) & \#\mathrm{totMods}(\theta) &= |\,\mathrm{totMods}(\theta)\,|
\end{aligned}
$$

This section has defined Bio-PEPA syntax and semantics for Bio-PEPA systems with levels, as well as some auxiliary definitions. The next section considers differences between PEPA and Bio-PEPA plus basic results about Bio-PEPA systems

3. Results about well-defined Bio-PEPA systems

First, it is important to note that the operational semantics define labelled transition systems and not labelled multi-transition systems as in PEPA. Hence if two different derivation trees generate the same transition in Bio-PEPA, there is only one transition and not two as there would be in PEPA. In PEPA, the multiplicity of transitions was required to obtain the correct rate. Since rates are calculated by a different mechanism in Bio-PEPA, this is not a concern. Hence multisets are not required when considering transitions or other definitions in Bio-PEPA.

Second, the operational semantics do not define any transitions for sequential components that have no level suffix – hence the components (α, κ) $\mathtt{op}$ S, $S_1 + S_2$ and C cannot perform any transitions whereas the components $((\alpha, \kappa)$ $\mathtt{op}$ $S)(l)$, $(S_1 + S_2)(l)$ and $C(l)$ may be able to. In effect, transitions are only defined for model components.

Next, results about Bio-PEPA models and systems are presented. The first result shows that well-definedness is preserved by transitions.

Lemma 1. *Let P be a Bio-PEPA model such that $P \xrightarrow{(\alpha,w)}_c P'$. Then if P is well-defined then so is P'.*

Proof. At the species level, any transition has the form $C(l) \xrightarrow{(\alpha,w)}_c C(l')$. This means that at the model level a transition has the form

$$C_1(l_1) \underset{\mathcal{L}_1}{\bowtie} \ldots \underset{\mathcal{L}_{p-1}}{\bowtie} C_p(l_p) \xrightarrow{(\alpha,w)}_c C_1(l'_1) \underset{\mathcal{L}_1}{\bowtie} \ldots \underset{\mathcal{L}_{p-1}}{\bowtie} C_p(l'_p)$$

showing that well-definedness is preserved. □

This result can clearly be repeated for Bio-PEPA systems with respect to the two transition relations $\rightarrow_s$ and $\rightarrow_{sc}$ since $\mathcal{T}$ is not changed in a transition. In the previous section, the object w appearing on the capability relation was defined to be a list or a multiset (since order was not important). The next lemma shows that for well-defined systems, it is a set.

Proposition 1. *Let P be a well-defined Bio-PEPA model. If $P \xrightarrow{(\alpha,w)}_c P'$ then w is a set.*

Proof. Consider $P \xrightarrow{(\alpha,w)}_c P'$ with w a multiset and not a set. Then there exists an $S\colon \mathtt{op}(l,\kappa)$ that appears twice (or more) in w. Hence by inspection of the operational semantics, at some point of the derivation of the transition, there exists a transition $Q_1 \underset{\mathcal{L}}{\bowtie} Q_2 \xrightarrow{\alpha, w_1::w_2}_c Q'$ with $S\colon \mathtt{op}(l,k)$ in both w_1 and w_2. Hence in Q_1 there is a $C_i \overset{\mathit{def}}{=} \ldots + (\alpha,\kappa)\ \mathtt{op}\ S + \ldots$ and in Q_2 there is a $C_j \overset{\mathit{def}}{=} \ldots + (\alpha,\kappa)\ \mathtt{op}\ S + \ldots$. By definition of well-definedness, S is C_i and S is C_j hence $i = j$. But this contradicts well-definedness, since a sequential component can only appear once in P. $\square$

Next, the impossibility of two transitions from the same model labelled with the same reaction name is shown.

Proposition 2. *Let P be a well-defined Bio-PEPA model. If $P \xrightarrow{(\alpha,w_1)}_c P'$ and $P \xrightarrow{(\alpha,w_2)}_c P''$ are two transitions, then $w_1 = w_2$ and $P' = P''$.*

Proof. Consider $P \overset{\mathit{def}}{=} C_1(l_1) \underset{\mathcal{L}_1}{\bowtie} \ldots \underset{\mathcal{L}_{m-1}}{\bowtie} C_m(l_m)$ with $P' \overset{\mathit{def}}{=} C_1(l'_1) \underset{\mathcal{L}_1}{\bowtie} \ldots \underset{\mathcal{L}_{m-1}}{\bowtie} C_m(l'_m)$ and $P'' \overset{\mathit{def}}{=} C_1(l''_1) \underset{\mathcal{L}_1}{\bowtie} \ldots \underset{\mathcal{L}_{m-1}}{\bowtie} C_m(l''_m)$. Consider any C_i involved in the transition. By well-definedness C_i has exactly one subterm of the form $(\alpha,\kappa).C_i$ and hence only one transition is possible for C_i, namely $C_i(l_i) \xrightarrow{(\alpha,[C_i:\mathtt{op}(l_i,\kappa_i)])}_c C_i(l)$ so $l = l'_i = l''_i$ and therefore $P' = P''$. This transition appears in the derivation tree of the final transition and contributes $C_i\colon \mathtt{op}(l_i,\kappa_i)$ to w_1 and w_2 hence, considering every species that takes part in the reaction, $w_1 = w_2$. $\square$

If it is not possible to have two transitions with the same reaction name, under what conditions is it possible to have two transitions between the same two processes with different reaction names?

Proposition 3. *Let P be a well-defined Bio-PEPA model. If $P \xrightarrow{(\alpha_1,w_1)}_c P'$ and $P \xrightarrow{(\alpha_2,w_2)}_c P'$ with $\alpha_1 \neq \alpha_2$ then $\mathit{reacts}((\alpha_1,w_1)) = \mathit{reacts}((\alpha_2,w_2))$ and $\mathit{prods}((\alpha_1,w_1)) = \mathit{prods}((\alpha_2,w_2))$.*

Proof. P is well-defined and can be written as $P \overset{\mathit{def}}{=} C_1(l_1) \underset{\mathcal{L}_1}{\bowtie} \ldots \underset{\mathcal{L}_{m-1}}{\bowtie} C_m(l_m)$. Since $P \xrightarrow{(\alpha_1,w_1)}_c P'$ by the operational semantics, $P' \overset{\mathit{def}}{=} C'_1(l'_1) \underset{\mathcal{L}_1}{\bowtie} \ldots \underset{\mathcal{L}_{m-1}}{\bowtie} C'_m(l'_m)$. Each C_i is well-defined, so $C_i = C'_i$, however l_i and l'_i may differ.

Let $C_i \in \mathit{reacts}((\alpha_1,w_1))$, hence $C_i{:}{\downarrow}(l_i,\kappa) \in w_1$, so in the derivation of $P \xrightarrow{(\alpha_1,w_1)}_c P'$, there is a transition $C_i(l_i) \xrightarrow{(\alpha_1,[C_i:\downarrow(l_i,\kappa)])}_c C_i(l_i - \kappa)$ with $\kappa > 0$. Both transitions result in

P', so there must be a similar transition $C_i(l_i) \xrightarrow{(\alpha_2,[C_i:\downarrow(l_i,\kappa)])}_c C_i(l_i - \kappa)$ in the derivation tree of $P \xrightarrow{(\alpha_2,w_2)}_c P'$, since this is the only way for the level to increase from l_i to $l_i - \kappa$. Hence $C_i \in reacts((\alpha_2, w_2))$. Likewise $prods((\alpha_1, w_1)) = prods((\alpha_2, w_2))$ can be proved in a similar fashion. $\square$

It is not possible to obtain a similar result for *mods*, *enzs*, *inhibs* and *totMods* since the level of a modifier does not change and hence it is not possible to determine which modifiers (if any) have been involved in obtaining P'.

By Proposition 2, it is not possible to have two distinct transitions $P \xrightarrow{(\alpha,w_1)}_c P'$ and $P \xrightarrow{(\alpha,w_2)}_c P'$ and this fact can be used to reason about the stochastic relation. If, in a model, it is necessary to have two reactions with the same reactants and the same product but where one involves in a modifier, these reactions must have different names. This is a result of the way in which the cooperation operator is defined. If a reaction name appears in the cooperation set, all species capable of that reaction must take part in the reaction.

Proposition 4. *Let $\mathcal{P}$ be a well-defined Bio-PEPA system. If $\mathcal{P} \xrightarrow{(\alpha,r_1)}_s \mathcal{P}_1$ and $\mathcal{P} \xrightarrow{(\alpha,r_2)}_s \mathcal{P}_2$ are two transitions then $r_1 = r_2$ and $\mathcal{P}_1 = \mathcal{P}_2$*

Proof. If $\mathcal{P}$ is well-defined, then $\pi(\mathcal{P})$, the model in the system $\mathcal{P}$, is a well-defined model component. By Proposition 2 and application of the rule `Final`, the result follows. $\square$

Since equivalences are the focus of this paper, pairs of Bio-PEPA systems will be compared. It is convenient to consider contexts that include all the definitions for both systems.

Without going into too much tedious syntactic detail, the following can be stated where "name" refers to the part of a definition before the "=" symbol for any definition appearing in the context.

Definition 11. The context, $\mathcal{V}$, $\mathcal{N}$, $\mathcal{K}$, $\mathcal{F}$ and *Comp*, of a Bio-PEPA system is *well-defined with respect to names* if each name only appears once within the context.

Definition 12. $\langle \mathcal{V}_1, \mathcal{N}_1, \mathcal{K}_1, \mathcal{F}_1, Comp_1, P_1 \rangle$ and $\langle \mathcal{V}_2, \mathcal{N}_2, \mathcal{K}_2, \mathcal{F}_2, Comp_2, P_2 \rangle$ are *compatible* if $\mathcal{V}_1 \cup \mathcal{V}_2$, $\mathcal{N}_1 \cup \mathcal{N}_2$, $\mathcal{K}_1 \cup \mathcal{K}_2$, $\mathcal{F}_1 \cup \mathcal{F}_2$ and $Comp_1 \cup Comp_2$ are well-defined with respect to names.

Definition 13. Given two compatible systems $\mathcal{P}_1 = \langle \mathcal{V}_1, \mathcal{N}_1, \mathcal{K}_1, \mathcal{F}_1, Comp_1, P_1 \rangle$ and $\mathcal{P}_2 = \langle \mathcal{V}_2, \mathcal{N}_2, \mathcal{K}_2, \mathcal{F}_2, Comp_2, P_2 \rangle$, then $\mathcal{V}$, $\mathcal{N}$, $\mathcal{K}$, $\mathcal{F}$ and $Comp$ forms a *covering context* for $\mathcal{P}_1$ and $\mathcal{P}_2$ if $\mathcal{V}_1 \cup \mathcal{V}_2 \subseteq \mathcal{V}$, $\mathcal{N}_1 \cup \mathcal{N}_2 \subseteq \mathcal{N}$, $\mathcal{K}_1 \cup \mathcal{K}_2 \subseteq \mathcal{K}$, $\mathcal{F}_1 \cup \mathcal{F}_2 \subseteq \mathcal{F}$ and $Comp_1 \cup Comp_2 \subseteq Comp$.

Given two Bio-PEPA systems $\langle \mathcal{T}_1, P_1 \rangle$ and $\langle \mathcal{T}_2, P_2 \rangle$ which are not compatible, then by modifying names in $\langle \mathcal{T}_2, P_2 \rangle$ giving $\langle \mathcal{T}_2', P_2' \rangle$ it is possible to make them compatible. For the rest of the document, compatibility is assumed without explicit modification of names.

The notation $\mathcal{T}(\mathcal{P})$ will be used for $\{\mathcal{P} \mid \mathcal{P} = \langle \mathcal{T}, P \rangle \text{ and } \mathcal{P} \in \tilde{\mathbf{P}}\}$, namely any well-defined Bio-PEPA system with covering context $\mathcal{T}$.

To consider these results in a biological context, first note that the definition of well-defined imposes conditions on species and models to make them biologically realistic. For example, reaction names in species must be distinct to reflect the fact that for a specific reaction, there is only one way for a species to be involved in that reaction. Moreover, species syntax is constrained so that species can be involved in reactions but cannot suddenly become another species. The results in this section come from these basic principles and capture the idea that we do not distinguish specific molecules or concentrations within a species. Hence in both transition systems, there is exactly one transition for a specific reaction capturing the idea that it is not necessary to care about the identity of the individual items that make up a species. This abstraction is beneficial for modelling biological systems because it avoids something that can lead to exponential blow-up of the transition system. The last part of this section has considered formally how to name the various components in a Bio-PEPA model to remove possible conflicts in naming and this aids in building well-behaved models which can be combined together, thereby supporting an important aspect of systems biology.

This section has covered the basic results that will be used in the rest of the paper.

Next, the issue of equivalences is considered. For the rest of the paper, unless explicitly mentioned, Bio-PEPA models and system will be assumed to be well-defined.

4. Equivalences for Bio-PEPA based on PEPA equivalences

In PEPA, there are three distinct equivalences [8]. Isomorphism is a very strict equivalence that requires the structure of transition systems to be identical. Strong bisimilarity considers actions and rates based on the standard definition of bisimilarity. Strong equivalence considers actions and rates to equivalence classes in the style of Larsen and Skou [25].

In the following subsections, equivalences over both the stochastic and capability relation as well as equivalences induced from Bio-PEPA models to Bio-PEPA equivalences are considered.

4.1. Equivalences over the stochastic relation

In this section, three equivalences over the stochastic relation are considered and applied to Bio-PEPA since the stochastic relation is most similar to the transition relation of PEPA (the difference is whether multi-transitions are used). For the remainder of this section, consider two Bio-PEPA systems with a covering context,

$$\mathcal{P}_1 = \langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, P_1 \rangle \qquad \text{and} \qquad \mathcal{P}_2 = \langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, P_2 \rangle.$$

Hence $\pi(\mathcal{P}_1) = P_1$ and $\pi(\mathcal{P}_2) = P_2$.

In the definitions of these equivalences for PEPA, two auxiliary definitions are used. One is the multiset of the current activities. As mentioned above, there is no multiplicity of transitions in Bio-PEPA and hence $\mathcal{A}(.)$ is suitable to describe the current set of activities. The other definition is the apparent rate[5] for a particular action. By Proposition 4, in well-defined Bio-PEPA systems there is only one transition from a system with a specific action name, and hence only one way in which that action can occur. Hence, apparent rate calculations are not necessary.

[5]The apparent rate is a calculation of the rate for an action α taking into account all possible ways in which this action can occur, and is dependent on the syntax of the model performing α. See [8] for details.

The first equivalence to be defined is isomorphism[6].

Definition 14. A function $\mathcal{F} : ds(\mathcal{P}_1) \rightarrow ds(\mathcal{P}_2)$ is a *system isomorphism* between Bio-PEPA systems $\mathcal{P}_1$ and $\mathcal{P}_2$ if $\mathcal{F}$ is an injective function and for any system $\mathcal{P}'_1$, $\mathcal{A}(\mathcal{P}'_1) = \mathcal{A}(\mathcal{F}(\mathcal{P}'_1))$ and for all $(\alpha, r) \in \Gamma$

$$\{\mathcal{P}'_2 \mid \mathcal{F}(\mathcal{P}'_1) \xrightarrow{(\alpha,r)}_s \mathcal{P}'_2\} = \{\mathcal{F}(\mathcal{P}''_1) \mid \mathcal{P}'_1 \xrightarrow{(\alpha,r)}_s \mathcal{P}''_1\}$$

Definition 15. Two components, $\mathcal{P}_1$ and $\mathcal{P}_2$ are *system isomorphic*, $\mathcal{P}_1 =_s \mathcal{P}_2$, if there exists a system isomorphism $\mathcal{F}$ between them such that $\mathcal{D}(\mathcal{F}(\mathcal{P}_1)) = \mathcal{D}(\mathcal{P}_2)$.

Strong bisimilarity is defined for PEPA, and this definition is now given in terms of the stochastic relation. In the PEPA definition of this equivalence, there is a clause requiring that the apparent rates of the two components be equal. For the same reason as given above, because there is at most one transition from a model with a given action name, this is not required.

Definition 16. A binary relation $\mathcal{R}$ over systems is a *system bisimulation* if $(\mathcal{P}_1, \mathcal{P}_2) \in \mathcal{R}$ implies that for all $\alpha \in \mathcal{A}$

1. whenever $\mathcal{P}_1 \xrightarrow{(\alpha,r)}_s \mathcal{P}'_1$ then for some $\mathcal{P}'_2$, $\mathcal{P}_2 \xrightarrow{(\alpha,r)}_s \mathcal{P}'_2$ and $(\mathcal{P}'_1, \mathcal{P}'_2) \in \mathcal{R}$
2. whenever $\mathcal{P}_2 \xrightarrow{(\alpha,r)}_s \mathcal{P}'_2$ then for some $\mathcal{P}'_1$, $\mathcal{P}_1 \xrightarrow{(\alpha,r)}_s \mathcal{P}'_1$ and $(\mathcal{P}'_1, \mathcal{P}'_2) \in \mathcal{R}$

As is standard, it can be shown that a system bisimulation is an equivalence relation. The semantic equivalence is defined to be the largest system bisimulation, in the usual manner.

Definition 17. $\mathcal{P}_1$ and $\mathcal{P}_2$ are *system bisimilar*, $\mathcal{P}_1 \sim_s \mathcal{P}_2$, if $(\mathcal{P}_1, \mathcal{P}_2) \in \mathcal{R}$ for some system bisimulation $\mathcal{R}$. $\sim_s = \bigcup\{\mathcal{R} \mid \mathcal{R}$ is a system bisimulation$\}$.

Finally, a definition based on the strong equivalence of PEPA can be made.

[6]The names of these equivalences are the same as those in [7] although the definitions differ. This is addressed in Section 4.4.

Definition 18. An equivalence relation $\mathcal{R}$ over $\tilde{\mathcal{P}}$ is a *system equivalence* if whenever $(\mathcal{P}_1, \mathcal{P}_2) \in \mathcal{R}$ then for $\alpha \in \mathcal{A}$ and for all $S \in \tilde{\mathcal{P}}/\mathcal{R}$,

$$\sum \{r \mid \mathcal{P}_1 \xrightarrow{(\alpha,r)} \mathcal{P}'_1, \mathcal{P}'_1 \in S\} = \sum \{s \mid \mathcal{P}_2 \xrightarrow{(\alpha,s)} \mathcal{P}'_2, \mathcal{P}'_2 \in S\}.$$

Definition 19. $\mathcal{P}_1$ and $\mathcal{P}_2$ are *system equivalent*, $\mathcal{P} \cong_s \mathcal{P}_2$, if $(\mathcal{P}_1, \mathcal{P}_2) \in \mathcal{R}$ for some system equivalence $\mathcal{R}$. $\cong_s = \bigcup\{\mathcal{R} \mid \mathcal{R}$ is a system equivalence$\}$.

This definition is the most interesting of the three equivalences defined for PEPA since it induces a (strongly or ordinarily) lumpable partition on the underlying CTMC [8]. Hence, it would be desirable to prove some properties such as congruence. However, before doing this, a comparison of the three equivalences is necessary and due to the constrained nature of a well-defined Bio-PEPA component, it can be shown that all of the above semantic equivalences are identical.

Theorem 1. *Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two well-defined Bio-PEPA systems, then*

$$\mathcal{P}_1 =_s \mathcal{P}_2 \quad \Leftrightarrow \quad \mathcal{P}_1 \sim_s \mathcal{P}_2 \quad \Leftrightarrow \quad \mathcal{P}_1 \cong_s \mathcal{P}_2.$$

Proof. Using Proposition 4 there is exactly one transition of the form $\mathcal{P} \xrightarrow{(\alpha,r)}$ for a given $\mathcal{P}$. This fact can be used to show that a system bisimulation and a system equivalence can be transformed into an isomorphism. □

With this result, congruence would need only to be proved for one of these semantic equivalences.

4.2. Equivalences over the capability relation

The labels of the capability relation have a more complex structure than those of the stochastic relation and hence there are more possibilities for equivalences since it is possible to ignore parts of the structure when defining an equivalence, as will be demonstrated in Section 5. The step taken in the current section is to consider equivalences over the capability relation, matching the first three defined above based on PEPA equivalences, and these consider the full label.

Capability isomorphism is defined in the same manner as system isomorphism, but instead of requiring a match on (α, r), the match is required on (α, w) since this is the form the labels take in the capability relation. Moreover, it is defined over Bio-PEPA models rather than Bio-PEPA systems.

Definition 20. A function $\mathcal{F} : ds(P_1) \to ds(P_2)$ is a *capability isomorphism* between P_1 and P_2 if $\mathcal{F}$ is an injective function and for any component P', $\mathcal{A}(P') = \mathcal{A}(\mathcal{F}(P'))$ and for all (α, w),

$$\{P_2' \mid \mathcal{F}(P') \xrightarrow{(\alpha,w)}_c P_2'\} = \{\mathcal{F}(P'') \mid P' \xrightarrow{(\alpha,w)}_c P''\}$$

Definition 21. Two components, P_1 and P_2 are *capability isomorphic*, $P_1 =_c P_2$, if there exists a capability isomorphism $\mathcal{F}$ between them such that $\mathcal{D}(\mathcal{F}(P_1)) = \mathcal{D}(P_2)$.

Again, capability bisimulation has a similar definition to system bisimulation except for the label structure and the use of Bio-PEPA models instead of systems.

Definition 22. A binary relation $\mathcal{R}$ over systems is a *capability bisimulation* if $(P_1, P_2) \in R$ implies that for all $\alpha \in \mathcal{A}$

1. whenever $P_1 \xrightarrow{(\alpha,w)}_c P_1'$ then for some P_2', $P_2 \xrightarrow{(\alpha,w)}_c P_2'$ and $(P_1', P_2') \in \mathcal{R}$
2. whenever $P_2 \xrightarrow{(\alpha,w)}_c P_2'$ then for some P_1', $P_1 \xrightarrow{(\alpha,w)}_c P_1'$ and $(P_1', P_2') \in \mathcal{R}$

Definition 23. P_1 and P_2 are *capability bisimilar*, $P_1 \sim_c P_2$, is $(P_1, P_2) \in \mathcal{R}$ for some capability bisimulation $\mathcal{R}$. $\sim_c = \bigcup\{\mathcal{R} \mid \mathcal{R}$ a capability bisimulation$\}$.

This has been defined without reference to rates since rate values cannot be calculated in this context. It is not clear how to define an equivalence like the system equivalence in the previous section as there is no obvious way to calculate rates, hence there is no reason to proceed with this. The two equivalences that have been defined can be compared.

Theorem 2. *Let P_1 and P_2 be two well-defined Bio-PEPA components. Then*

$$P_1 =_c P_2 \quad \Leftrightarrow \quad P_1 \sim_c P_2.$$

Proof. By Proposition 2 there is at most one α-labelled transition between any two components, and hence one can translate a bisimulation to the appropriate isomorphism. □

These two equivalences are very strict, and require an exact match of the list w. Considering the example from Section 2, and the system

$$I_1(0) \underset{\{\alpha_1,\beta_1\}}{\bowtie} ((A(n) \underset{\{\alpha_2\}}{\bowtie} I_2(0)) \underset{\{\beta_2\}}{\bowtie} B(0))$$

it can be shown that they are not capability bisimilar since the label on the only transition from this new model is $(\alpha_1, I_1{:}\uparrow(0,1){::}A{:}\downarrow(n,1))$ which differs from the label on the transition in the original example. If w is treated as a set, then this removes this problem (and allows some standard properties such as commutativity to be proved). However, it is still a strict equivalence because it requires the same species names to appear, with the same prefix, the same level and the same stoichiometry.

It is difficult to think of a non-pathological example where systems of different species might be equated by this equivalence. At best, there are equations such as $M \sim_c M \underset{\emptyset}{\bowtie} D(0)$ where D is a species which can only act as a reactant, and shares no reaction names with any of the species in M. In light of this strictness, in Section 5 less strict equivalences are considered over the capability relation by introducing a function over the labels of the transitions.

4.3. Induced equivalences

There are two transition relations for which equivalences are defined in the style of Bio-PEPA. An equivalence over model components can lead to one over Bio-PEPA systems in the following way.

Definition 24. Let $\equiv$ be an equivalence over model components. Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two Bio-PEPA systems with a covering context $\mathcal{T}$. The induced equivalence $\equiv^I$ is defined as $\mathcal{P}_1 \equiv^I \mathcal{P}_2$ if $\pi(\mathcal{P}_1) \equiv \pi(\mathcal{P}_2)$

Since there is a covering context $\mathcal{T}$, results can be proved that relate the induced equivalences with those directly defined on the stochastic relation.

Proposition 5. *Let $\mathcal{P}_i = \langle \mathcal{V}_i, \mathcal{N}_i, \mathcal{K}_i, \mathcal{F}_i, Comp, P_i \rangle$, $i = 1, 2$ be two well-defined Bio-PEPA systems with a covering context $\mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp$. Then*

$$\mathcal{P}_1 =_s \mathcal{P}_2 \quad \Leftrightarrow \quad \mathcal{P}_1 =_c^I \mathcal{P}_2$$

Proof. Consider two systems $\langle \mathcal{T}, P_1 \rangle$ and $\langle \mathcal{T}, P_2 \rangle$ with $P_1 =_c P_2$ and consider a transition $P_1 \xrightarrow{(\alpha,w)} P'_1$ whose image under the isomorphism is the transition $P_2 \xrightarrow{(\alpha,w)} P'_2$.

The use of the rule `Final` gives two transitions $\mathcal{P}_1 \xrightarrow{(\alpha,r_1)} \mathcal{P}'_1$ and $\mathcal{P}_2 \xrightarrow{(\alpha,r_2)} \mathcal{P}'_2$ where $\mathcal{P}_i = \langle \mathcal{T}, P_i \rangle$ and $\mathcal{P}'_i = \langle \mathcal{T}, P'_i \rangle$. Then

$$r_1 = r_\alpha[w, \mathcal{N}, \mathcal{K}] = f_\alpha[w, \mathcal{N}, \mathcal{K}]/h_1 \quad \text{and} \quad r_2 = r_\alpha[w, \mathcal{N}, \mathcal{K}] = f_\alpha[w, \mathcal{N}, \mathcal{K}]/h_2.$$

Expanding this by using the definition of f_α,

$$r_1 = f_\alpha(k_1, \ldots, k_m, C_1, \ldots, C_n)[w, \mathcal{N}, \mathcal{K}]/h_1 = f_\alpha(k_1, \ldots, k_m, (l_1 h_1)^{\kappa_1}, \ldots, (l_n h_1)^{\kappa_n})/h_1$$

$$r_2 = f_\alpha(k_1, \ldots, k_m, C_1, \ldots, C_n)[w, \mathcal{N}, \mathcal{K}]/h_2 = f_\alpha(k_1, \ldots, k_m, (l_1 h_2)^{\kappa_1}, \ldots, (l_n h_2)^{\kappa_n})/h_2$$

Note that it must be the case that $\{C_1, \ldots, C_n\} \subseteq \mathit{component}(P_1) \cap \mathit{component}(P_2)$ otherwise the function will not be defined. The same constant values from $\mathcal{K}$ are used for both r_1 and r_2. Since the values l_i and κ_i for each component C_i come from the string w and this is the same for r_1 and r_2, these are identical, so the only difference can occur is between h_1 and h_2. All components in $\mathcal{P}_1$ have the same step size, and all components in $\mathcal{P}_2$ have the same step size so it remains to show that $h_1 = h_2$. By the definition of a covering context, the two contexts must be compatible, so names are unique, therefore the definitions of h_1 and h_2 which use the same name (in fact names, because step size is recorded for all components) must be the same value.

Therefore $r_1 = r_2$, hence there is a one-to-one correspondence between transitions under $\rightarrow_c$ and $\rightarrow_s$ giving the result. $\square$

Proposition 6. *Let $\mathcal{P}_i = \langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, P_i \rangle$, $i = 1, 2$ be two Bio-PEPA systems with the step size h the same across all components. Then*

$$\mathcal{P}_1 \sim_s \mathcal{P}_2 \quad \Leftrightarrow \quad \mathcal{P}_1 \sim_c^I \mathcal{P}_2$$

Proof. Similar to previous proof. $\square$

Corollary 1. *Let $\mathcal{P}_i = \langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, P_i \rangle$, $i = 1, 2$ be two Bio-PEPA systems with the step size h the same across all components. Then*

$$\mathcal{P}_1 =_c^I \mathcal{P}_2 \quad \Leftrightarrow \quad \mathcal{P}_1 \sim_c^I \mathcal{P}_2$$

Note that transitions that match on the reaction name α will use the same f_α which will obviously constrain how the rate can vary. Later in this document, an equivalence will be defined where this is not the case.

4.4. Existing equivalences defined for Bio-PEPA

Equivalences are defined in the original Bio-PEPA paper [7]. These differ from those defined here and this section discusses these differences.

First consider the two isomorphisms in [7]. They both contain a condition of the form $\mathcal{A}(X) = \mathcal{A}(\mathcal{F}(X))$ which refers to a different definition of $\mathcal{A}(.)$ to the one given in this current paper, and the definition in [7] does not take the current level into account. This means that all actions (whether possible or not) must appear in both systems. In the definition in the current paper, only actions that are possible with respect to the current level are considered. This makes the isomorphisms in the current paper more general in the sense that they equate slightly more models and systems in a more standard manner. For the definition of $\mathcal{A}(.)$ given here, when applied to $A(0) \underset{\{\alpha_1,\alpha_2\}}{\bowtie} ((I_1(n_1) \underset{\emptyset}{\bowtie} I_2(n_2)) \underset{\{\beta_1,\beta_2\}}{\bowtie} B(0))$, $\{\beta_1, \beta_2\}$ is obtained. Using the definition from [7], it would give $\{\alpha_1, \alpha_2, \beta_1, \beta_2\}$. This means that under the definition from [7], $A(0) \underset{\{\alpha_1,\alpha_2\}}{\bowtie} ((I_1(n_1) \underset{\emptyset}{\bowtie} I_2(n_2)) \underset{\{\beta_1,\beta_2\}}{\bowtie} B(0))$ and $(I_1(n_1) \underset{\emptyset}{\bowtie} I_2(n_2)) \underset{\{\beta_1,\beta_2\}}{\bowtie} B(0)$ would not be system isomorphic although it is reasonable to expect them to be.

Additionally, in the definition of component isomorphism in [7], the isomorphism function is applied to the string w but this is not necessary for a reasonable or standard definition of isomorphism, as is given by capability isomorphism in the current paper. Moreover, this application of the isomorphism to w requires it to be extended so that it is defined over components, although this is not done in [7]. Later in this paper, it will be shown how it is possible to map between components in equivalence definitions.

The two strong capability bisimilarity definitions in the original Bio-PEPA paper abstract away from the details of the string w, either by ignoring it altogether or by matching on the counts of reactants, products, enzymes, inhibitors and other modifiers, each totalled separately. In the current paper, the definition of capability bisimilarity is the more natural and standard definition of matching on (α, w). Both of the strong capability bisimilarities can be expressed in the more general equivalence that defined in Section 5. The strong stochastic bisimilarity in the original paper is identical to the one here.

5. A more general definition of bisimilarity

In the previous three sections, equivalences based on those defined for PEPA have been considered. It has been demonstrated that for the stochastic relation, all three equivalences equate the same Bio-PEPA systems, and for the capability relation, both equivalences defined equate the same Bio-PEPA models. Moreover, the equivalences on the capability relation induce those on the stochastic relation. This means, essentially that there is one equivalence, and a strict one at that. It is strict in the sense that exact matching is required and also because the limited branching in Bio-PEPA models means that there is limited scope for variations in behaviour.

In this section, a different definition is introduced based on bisimilarity. This equivalence is defined over the $\rightarrow_{sc}$ relation since it contains all relevant information. A function is introduced into the definition of bisimulation to allow for different interpretations of the transition label (α, w) as well as information in the target model and the source model. This function may also refer to the information given by the covering context since this is fixed.

This function takes three arguments: the transition label and two Bio-PEPA models. In the following, let $\mathcal{P}_1 = \langle \mathcal{T}, P_1 \rangle$ and $\mathcal{P}_2 = \langle \mathcal{T}, P_2 \rangle$ be Bio-PEPA systems with $\mathcal{T}$ a covering context.

Definition 25. Let $g : \Theta \times \mathbf{P} \times \mathbf{P} \rightarrow X$ be a function with arbitrary non-empty range X. A binary relation $\mathcal{R}$ over $\mathcal{T}(\mathcal{P})$ is a *g-bisimulation* if $(\langle \mathcal{T}, P_1 \rangle, \langle \mathcal{T}, P_2 \rangle) \in \mathcal{R}$ implies that for all $\theta_1 \in \Theta$

1. whenever $\langle \mathcal{T}, P_1 \rangle \xrightarrow{\theta_1}_{sc} \langle \mathcal{T}, P'_1 \rangle$ then for some $\langle \mathcal{T}, P'_2 \rangle$ and θ_2, $\langle \mathcal{T}, P_2 \rangle \xrightarrow{\theta_2}_{sc} \langle \mathcal{T}, P'_2 \rangle$, $g(\theta_1, P_1, P'_1) = g(\theta_2, P_2, P'_2)$ and $(\langle \mathcal{T}, P'_1 \rangle, \langle \mathcal{T}, P'_2 \rangle) \in \mathcal{R}$
2. whenever $\langle \mathcal{T}, P_2 \rangle \xrightarrow{\theta_2}_{sc} \langle \mathcal{T}, P'_2 \rangle$ then for some $\langle \mathcal{T}, P'_1 \rangle$ and θ_1, $\langle \mathcal{T}, P_1 \rangle \xrightarrow{\theta_1}_{sc} \langle \mathcal{T}, P'_1 \rangle$, $g(\theta_1, P_1, P'_1) = g(\theta_2, P_2, P'_2)$ and $(\langle \mathcal{T}, P'_1 \rangle, \langle \mathcal{T}, P'_2 \rangle) \in \mathcal{R}$

Definition 26. $\mathcal{P}_1$ and $\mathcal{P}_2$ are *g-bisimilar*, $\mathcal{P}_1 \sim_g \mathcal{P}_2$, if $(\mathcal{P}_1, \mathcal{P}_2) \in \mathcal{R}$ for some g-bisimulation $\mathcal{R}$. $\sim_g = \bigcup\{\mathcal{R} \mid \mathcal{R}$ a g-bisimulation$\}$.

A number of functions can be defined to give us equivalences of interest. Let $\theta = (\alpha, w)$ be the transition label, P the source model and P' the target model.

It is straightforward to express the bisimilarities defined earlier in this paper.

System bisimilarity (Definition 17): $g_s(\theta, P, P') = (action(\theta), r_{action(\theta)}[list(\theta), \mathcal{N}, \mathcal{K}])$

Capability bisimilarity (Definition 23): $g_c(\theta, P, P') = \theta = (\alpha, w)$

The bisimilarities defined in the original Bio-PEPA paper [7] can be expressed in the following manner.

Strong stochastic bisimilarity:

$$g_{ssb}(\theta, P, P') = (action(\theta), r_{action(\theta)}[list(\theta), \mathcal{N}, \mathcal{K}]) = g_s(\theta, P, P')$$

Strong capability bisimilarity:

$$g_{scb}(\theta, P, P') = (action(\theta), \#reacts(\theta), \#prods(\theta), \#enzs(\theta), \#inhibs(\theta), \#mods(\theta))$$

Strong capability bisimilarity (2):

$$g_{scb2}(\theta, P, P') = action(\theta)$$

An alternative name for the last equivalence is qualitative bisimilarity since all quantitative aspects of the transitions are ignored. By Theorems 1 and 2, the isomorphisms from the original Bio-PEPA paper can also be expressed as g-bisimilarities.

System isomorphism: $g_{si}(\theta, P, P') = g_s(\theta, P, P')$

Component isomorphism: $g_{ci}(\theta, P, P') = g_c(\theta, P, P')$

A question of interest is whether system equivalence (as given in Definition 18) can be expressed as g-bisimilarity by finding suitable g. System equivalence is defined in the style of Larsen and Skou [25] and requires summation of rates. Assuming $\mathcal{R} \subseteq \tilde{\mathbf{P}} \times \tilde{\mathbf{P}}$ is an equivalence relation then the following definition can be made.

System equivalence: $g_{se}(\theta, P, P') = \Sigma \{ r \mid \langle \mathcal{T}, P \rangle \xrightarrow{(\text{action}(\theta), r)} \langle \mathcal{T}, Q \rangle, Q \in [P']_{\mathcal{R}} \}$

where $[P']_{\mathcal{R}}$ is the equivalence class of P' with respect to $\mathcal{R}$. In the specific case of well-defined Bio-PEPA systems, by Proposition 4, there is only one transition from P for a specific α and this definition is unnecessary. However, if an equivalence definition requires the summation over the rates on many transitions with targets in the same equivalence class, this definition illustrates that the Larsen and Skou style is possible.

Finally, it is possible to use additional functions to define g. For example, it may be desirable to link together the names of species. Given a function G which maps from component names to $\mathbb{N}$, let any components which are to be regarded as the same, be mapped to the same number. Alternatively G could map to a set of component names that do not appear in $\mathcal{T}$. The following g-bisimilarity can be defined.

Reactant-product bisimilarity: $g_{rp}(\theta, P, P') = (action(\theta), G(\text{reacts}(\theta)), G(\text{prods}(\theta)))$

It may also be of interest to relate action names. These are simply tags for identifying reactions and have no inherent meaning. Using the same approach to defining G as above, the following quantitative equivalence illustrates how this can be done.

Action-relabelling bisimilarity $g_{al}(\theta, P, P') = G(action(\theta))$

6. Congruence of *g*-bisimilarity

As always, congruence with respect to the operators of a process algebra is important to establish, as this allows various manipulations. For this type of parameterised bisimilarity, it is important to consider which functions allow the bisimilarity to be a congruence. The following definitions are useful. In the next definition, Proposition 1 is used to treat w as a set.

Definition 27. A function $g : \Theta \times \mathbf{P} \times \mathbf{P} \to X$ is *set-stable* if for any valid w,

$$g((\alpha_1, w_1), P_1, P_1') = g((\alpha_2, w_2), P_2, P_2') \quad \Rightarrow \quad g((\alpha_1, w_1 \cup w), P_1, P_1') = g((\alpha_2, w_2 \cup w), P_2, P_2').$$

Definition 28. A function $g : \Theta \times \mathbf{P} \times \mathbf{P} \to X$ is *species-blind* if g ignores the species/ component names that appear in w.

Three results will be proved about congruence. Species-blindness is not necessary as long as there is a suitable mapping from component names for S_1 to those for S_2, so that species name is not used as a way to distinguish transition labels.

Theorem 3. *Let S_1, S_2 and S be Bio-PEPA sequential components and g be species-blind. If $\langle \mathcal{T}, S_1(l_1)\rangle \sim_g \langle \mathcal{T}, S_2(l_2)\rangle$ then*

1. $\langle \mathcal{T}, (S_1 + S)(l_1)\rangle \sim_g \langle \mathcal{T}, (S_2 + S)(l_2)\rangle$ *and* $\langle \mathcal{T}, (S + S_1)(l_1)\rangle \sim_g \langle \mathcal{T}, (S + S_2)(l_2)\rangle$
2. $\langle \mathcal{T}, C_1(l_1)\rangle \sim_g \langle \mathcal{T}, C_2(l_2)\rangle$ *where* $C_i \stackrel{\text{def}}{=} S_i$ *for* $i = 1, 2$

Proof.

Choice: The proof proceeds by induction on the height of the transition derivation. Since the rule `Enrich` can only be applied as the last step of the derivation tree, the second last step is considered. Only one case will be considered here; the other case is similar. It must be shown that for any transition $\langle \mathcal{T}, (S_1 + S)(l_1)\rangle \xrightarrow{\theta_1}_{sc} \langle \mathcal{T}, S_1'(l_1')\rangle$, there exists a transition $\langle \mathcal{T}, (S_2 + S)(l_2)\rangle \xrightarrow{\theta_2}_{sc} \langle \mathcal{T}, S_2'(l_2')\rangle$ with $g(\theta_1, (S_1 + S)(l_1), S_1'(l_1')) = g(\theta_2, (S_2 + S)(l_2), S_2'(l_2'))$ and $\langle \mathcal{T}, S_1'(l_1')\rangle \sim_g \langle \mathcal{T}, S_2'(l_2')\rangle$.

Consider the first transition. By a one-step shorter inference, $(S_1 + S)(l_1) \xrightarrow{\theta_1}_c S_1'(l_1')$, and by a two-step shorter inference $S_1(l_1) \xrightarrow{\theta_1}_c S_1'(l_1')$. By applying `Enrich` to this transition, the transition $\langle \mathcal{T}, S_1(l_1)\rangle \xrightarrow{\theta_1}_{sc} \langle \mathcal{T}, S_1'(l_1')\rangle$ is obtained.

Since $\langle \mathcal{T}, S_1(l_1)\rangle \sim_g \langle \mathcal{T}, S_2(l_2)\rangle$ there exists the transition $\langle \mathcal{T}, S_2(l_2)\rangle \xrightarrow{\theta_2}_{sc} \langle \mathcal{T}, S_2'(l_2')\rangle$ with $g(\theta_1, S_1(l_1), S_1'(l_1')) = g(\theta_2, S_2(l_2), S_1'(l_2'))$ and $\langle \mathcal{T}, S_1'(l_1')\rangle \sim_g \langle \mathcal{T}, S_2'(l_2')\rangle$. Therefore by a shorter inference $S_2(l_2) \xrightarrow{\theta_2}_{sc} S_2'(l_2')$ and by an application of `Choice` followed by an application of `Enrich`, $\langle \mathcal{T}, (S_2 + S)(l_2)\rangle \xrightarrow{\theta_2}_{sc} \langle \mathcal{T}, S_2'(l_2')\rangle$. Since g is species-blind, $g(\theta_1, (S_1 + S)(l_1), S_1'(l_1')) = g(\theta_2, (S_2 + S)(l_2), S_2'(l_2'))$. Hence all conditions for g-bisimilarity are fulfilled.

Constant: Similar to previous case. □

Next, the prefixes are considered. There are various ways in which this theorem can be stated but here the most minimal statement has been chosen. The corollary illustrates a more general condition.

Theorem 4. *Let S_1, S_2 and S be Bio-PEPA sequential agents and g a species-blind function. Then*

1. $\langle \mathcal{T}, ((\alpha, \kappa)\ \mathtt{op}\ S_1)(l)\rangle \sim_g \langle \mathcal{T}, ((\alpha, \kappa)\ \mathtt{op}\ S_2)(l)\rangle$ *for* $\mathtt{op} \in \{\oplus, \ominus, \odot\}$ *if* $\langle \mathcal{T}, S_1(l)\rangle \sim_g \langle \mathcal{T}, S_2(l)\rangle$
2. $\langle \mathcal{T}, ((\alpha, \kappa){\uparrow}S_1)(l)\rangle \sim_g \langle \mathcal{T}, ((\alpha, \kappa){\uparrow}S_2)(l)\rangle$ *if* $\langle \mathcal{T}, S_1(l + \kappa)\rangle \sim_g \langle \mathcal{T}, S_2(l + \kappa)\rangle$

3. $\langle \mathcal{T}, ((\alpha,\kappa)\downarrow S_1)(l)\rangle \sim_g \langle \mathcal{T}, ((\alpha,\kappa)\downarrow S_2)(l)\rangle$ *if* $\langle \mathcal{T}, S_1(l-\kappa)\rangle \sim_g \langle \mathcal{T}, S_2(l-\kappa)\rangle$

Proof.

Modifier prefix: There are three cases to consider here but they can be covered by the rule

$$((\alpha,k)\ \mathsf{op}\ S)(l) \xrightarrow{(\alpha,[S:\ \mathsf{op}\ (l,k)])}_c S(l).$$

Using this rule, assuming l is large enough,

$$((\alpha,\kappa)\ \mathsf{op}\ S_1)(l) \xrightarrow{(\alpha,[S_1:\ \mathsf{op}\ (l,\kappa)])}_c S_1(l) \quad \text{and} \quad ((\alpha,\kappa)\ \mathsf{op}\ S_2)(l) \xrightarrow{(\alpha,[S_2:\ \mathsf{op}\ (l,\kappa)])}_c S_2(l)$$

and using `Enrich`

$$\langle \mathcal{T}, ((\alpha,\kappa)\ \mathsf{op}\ S_1)(l)\rangle \xrightarrow{(\alpha,[S_1:\ \mathsf{op}\ (l,\kappa)])}_{sc} \langle \mathcal{T}, S_1(l)\rangle \quad \text{and}$$

$$\langle \mathcal{T}, ((\alpha,\kappa)\ \mathsf{op}\ S_2)(l)\rangle \xrightarrow{(\alpha,[S_2:\ \mathsf{op}\ (l,\kappa)])}_{sc} \langle \mathcal{T}, S_2(l)\rangle$$

and the result follows since g is species-blind.

Product prefix: Consider the rule $((\alpha,k)\uparrow S)(l) \xrightarrow{(\alpha,[S:\uparrow(l,k)])}_c S(l+k)$ with $0 \geq \kappa \geq N-\kappa$. Using this rule,

$$((\alpha,\kappa)\uparrow S_1)(l) \xrightarrow{(\alpha,[S_1:\uparrow(l,\kappa)])}_c S_1(l+k) \quad \text{and} \quad ((\alpha,\kappa)\uparrow S_2)(l) \xrightarrow{(\alpha,[S_2:\uparrow(l,\kappa)])}_c S_2(l+k)$$

and using `Enrich`

$$\langle \mathcal{T}, ((\alpha,\kappa)\uparrow S_1)(l)\rangle \xrightarrow{(\alpha,[S_1:\uparrow(l,\kappa)])}_{sc} \langle \mathcal{T}, S_1(l+k)\rangle \quad \text{and}$$

$$\langle \mathcal{T}, ((\alpha,\kappa)\uparrow S_2)(l)\rangle \xrightarrow{(\alpha,[S_2:\uparrow(l,\kappa)])}_{sc} \langle \mathcal{T}, S_2(l+k)\rangle.$$

Since $\langle \mathcal{T}, S_1(l+k)\rangle \sim_g \langle \mathcal{T}, S_2(l+k)\rangle$ the conclusion is that $\langle \mathcal{T}, ((\alpha,\kappa)\uparrow S_1)(l)\rangle \sim_g \langle \mathcal{T}, ((\alpha,\kappa)\uparrow S_2)(l)\rangle$ as long as g does not refer to species names S_1 and S_2.

Reactant prefix: This case can be proved similarly to the previous one. □

Corollary 2. *Given two Bio-PEPA sequential components S_1 and S_2 with the same maximum level N such that $\langle \mathcal{T}, S_1(l)\rangle \sim_g \langle \mathcal{T}, S_2(l)\rangle$ for all $0 \leq l \leq N$ and g a species-blind function then*

1. $\langle \mathcal{T}, ((\alpha,\kappa)\downarrow S_1)(l)\rangle \sim_g \langle \mathcal{T}, ((\alpha,\kappa)\downarrow S_2)(l)\rangle$

2. $\langle \mathcal{T}, ((\alpha,\kappa){\uparrow}S_1)(l)\rangle \sim_g \langle \mathcal{T}, ((\alpha,\kappa){\uparrow}S_2)(l)\rangle$

Finally, the cooperation case can be considered. This requires a condition on the function g, namely that adding the same information about reactions preserves equality under g.

Theorem 5. *Let P_1, P_2 and Q be Bio-PEPA model components. If $\langle \mathcal{T}, P_1\rangle \sim_g \langle \mathcal{T}, P_2\rangle$ with g both set-stable and species-blind, then $\langle \mathcal{T}, P_1 \underset{\mathcal{L}}{\bowtie} Q\rangle \sim_g \langle \mathcal{T}, P_2 \underset{\mathcal{L}}{\bowtie} Q\rangle$ and $\langle \mathcal{T}, Q \underset{\mathcal{L}}{\bowtie} P_1\rangle \sim_g \langle \mathcal{T}, Q \underset{\mathcal{L}}{\bowtie} P_2\rangle$*

Proof. There are three rules to consider and a suitable bisimulation needs to be constructed. Let $\mathcal{R} = \{(\langle \mathcal{T}, P_1 \underset{\mathcal{L}}{\bowtie} Q\rangle, \langle \mathcal{T}, P_2 \underset{\mathcal{L}}{\bowtie} Q\rangle) \mid \langle \mathcal{T}, P_1\rangle \sim_g \langle \mathcal{T}, P_2\rangle\}$. There are three cases but only the case of synchronisation, namely $\alpha \in L$, will be proved. The other cases are similar but simpler.

Consider a transition $\langle \mathcal{T}, P_1 \underset{\mathcal{L}}{\bowtie} Q\rangle \xrightarrow{(\alpha_1,w_1)}_{sc} \langle \mathcal{T}, P_1' \underset{\mathcal{L}}{\bowtie} Q'\rangle$. If a second transition $\langle \mathcal{T}, P_2 \underset{\mathcal{L}}{\bowtie} Q\rangle \xrightarrow{(\alpha_2,w_2)}_{sc} \langle \mathcal{T}, P_2' \underset{\mathcal{L}}{\bowtie} Q'\rangle$ such that $g((\alpha_1,w_1),P_1,P_1') = g((\alpha_2,w_2),P_2,P_2')$ and $\langle \mathcal{T}, P_1'\rangle \sim_g \langle \mathcal{T}, P_2'\rangle$ can be found, then $\mathcal{R}$ is a g-bisimulation.

Considering the first transition above, by a two-step shorter inference $P_1 \xrightarrow{(\alpha_1,w_1')}_c P_1'$ and $Q \xrightarrow{(\alpha_1,w'')}_c Q'$ with $w_1 = w_1' \cup w''$.

By applying `Enrich` to the transition from P_1, using the fact that $\langle \mathcal{T}, P_1\rangle \sim_g \langle \mathcal{T}, P_2\rangle$ to obtain the matching transition, $g((\alpha_1,w_1'),P_1,P_1') = g((\alpha_2,w_2'),P_2,P_2')$ and $\langle \mathcal{T}, P_1'\rangle \sim_g \langle \mathcal{T}, P_2'\rangle$ and considering a one-step shorter inference on the derivation of the matching transition, hence $P_2 \xrightarrow{(\alpha_2,w_2')}_c P_2'$.

Combining this and the transition for Q, and applying the rule `Enrich`, gives the transition $\langle \mathcal{T}, P_2 \underset{\mathcal{L}}{\bowtie} Q\rangle \xrightarrow{(\alpha_2,w_2'\cup w'')}_{sc} \langle \mathcal{T}, P_2' \underset{\mathcal{L}}{\bowtie} Q'\rangle$. Since g is set-stable and species-blind $g((\alpha_1,w_1'\cup w''),P_1 \underset{\mathcal{L}}{\bowtie} Q, P_1' \underset{\mathcal{L}}{\bowtie} Q) = g((\alpha_2,w_2'\cup w''),P_2 \underset{\mathcal{L}}{\bowtie} Q, P_2' \underset{\mathcal{L}}{\bowtie} Q)$ and the result follows. □

Note that the set-stability condition imposed on g is reasonable. One way in which to break it is for g to count the number of elements in the set, and to return 0 if this number is less than or equal to a value x, and 1 if this number is greater than x. Then if $|w_1'| = x - 1$ and $|w_2'| = x$, set-stability would not hold. It seems unlikely that this type of function would return an equivalence of interest.

7. Weak g-bisimilarity

In this section, a weak form of g-bisimulation is considered where it is possible to abstract from certain actions. This type of equivalence has not been considered for Bio-PEPA although weak isomorphism has been defined for PEPA [8].

By allowing the function g to map to a special value τ, it is possible to abstract away from such reactions that are mapped to τ in the style of weak CCS as defined by Milner [19]. To start, new transition notation is required.

Definition 29. Let $g : \Theta \times \mathbf{P} \times \mathbf{P} \to X \cup \{\tau\}$ be a function where X is a non-empty, arbitrary set. Then

1. $P \xrightarrow{g(\theta_1,P,P_1)\,\ldots\,g(\theta_n,P_{n-1},P')}_g P'$ represents $P \xrightarrow{\theta_1}_c P_1 \xrightarrow{\theta_2}_c \ldots \xrightarrow{\theta_{n-1}}_c P_{n-1} \xrightarrow{\theta_n}_c P'$.
2. $P \overset{\phi_1\ldots\phi_n}{\Longrightarrow}_g P'$ represents $P\ (\xrightarrow{\tau}_g)^* \xrightarrow{\phi_1}_g (\xrightarrow{\tau}_g)^* \ \ldots\ (\xrightarrow{\tau}_g)^* \xrightarrow{\phi_n}_g (\xrightarrow{\tau}_g)^*\ P'$ for $\phi_i \in X \cup \{\tau\}$, $i = 1, \ldots, n$.

Note that $P \Rightarrow_g P'$ allows for none, one or more τ transitions, hence $P \Rightarrow_g P$.

This notation can be extended to Bio-PEPA systems in the obvious manner. It is then possible to define the following equivalence in the style of weak bisimulation equivalence.

Definition 30. Let $g : \Theta \times \mathbf{P} \times \mathbf{P} \to X \cup \{\tau\}$ be a function where X is a non-empty, arbitrary set. A binary relation $\mathcal{R}$ over $\mathcal{T}(\mathcal{P})$ is a *weak g-bisimulation* if $(\langle \mathcal{T}, P_1\rangle, \langle \mathcal{T}, P_2\rangle) \in \mathcal{R}$ implies that for all $\phi \neq \tau$,

1. whenever $\langle \mathcal{T}, P_1\rangle \xrightarrow{\phi}_g \langle \mathcal{T}, P_1'\rangle$ then for some $\langle \mathcal{T}, P_2'\rangle$, $\langle \mathcal{T}, P_2\rangle \overset{\phi}{\Rightarrow}_g \langle \mathcal{T}, P_2'\rangle$, and $(\langle \mathcal{T}, P_1'\rangle, \langle \mathcal{T}, P_2'\rangle) \in \mathcal{R}$
2. whenever $\langle \mathcal{T}, P_2\rangle \xrightarrow{\phi}_g \langle \mathcal{T}, P_2'\rangle$ then for some $\langle \mathcal{T}, P_1'\rangle$, $\langle \mathcal{T}, P_1\rangle \overset{\phi}{\Rightarrow}_g \langle \mathcal{T}, P_1'\rangle$, and $(\langle \mathcal{T}, P_1'\rangle, \langle \mathcal{T}, P_2'\rangle) \in \mathcal{R}$
3. whenever $\langle \mathcal{T}, P_1\rangle \xrightarrow{\tau}_g \langle \mathcal{T}, P_1'\rangle$ then for some $\langle \mathcal{T}, P_2'\rangle$, $\langle \mathcal{T}, P_2\rangle \Rightarrow_g \langle \mathcal{T}, P_2'\rangle$, and $(\langle \mathcal{T}, P_1'\rangle, \langle \mathcal{T}, P_2'\rangle) \in \mathcal{R}$
4. whenever $\langle \mathcal{T}, P_2\rangle \xrightarrow{\tau}_g \langle \mathcal{T}, P_2'\rangle$ then for some $\langle \mathcal{T}, P_1'\rangle$, $\langle \mathcal{T}, P_1\rangle \Rightarrow_g \langle \mathcal{T}, P_1'\rangle$, and $(\langle \mathcal{T}, P_1'\rangle, \langle \mathcal{T}, P_2'\rangle) \in \mathcal{R}$

Definition 31. $\mathcal{P}_1$ and $\mathcal{P}_2$ are *weakly g-bisimilar*, $\mathcal{P}_1 \approx_g \mathcal{P}_2$, if $(\mathcal{P}_1, \mathcal{P}_2) \in \mathcal{R}$ for some weak g-bisimulation $\mathcal{R}$. $\approx_g = \bigcup\{\mathcal{R} \mid \mathcal{R}$ is a weak g-bisimulation$\}$.

As before, congruence is of interest for this equivalence.

Theorem 6. *Let S_1, S_2 and S be Bio-PEPA sequential components and let g be species-blind. If $\langle \mathcal{T}, S_1(l_1)\rangle \approx_g \langle \mathcal{T}, S_2(l_2)\rangle$ then $\langle \mathcal{T}, C_1(l_1)\rangle \approx_g \langle \mathcal{T}, C_2(l_2)\rangle$ where $C_i \stackrel{\text{def}}{=} S_i$ for $i = 1, 2$.*

Proof. For $\langle \mathcal{T}, C_1(l_1)\rangle \xrightarrow{\phi}_g \langle \mathcal{T}, S_1'(l_1')\rangle$, the transition $\langle \mathcal{T}, S_2(l_2)\rangle \stackrel{\phi}{\Longrightarrow}_g \langle \mathcal{T}, S_2'(l_2')\rangle$ can be obtained in the standard manner. This transition is a sequence of transitions of the form $\langle \mathcal{T}, S_2(l_2)\rangle \xrightarrow{\tau}_g \langle \mathcal{T}, S''(l'')\rangle\, (\xrightarrow{\tau}_g)^* \xrightarrow{\phi}_g (\xrightarrow{\tau}_g)^*\, \langle \mathcal{T}, S_2'(l_2')\rangle$. A shorter inference for first transition followed by application of rules `constant` and then `Enrich` gives $\langle \mathcal{T}, C_2(l_2)\rangle \xrightarrow{\tau}_g \langle \mathcal{T}, S''(l'')\rangle$ since g is species-blind, hence $\langle \mathcal{T}, C_2(l_2)\rangle \stackrel{\phi}{\Longrightarrow}_g \langle \mathcal{T}, S_2'(l_2')\rangle$ as required. □

Theorem 7. *Let S_1, S_2 and S be Bio-PEPA sequential components and g a species-blind function. Then*

1. $\langle \mathcal{T}, ((\alpha, \kappa)\ \mathsf{op}\ S_1)(l)\rangle \approx_g \langle \mathcal{T}, ((\alpha, \kappa)\ \mathsf{op}\ S_2)(l)\rangle$ *for* $\mathsf{op} \in \{\oplus, \ominus, \odot\}$ *if* $\langle \mathcal{T}, S_1(l)\rangle \approx_g \langle \mathcal{T}, S_2(l)\rangle$
2. $\langle \mathcal{T}, ((\alpha, \kappa){\uparrow} S_1)(l)\rangle \approx_g \langle \mathcal{T}, ((\alpha, \kappa){\uparrow} S_2)(l)\rangle$ *if* $\langle \mathcal{T}, S_1(l + \kappa)\rangle \approx_g \langle \mathcal{T}, S_2(l + \kappa)\rangle$
3. $\langle \mathcal{T}, ((\alpha, \kappa){\downarrow} S_1)(l)\rangle \approx_g \langle \mathcal{T}, ((\alpha, \kappa){\downarrow} S_2)(l)\rangle$ *if* $\langle \mathcal{T}, S_1(l - \kappa)\rangle \approx_g \langle \mathcal{T}, S_2(l - \kappa)\rangle$

Proof. In each case, there are both $\langle \mathcal{T}, ((\alpha, \kappa)\ \mathsf{op}\ S_1)(l)\rangle \xrightarrow{(\alpha, [S_1 :\ \mathsf{op}\ (l, \kappa)])}_{sc} \langle \mathcal{T}, S_1(l')\rangle$ and $\langle \mathcal{T}, ((\alpha, \kappa)\ \mathsf{op}\ S_2)(l)\rangle \xrightarrow{(\alpha, [S_2 :\ \mathsf{op}\ (l, \kappa)])}_{sc} \langle \mathcal{T}, S_2(l')\rangle$ where l' has the appropriate arithmetic relationship with l. The equality $g((\alpha, [S_1 :\ \mathsf{op}\ (l, \kappa)]), ((\alpha, \kappa)\ \mathsf{op}\ S_1)(l), S_1(l'))$ $= g((\alpha, [S_2 :\ \mathsf{op}\ (l, \kappa)]), ((\alpha, \kappa)\ \mathsf{op}\ S_2)(l), S_2(l'))$ holds since g is species-blind and the results follow. □

Corollary 3. *Given two Bio-PEPA sequential components S_1 and S_2 with the same maximum level N such that $\langle \mathcal{T}, S_1(l)\rangle \approx_g \langle \mathcal{T}, S_2(l)\rangle$ for all $0 \leq l \leq N$ and g a species-blind function then*

1. $\langle \mathcal{T}, ((\alpha, \kappa){\downarrow} S_1)(l)\rangle \approx_g \langle \mathcal{T}, ((\alpha, \kappa){\downarrow} S_2)(l)\rangle$

2. $\langle \mathcal{T}, ((\alpha,\kappa){\uparrow}S_1)(l)\rangle \approx_g \langle \mathcal{T}, ((\alpha,\kappa){\uparrow}S_2)(l)\rangle$

Next, the cooperation case can be considered.

Theorem 8. *Let P_1, P_2 and Q be Bio-PEPA model components. If $\langle \mathcal{T}, P_1\rangle \approx_g \langle \mathcal{T}, P_2\rangle$ with g set-stable and species-blind, then both $\langle \mathcal{T}, P_1 \underset{\mathcal{L}}{\bowtie} Q\rangle \approx_g \langle \mathcal{T}, P_2 \underset{\mathcal{L}}{\bowtie} Q\rangle$ and $\langle \mathcal{T}, Q \underset{\mathcal{L}}{\bowtie} P_1\rangle \approx_g \langle \mathcal{T}, Q \underset{\mathcal{L}}{\bowtie} P_2\rangle$.*

Proof. Let $\mathcal{R} = \{(\langle \mathcal{T}, P_1 \underset{\mathcal{L}}{\bowtie} Q\rangle, \langle \mathcal{T}, P_2 \underset{\mathcal{L}}{\bowtie} Q\rangle) \mid \langle \mathcal{T}, P_1\rangle \approx_g \langle \mathcal{T}, P_2\rangle\}$. Considering synchronisation, namely $\alpha \in \mathcal{L}$, let $\langle \mathcal{T}, P_1 \underset{\mathcal{L}}{\bowtie} Q\rangle \xrightarrow{\phi}_g \langle \mathcal{T}, P_1' \underset{\mathcal{L}}{\bowtie} Q'\rangle$. By a two-step shorter inference, $P_1 \xrightarrow{(\alpha_1, w_1')}_c P_1'$ and $Q \xrightarrow{(\alpha_1, w'')}_c Q'$ with $w_1 = w_1' \cup w''$, and $\phi = g((\alpha_1, w_1), P_1 \underset{\mathcal{L}}{\bowtie} Q, P_1' \underset{\mathcal{L}}{\bowtie} Q')$.

By applying `Enrich` to the transition from P_1 and using $\langle \mathcal{T}, P_1\rangle \approx_g \langle \mathcal{T}, P_2\rangle$, then $\langle \mathcal{T}, P_2\rangle\, (\xrightarrow{\tau}_g)^*\, P \xrightarrow{\phi'}_g P'\, (\xrightarrow{\tau}_g)^*\, \langle \mathcal{T}, P_2'\rangle$ and $\phi' = g((\alpha_1, w_1'), P_1, P_1') = g((\alpha_2, w_2'), P, P')$, with $\langle \mathcal{T}, P_1'\rangle \approx_g \langle \mathcal{T}, P_2'\rangle$.

Considering a shorter inference for each transition, applying `coop3` then `Enrich`, the result is $\langle \mathcal{T}, P_2 \underset{\mathcal{L}}{\bowtie} Q\rangle\, (\xrightarrow{\tau}_g)^*\, P \underset{\mathcal{L}}{\bowtie} Q \xrightarrow{\phi''}_g P' \underset{\mathcal{L}}{\bowtie} Q'\ (\xrightarrow{\tau}_g)^*\ \langle \mathcal{T}, P_2' \underset{\mathcal{L}}{\bowtie} Q'\rangle$ where $\phi'' = g((\alpha_2, w_2' \cup w''), P \underset{\mathcal{L}}{\bowtie} Q, P' \underset{\mathcal{L}}{\bowtie} Q')$. Since g is set-stable and species-blind, $\phi'' = \phi$ as required. Likewise, because g is species-blind, all other labels evaluate to τ. The other two cases are similar. $\square$

In general, weak bisimulations are not congruences for choice between sequential components because a transition of the form $P \xrightarrow{\tau} P'$ can be matched by $Q \Rightarrow_g Q$ Hence $P + R \xrightarrow{\tau} P'$ where P' may not have the same actions as R whereas $Q + R \Rightarrow_g Q + R$ which obviously has the actions of R. However, in the context of Bio-PEPA, well-defined sequential components have a restricted form and it is possible to define an extension operator involving choice for which weak g-bisimulation is a congruence.

Definition 32. Let A and B be two well-defined species

$$A \stackrel{\text{def}}{=} \sum_{i=1}^{n} (\alpha_i, \kappa_i)\, \mathsf{op}_i\, A \quad\text{and}\quad B \stackrel{\text{def}}{=} \sum_{j=1}^{m} (\beta_j, \lambda_j)\, \mathsf{op}_j\, B$$

the extension of A by B is defined by

$$A\{B\} \stackrel{\text{def}}{=} \sum_{i=1}^{n} (\alpha_i, \kappa_i)\, \mathsf{op}_i\, A\{B\} + \sum_{j=1}^{m} (\beta_j, \lambda_j)\, \mathsf{op}_j\, A\{B\}$$

$$\begin{tikzcd} & I_1 \arrow[dr, "S_1"] & \\ A \arrow[ur, "R_1"] \arrow[dr, "R_2"'] & & B \\ & I_2 \arrow[ur, "S_2"'] & \end{tikzcd} \qquad A' \xrightarrow{R} I' \xrightarrow{S} B'$$

Figure 3: Species I_1 and I_2 modelled explicitly, and modelled as a single species I

This permits A to take on additional reaction capabilities, specifically those of B. For $A\{B\}$ to be well-defined, there must be no overlap in the reaction names of A and B, and for both A and B to be well-defined. $A\{B\}$ and $B\{A\}$ are isomorphic since their transition systems are structurally identical with matching actions.

Theorem 9. *Let g be species-blind and let $\langle \mathcal{T}, C_1(l)\rangle \approx_g \langle \mathcal{T}, C_2(l)\rangle$ for sequential Bio-PEPA components C_1 and C_2. If $C_1\{C\}$ and $C_2\{C\}$ are well-defined for C a sequential Bio-PEPA component and all species have maximum level n then $\langle \mathcal{T}, (C_1\{C\})(l)\rangle \approx_g \langle \mathcal{T}, (C_2\{C\})(l)\rangle$ and $\langle \mathcal{T}, (C\{C_1\})(l)\rangle \approx_g \langle \mathcal{T}, (C\{C_2\})(l)\rangle$.*

Proof. Consider a transition $\langle \mathcal{T}, (C_1\{C\})(l)\rangle \xrightarrow{\phi}_g \langle \mathcal{T}, (C_1\{C\}(l'))\rangle$ where it is the case that $\phi = g((\alpha, w), (C_1\{C\})(l), (C_1\{C\})(l'))$. If α appears in C, then there exists the transition $\langle \mathcal{T}, (C_2\{C\})(l)\rangle \xrightarrow{\phi}_g \langle \mathcal{T}, (C_2\{C\})(l')\rangle$ with $\phi = g((\alpha, w), (C_2\{C\})(l), (C_2\{C\})(l'))$ and hence $\langle \mathcal{T}, (C_2\{C\})(l)\rangle \overset{\phi}{\Longrightarrow}_g \langle \mathcal{T}, (C_2\{C\})(l')\rangle$.

If α appears in C_1 and C_2 then $\langle \mathcal{T}, C_1(l)\rangle \xrightarrow{\phi}_g \langle \mathcal{T}, C_1(l')\rangle$ and since $C_1(l) \approx_g C_2(l)$, then $\langle \mathcal{T}, C_2(l)\rangle \overset{\phi}{\Longrightarrow}_g \langle \mathcal{T}, C_2(l'')\rangle$ and $\langle \mathcal{T}, (C_2\{C\})(l)\rangle \overset{\phi}{\Longrightarrow}_g \langle \mathcal{T}, (C_2\{C\})(l'')\rangle$ as required. The other cases are shown similarly. □

Note that these results require g to be species-blind. As with the earlier forms of equivalence, as long as g maps species that appear in transitions that should match, to identical values, these congruence results will still hold.

In Section 9, weak g-bisimulation will be applied to a biological system.

8. Example: grouping of species

A specific aim of investigating equivalences in the context of modelling biological systems is to capture the ways in which biologists might choose to abstract informa-

$$
\begin{aligned}
A &\stackrel{def}{=} (\alpha_1, 1) \downarrow A + (\alpha_2, 1) \downarrow A \\
I_1 &\stackrel{def}{=} (\alpha_1, 1) \uparrow I_1 + (\beta_1, 1) \downarrow I_1 \\
I_2 &\stackrel{def}{=} (\alpha_2, 1) \uparrow I_2 + (\beta_2, 1) \downarrow I_2 \\
B &\stackrel{def}{=} (\beta_1, 1) \uparrow B + (\beta_2, 1) \uparrow B
\end{aligned}
\qquad
\begin{aligned}
A' &\stackrel{def}{=} (\alpha, 1) \downarrow A' \\
I' &\stackrel{def}{=} (\alpha, 1) \uparrow I' + (\beta, 1) \downarrow I' \\
B' &\stackrel{def}{=} (\beta, 1) \uparrow B'
\end{aligned}
$$

$$
M \stackrel{def}{=} A(n) \underset{\{\alpha_1,\alpha_2\}}{\bowtie} ((I_1(0) \underset{\emptyset}{\bowtie} I_2(0)) \underset{\{\beta_1,\beta_2\}}{\bowtie} B(0)) \qquad M' \stackrel{def}{=} A'(n) \underset{\{\alpha\}}{\bowtie} (I'(0) \underset{\{\beta\}}{\bowtie} B'(0))
$$

Table 1: Bio-PEPA model definitions

tion. Equivalences developed will capture these abstractions formally and the example illustrates this process. It considers the case where a number of species that play a similar role in a set of reactions are modelled as a single species as illustrated in Figure 3. The species I_1 and I_2 are known to have similar functions and for modelling simplicity, they can be grouped together simply as I. The labels on the reaction arrows give the rates for the reactions. Assuming mass action as the reaction kinetics for all reactions, then it is expected that $R = R_1 + R_2$ and $S = (S_1[I_1] + S_2[I_2])/[I]$ where $[C]$ denotes the concentration of a species C. Biologically, this can be motivated in the way that extracellular regulated kinase (ERK) is considered both in experiments and modelling. ERK actually refers to two different kinases ERK1 and ERK2, and in some experiments they can be identified separately and in others, together [26]. Hence, they are frequently modelled together as a single species [27, 10] but could be modelled separately as current research shows that they have distinct functions within the cell [28].

These two different models can be expressed in Bio-PEPA as given in Table 1. The first model is the running example from Section 2. Although A' and B' are the same species as A and B, they have different definitions and hence it is necessary to give them different names as a way to distinguish them. For the models, the initial levels are the maximum for A and A' and zero for the other components.

A covering context $\mathcal{T}$ is required and is given in Table 2. It is chosen to be minimal (in the sense that only the information needed for the example is included, for example, no constants k_i are used and the rates remain symbolic). All reactions use mass action

$$
\begin{array}{rcl}
\mathcal{T} & = & \mathcal{V}, \mathcal{K}, \mathcal{F}, \mathcal{N}, Comp \qquad \text{where} \qquad \mathcal{V} = \{v\} \qquad \mathcal{K} = \emptyset \\
\mathcal{F} & = & \{f_{\alpha_1} = fMA(r_1),\ f_{\alpha_2} = fMA(r_2),\ f_\alpha = fMA(r), \\
& & \quad f_{\beta_1} = fMA(s_1),\ f_{\beta_2} = fMA(s_2),\ f_\beta = fMA(s)\ \} \\
\mathcal{N} & = & \{A : H{=}h, N{=}n, V{=}v;\ I_1 : H{=}h, N{=}n, V{=}v;\ I_2 : H{=}h, N{=}n, V{=}v; \\
& & \quad B : H{=}h, N{=}n, V{=}v; \\
& & \quad A' : H{=}h, N{=}n, V{=}v;\ I' : H{=}h, N{=}n, V{=}v;\ B' : H{=}h, N{=}n, V{=}v\ \} \\
Comp & = & \{ \text{ model definitions for } A, I_1, I_2, B, A', I', B' \text{ from Table 1 } \}
\end{array}
$$

Table 2: Bio-PEPA context definitions

kinetics, hence the use of *fMA* (which is defined in Section 2). The two Bio-PEPA systems are then $\mathcal{P}' = \langle \mathcal{T}, M' \rangle$ and $\mathcal{P} = \langle \mathcal{T}, M \rangle$.

All species have the same step size h because of the necessity for conservation of mass. The same maximum concentration is assumed for all species.

To define the appropriate bisimulation, note that the reaction α is the analog of reactions α_1 and α_2, likewise for β with β_1 and β_2. Hence, a function is required to capture this.

$$
h_1(\gamma) \quad = \quad \begin{cases} \alpha & \text{if } \gamma = \alpha_i \quad i = 1, 2 \\ \beta & \text{if } \gamma = \beta_i \quad i = 1, 2 \\ \gamma & \text{otherwise} \end{cases}
$$

A function h_2 for rate calculation is required as well.

$$
h_2((\gamma, w), P) \quad = \quad \sum \Big\{ f_{\gamma'}[w', \mathcal{N}, \mathcal{K}]/h \mid P \xrightarrow{(\gamma', w')} \text{ and } h_1(\gamma) = h_1(\gamma') \Big\}
$$

Together these will be used to define g from which a g-bisimulation can be constructed.

$$
g(\theta, P, P') \quad = \quad (h_1(action(\theta)), h_2(\theta, P)).
$$

For the rest of this example, vector notation will be used instead of Bio-PEPA model notation. For model M, states will be represented by $(l_A, l_{I_1}, l_{I_2}, l_B)$ where the order of species is given by the subscripts, and for M', vectors of the form $(l_{A'}, l_{I'}, l_{B'})$ will be used.

To show these two models are g-bisimilar, it is necessary to show that the following relation is a g-bisimulation.

$$\mathcal{R} = \{(k - (j+l), j, l, n-k), (k - (j+l), j+l, n-k) \mid 0 \leq k \leq n, 0 \leq j+l \leq k\}$$

The rates will be considered symbolically in describing transitions after which the relationship necessary between the rates of the different models to obtain g-bisimilarity is considered. There are eight distinct cases and these are considered in Appendix A. For $\mathcal{R}$ to be a g-bisimulation, it is then necessary that $r = r_1 + r_2$ and $s = s_1 = s_2$ (which implies $s = (s_1 j + s_2 l)/(j+l)$).

Hence, this bisimulation captures the original idea of grouping which includes the relationship between the rates, namely $R = R_1 + R_2$ and $S = (S_1[I_1] + S_2[I_2])/[I]$. However, to obtain g-bisimilarity correctly, it is also necessary that $S_1 = S_2 = S$ which is not a feature of the original abstraction, although it does guarantee the correct relationship between rates. This requires further investigation.

The function defined above g is species-blind and set-stable (if all reactions are mass-action-based) and so the congruence results of the previous section apply. Hence, given another Bio-PEPA model P, $P \underset{\mathcal{L}}{\bowtie} M \sim_g P \underset{\mathcal{L}}{\bowtie} M'$. This introduces the possibility that other species might synchronise on the existing reactions. Of more interest, is the case where A and A' are modified to add the capability for the same new reactions, for example, adding $(\gamma, 2) \uparrow A$ and $(\gamma, 2) \uparrow A'$, and B and B' are modified to add capability for the same new reactions (but not necessarily the same new reactions as A and A'). Call the new model components M_1 and M'_1. These are still g-bisimilar, since the same transitions are possible in each case, and congruence can be used again, to give $P \underset{\mathcal{L}}{\bowtie} M_1 \sim_g P \underset{\mathcal{L}}{\bowtie} M'_1$. In this cooperation, it is now possible for shared reactions between P and the second operand that do not involve the grouped species to take place. In both cases where congruence is used, there is the opportunity to replace the more complex system with more states (M or M_1) with the simpler one (M' or M'_1) to obtain a smaller state space to work with.

$$
\begin{array}{lrcl}
v16 & EG + Grb2 & \rightarrow & EG{-}Grb2 \\
v17 & Sos + EG{-}Grb2 & \rightarrow & EG{-}Grb2{-}Sos \\
v18 & Ras{-}GDP + EG{-}Grb2{-}Sos & \rightarrow & EG{-}Grb2{-}Sos{-}Ras{-}GDP \\
v19 & EG{-}Grb2{-}Sos{-}Ras{-}GDP & \rightarrow & Ras{-}GTP + EG{-}Grb2{-}Sos \\
v20 & Ras{-}GTP^{*} + EG{-}Grb2{-}Sos & \rightarrow & EG{-}Grb2{-}Sos{-}Ras{-}GTP \\
v21 & EG{-}Grb2{-}Sos{-}Ras{-}GTP & \rightarrow & Ras{-}GDP + EG{-}Grb2{-}Sos \\
v22 & EG + Shc & \rightarrow & EG{-}Shc \\
v23 & EG{-}Shc & \rightarrow & EG{-}Shc^{*} \\
v24 & Grb2 + EG{-}Shc^{*} & \rightarrow & EG{-}Shc^{*}{-}Grb2 \\
v25 & Sos + EG{-}Shc^{*}{-}Grb2 & \rightarrow & EG{-}Shc^{*}{-}Grb2{-}Sos \\
v26 & Ras{-}GDP + EG{-}Shc^{*}{-}Grb2{-}Sos & \rightarrow & EG{-}Shc^{*}{-}Grb2{-}Sos{-}Ras{-}GDP \\
v27 & EG{-}Shc^{*}{-}Grb2{-}Sos{-}Ras{-}GDP & \rightarrow & Ras{-}GTP + EG{-}Shc^{*}{-}Grb2{-}Sos \\
v30 & Ras{-}GTP^{*} + EG{-}Shc^{*}{-}Grb2{-}Sos & \rightarrow & EG{-}Shc^{*}{-}Grb2{-}Sos{-}Ras{-}GTP \\
v31 & EG{-}Shc^{*}{-}Grb2{-}Sos{-}Ras{-}GTP & \rightarrow & Ras{-}GDP + EG{-}Shc^{*}{-}Grb2{-}Sos \\
v32 & EG{-}Shc^{*}{-}Grb2{-}Sos & \rightarrow & EG + Shc^{*}{-}Grb2{-}Sos \\
v33 & Shc^{*}{-}Grb2{-}Sos & \rightarrow & Shc^{*} + Grb2{-}Sos \\
v34 & EG{-}Grb2{-}Sos & \rightarrow & EG + Grb2{-}Sos \\
v35 & Grb2{-}Sos & \rightarrow & Grb2 + Sos \\
v36 & Shc^{*} & \rightarrow & Shc \\
v37 & EG{-}Shc^{*} & \rightarrow & EG + Shc^{*} \\
v38 & Shc^{*} + Grb2 & \rightarrow & Shc^{*}{-}Grb2 \\
v39 & EG{-}Shc^{*}{-}Grb2 & \rightarrow & EG + Shc^{*}{-}Grb2 \\
v40 & Shc^{*}{-}Grb2 + Sos & \rightarrow & Shc^{*}{-}Grb2{-}Sos \\
v41 & EG{-}Shc^{*} + Grb2{-}Sos & \rightarrow & EG{-}Shc^{*}{-}Grb2{-}Sos
\end{array}
$$

Figure 4: Alternative signalling pathways from the MAPK cascade activated from EGF receptors (*EG* is an abbreviation for $(EGF{-}EGFR^{*})2{-}GAP$ and the reaction names are taken from the original paper [29]).

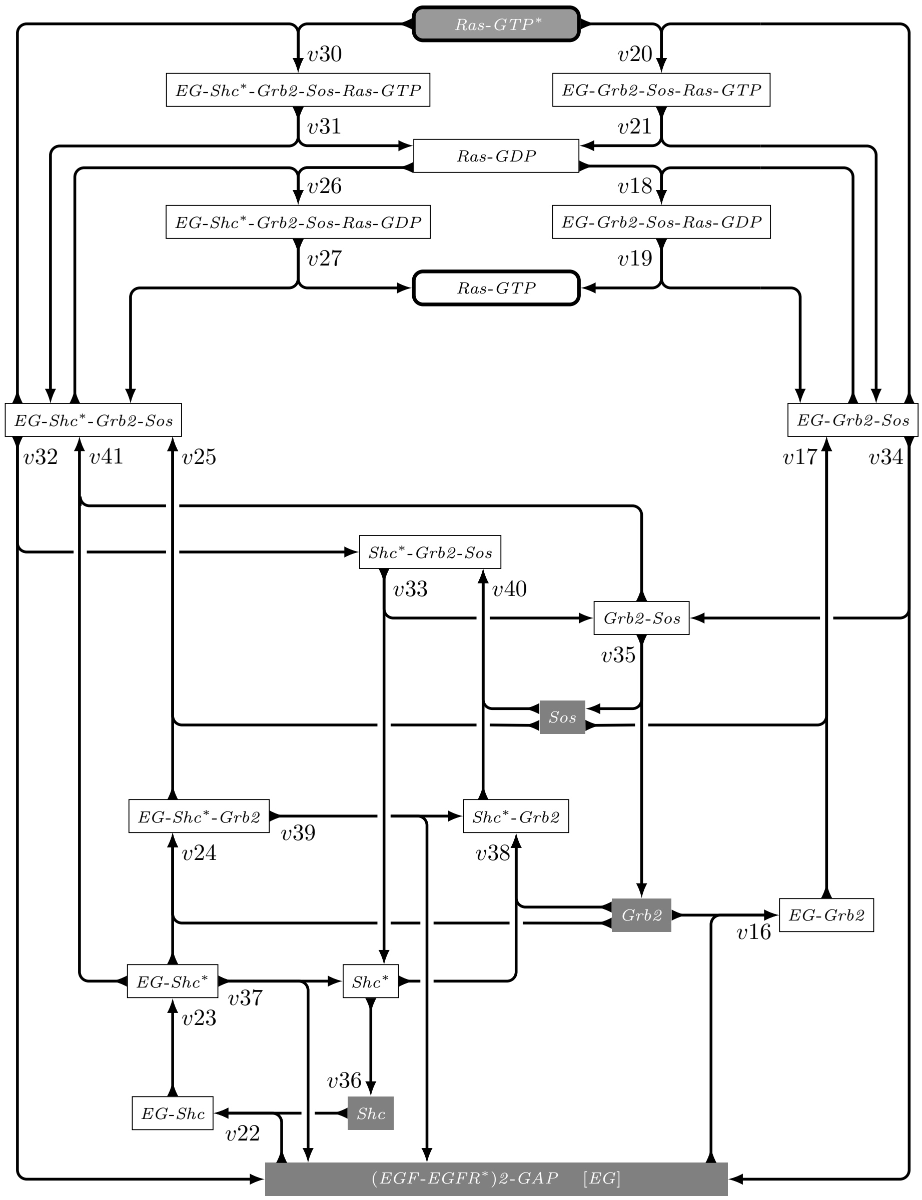

Figure 5: Diagrammatic representation of the reactions in Figure 4 (after [31]). Species in dark boxes with white text are present initially. Rounded rectangles show the sink and the source. The *Shc*-dependent pathway is on the left of the figure and the *Shc*-independent pathway is on the right. Species that are used in both pathways (*EG*, *Grb2*, *Sos* and *Grb2-Sos*) are to the right of the middle

.

9. Example: alternative signalling pathways

Biological research has shown that there can be distinct signalling pathways with the same function within cells [30]. This confers a benefit to an organism since failure of one pathway may reduce function but not totally remove it, mitigating the effect.

In this section, an example of alternative pathways is considered in the MAPK kinase signalling cascade which is triggered by EGF binding to receptors on the cell's external surface [29]. Schoeberl *et al* developed an ODE-based model which was further investigated by Gong and Zhang [31]. For reasons of space, the whole model will not be considered but rather a small fragment that captures the two alternative pathways is considered. Additionally, only reactions originating from the cell-surface will be considered, although internalised reactions are considered in the full model.

In the model, there are two pathways, one independent of the protein *Shc* and one dependent on *Shc*. Both pathways start with the presence of a complex which consists of a GAP protein and a phosphorylated dimer of EGF bound to the receptor EGFR. They both end with the activation of Ras-GTP which leads to the activation of the MAP kinase cascade. By experimentation with the model of Schoeberl *et al*, Gong and Zhang have shown that the *Shc*-dependent pathway is redundant but dominant [31]. The reactions of interest are given in Figure 4 and these are shown diagrammatically in Figure 5. The aim here is to consider the model at a structural level only (ignoring rates) to show that the pathways are redundant, in the sense that either pathway has the same output. This requires the construction of two Bio-PEPA models, one for each pathway, and the construction of a semantic equivalence to show that in terms of features of interest, the pathways are the same. The two Bio-PEPA models are given in the appendix.

The states of the labelled transition systems of the Bio-PEPA models can be viewed as vectors. The *Shc*-dependent model has 17 species and hence a state $\mathbf{t} \in \mathbb{N}^{17}$. The *Shc*-dependent model has 11 species and hence a state $\mathbf{u} \in \mathbb{N}^{11}$. The ordering of these vectors will be discussed later in this section.

A function is required for the equivalence definition. First, an auxiliary function is defined to determine the current change in levels for a species from the string w. Let

$$\Delta(A, w) \quad = \quad \begin{cases} \kappa & \text{if } A \uparrow :(l, \kappa) \text{ appears in } w \\ -\kappa & \text{if } A \downarrow :(l, \kappa) \text{ appears in } w \\ 0 & \text{otherwise} \end{cases}$$

Then, the following generic definition can be made

$$g_{\{A_1,\ldots,A_m\}}((\alpha, w), P, P') \quad = \quad \begin{cases} (\Delta(A_1, w), \ldots, \Delta(A_m, w)) & \text{if any } A_i \text{ appears in } w \\ \tau & \text{otherwise} \end{cases}$$

which provides a list of change values for the species listed only if one of them changes. This can now be used as the basis of the equivalence that will be used to compare the two pathways.

In considering which are the crucial species of the example, it would appear that one could start with the species at the start of the pathway *EG* and that at the end *Ras-GTP*. However, by use of the Bio-PEPA Eclipse Plugin[7], in both models, we can identify a single source *Ras-GTP** and a single sink *Ras-GTP*. These are indicated by the rounded rectangles in Figure 5.

Further investigation using the Infer Invariants option in the Bio-PEPA Plugin [32, 33] shows that the *Shc*-independent model has a single loop involving reactions that decrease and increase *EG* ($v16$ and $v34$, respectively) that returns the system to the same state. Similarly, the *Shc*-dependent model has four loops of this type and one other loop that does not affect *EG*. Both models have four reactions that are not part of loops and these are the reactions that involve the three *Ras* proteins and the complexes they form. Furthermore, the *Shc*-independent model has 4 invariants over species, all of which involve *EG-Grb2-Sos-Ras-GTP* and *EG-Grb2-Sos-Ras-GDP*. Each invariant involves one of the species that is initially present, and hence the invariant has the value of the quantity of that species. The *Shc*-independent model has 5 invariants, again where each involves one of the species that is initially present. All five invariants involve both *EG-Shc***-Grb2-Sos-Ras-GTP* and *EG-Shc***-Grb2-Sos-Ras-GDP*.

[7] `www.biopepa.org`

$\mathbf{t}$	*Shc*-dependent	$\mathbf{u}$	*Shc*-independent
t_1	*Ras-GTP**	u_1	*Ras-GTP**
t_2	*Ras-GDP*	u_2	*Ras-GDP*
t_3	*Ras-GTP*	u_3	*Ras-GTP*
t_4	*EG-Shc*-Grb2-Sos-Ras-GTP*	u_4	*EG-Grb2-Sos-Ras-GTP*
t_5	*EG-Shc*-Grb2-Sos-Ras-GDP*	u_5	*EG-Grb2-Sos-Ras-GDP*
t_6	*EG*	u_6	*EG*
t_7	*Grb2*	u_7	*Grb2*
t_8	*Sos*	u_8	*Sos*
t_9	*Grb2-Sos*	u_9	*Grb2-Sos*
t_{10}	*EG-Shc*-Grb2*	u_{10}	*EG-Grb2*
t_{11}	*EG-Shc*-Grb2-Sos*	u_{11}	*EG-Grb2-Sos*
t_{12}	*Shc*		
t_{13}	*Shc**		
t_{14}	*EG-Shc*		
t_{15}	*EG-Shc**		
t_{16}	*Shc*-Grb2*		
t_{17}	*Shc*-Grb2-Sos*		

Figure 6: Ordering of vectors for *Shc*-dependent and *Shc*-independent models

From this, it is possible to obtain an alternative view of the pathways. Each pathway takes a quantity of *Ras-GTP** and through a sequence of reactions involving a fixed amount of *EG*, *Grb2*, *Sos*, and in the case of the *Shc*-dependent pathway, *Shc*, turns the *Ras-GTP** into the same amount of *Ras-GTP* without losing any *EG*, *Grb2*, *Sos* (and *Shc*). This is the viewpoint that will be taken here in formally identifying the structural similarities in the two models.

A suitable relation must be constructed with pairs of state vectors from each model. The ordering of vectors is given in Figure 6. Hence t_1 is the count for *Ras-GTP** and t_{17} the count for *Shc*-Grb2-Sos* in the state vector $\mathbf{t}$.

The relation over which the semantic equivalence will be checked is defined over $\{0,\ldots,N\}^{17} \times \{0,\ldots,N\}^{11}$ and specified as $\mathcal{R} = \{(\mathbf{t},\mathbf{u}) \mid t_i = u_i \text{ for } i = 1,\ldots,5\}$. This relation matches states with the same level of *Ras-GTP**, *Ras-GDP*, *Ras-GTP*, as well as where *EG-Shc*-Grb2-Sos-Ras-GTP* and *EG-Grb2-Sos-Ras-GTP* are the same, and *EG-Shc*-Grb2-Sos-Ras-GDP* and *EG-Grb2-Sos-Ras-GDP* are the same, for the reasons described above. Let $E = \{\textit{Ras-GTP}^*, \textit{Ras-GDP}, \textit{Ras-GTP}\}$.

To show that $\mathcal{R}$ is a weak g_E-bisimulation, case analysis is required. A maximum level of N is assumed for all species, and at the start of the system *EG*, *Grb2*, *Sos*, *Shc* (in the *Shc*-dependent model) are at this level and all other species are at 0. For convenience, transitions will be labelled with $g_E(\theta)$ rather than θ. There is an invariant involving the first 5 elements of the vector for both models. Hence $t_1+t_2+t_3+t_4+t_5 = N = u_1+u_2+u_3+u_4+u_5$. The first transitions to consider are where the vectors have the form $(k_1, k_2, k_3, k_4, k_5, \ldots)$ for $1 \le k_i \le N-1$, $i = 1, \ldots, 5$

1. $(k_1, k_2, k_3, k_4, k_5, x_6, \ldots, x_{17}) \xrightarrow{(-1,0,0)}_g (k_1\text{-}1, k_2, k_3, k_4\text{+}1, k_5, x'_6, \ldots, x'_{17})$ which occurs when reaction $v30$ is triggered and a similar transition is required for $(k_1, k_2, k_3, k_4, k_5, y_6, \ldots, y_{11})$. The matching reaction for the *Shc*-independent model is $v20$. However, it requires that there is at least one level of *EG-Grb2-Sos* available. By an analysis of the invariants, it can be shown that if no *EG-Grb2-Sos* is available then it is possible through a sequence of τ actions to obtain at least one level of this species. The reactions that are involved are $v35$, $v16$ and $v17$. This gives the transitions of the form

$$\begin{array}{lcl} (k_1, k_2, k_3, k_4, k_5, y_4, \ldots, y_{11}) & \xrightarrow{\tau}_g & (k_1, k_2, k_3, k_4, k_5, y'_6, \ldots, y'_{11}) \\ & \vdots & \vdots \\ & \xrightarrow{\tau}_g & (k_1, k_2, k_3, k_4, k_5, y''_6, \ldots, y''_{11}) \\ & \xrightarrow{(-1,0,0)}_g & (k_1\text{-}1, k_2, k_3, k_4\text{+}1, k_5, y'''_6, \ldots, y'''_{11}) \end{array}$$

with between none and three intermediate transitions, and these transition provide a suitable match.
For $(k_1, k_2, k_3, k_4, k_5, y_6, \ldots, y_{11}) \xrightarrow{(-1,0,0)}_g (k_1\text{-}1, k_2, k_3, k_4\text{+}1, k_5, y'_6, \ldots, y'_{11})$, a similar argument can be made in terms of the invariants of the system that eventually *EG-Shc*-Grb2-Sos* will be produced and reaction $v30$ will occur.

$$\begin{array}{lcl} (k_1, k_2, k_3, k_4, k_5, x_6, \ldots, x_{17}) & \xrightarrow{\tau\ldots\tau}_g & (k_1, k_2, k_3, k_4, k_5, x'_6, \ldots, x'_{17}) \\ & \xrightarrow{(-1,0,0)}_g & (k_1\text{-}1, k_2, k_3, k_4\text{+}1, k_5, x''_6, \ldots, x''_{17}) \end{array}$$

2. $(k_1, k_2, k_3, k_4, k_5, x_6, \ldots, x_{17}) \xrightarrow{(0,1,0)}_g (k_1, k_2\text{+}1, k_3, k_4\text{-}1, k_5, x'_6, \ldots, x'_{17})$ which occurs when reaction $v31$ is triggered and a similar transition is required for the

state $(k_1, k_2, k_3, k_4, k_5, y_6, \ldots, y_{11})$. The matching reaction for the *Shc*-independent model is $v21$. Since there is at least one level of k_4 the reaction is possible and hence a matching transition can be found. The symmetric case is proved similarly.

3. $(k_1, k_2, k_3, k_4, k_5, x_6, \ldots, x_{17}) \xrightarrow{(0,-1,0)}_g (k_1, k_2\text{-}1, k_3, k_4, k_5{+}1, x'_6, \ldots, x'_{17})$. This is proved in a similar fashion to the first case, considering reactions $v26$ and $v18$ and arguing that through a sequence of τ actions, the required reactant species will be formed if it was not already present.
4. $(k_1, k_2, k_3, k_4, k_5, x_6, \ldots, x_{17}) \xrightarrow{(0,0,1)}_g (k_1, k_2, k_3{+}1, k_4, k_5\text{-}1, x'_6, \ldots, x'_{17})$. This is proved in a similar fashion to the second case.
5. $(k_1, k_2, k_3, k_4, k_5, x_6, \ldots, x_{17}) \xrightarrow{\tau}_g (k_1, k_2, k_3, k_4, k_5, x'_6, \ldots, x'_{17})$. This transition can be matched by $(k_1, k_2, k_3, k_4, k_5, y_6, \ldots, y_{11}) \Rightarrow_g (k_1, k_2, k_3, k_4, k_5, y_6, \ldots, y_{11})$.

For the cases where some of the k_i are zero or N, reactions will not be possible but they will not be possible in both models so this does not violate the conditions of the bisimulation.

Hence, $\mathcal{R}$ is a weak g_E-bisimulation. This means that the two models have been compared in terms of the main role of the pathways, and it has been shown that they produce the same outputs from the same inputs. This was achieved by being able to abstract from the reactions that played a similar role in both systems, namely forming complexes that could then be used in the production of *Ras-GTP*. Although this approach was applied to a single system (where it was known that there were redundant pathways), it appears robust enough to be used in other situations to establish the presence of such alternative pathways,

10. Related work

The use of process algebras for modelling systems biology has increased rapidly since the first paper proposed their use [34]. Approaches include the κ-calculus [3], stochastic π-calculus [4, 2], Beta-binders [5] and Bio-Ambients [6]. Most of these approaches use stochastic simulation as their analysis tool, and few approaches have considered the use of semantic equivalences.

In the case where equivalences have been used, these are mostly qualitative, in the sense that they do not consider stochastic aspects of the systems that are being modelled. Note that g-bisimilarity and its weak variant can be either qualitative or quantitative depending on whether rates are calculated by g.

Quantitative equivalences are defined for Bio-PEPA [7] (as discussed in Section 4.4) and an example is given showing strong stochastic bisimilarity between a system with enzymes and the Michaelis-Menten abstraction of the system. In both systems, species have two levels only.

Most of the qualitative equivalences that appear in the literature are based on Milner's (strong) bisimulation or weak bisimulation [19]. Weak bisimulation is shown to be a congruence for the bio-κ-calculus as is a context bisimulation which allows for the modelling of cell interaction [35]. Weak bisimulation has also been used to show that CCS specifications of elements of lactose operon regulation have the same behaviour as more detailed models [36]. In an example of biological modelling using hybrid systems, bisimulation is used to quotient the state space with respect to a subset of variables as a technique for state space reduction [37]. Bisimulation has also been used in the comparison of ambient-style models and membrane-style models [38] and the comparison of a term-rewriting calculus, the Calculus of Looping Systems and a simple brane calculus, PEP [39].

Compression bisimilarity has been defined over Bio-PEPA systems [22]. It is based on the idea that different discretisations of a system should have the same behaviour assuming sufficient levels. Conditions for congruence are identified and the equivalence is illustrated via a substrate-enzyme example.

Another equivalence defined for Bio-PEPA is fast-slow bisimilarity which abstracts from those reactions that are considered to be fast, taking inspiration from the Quasi-Steady-State Assumption [20]. Comparison of fast-slow bisimilarity with weak g-bisimilarity is future work but it seems likely that fast-slow bisimilarity or a variant will be expressible as weak g-bsimilarity with the choice of appropriate function.

11. Conclusion

This paper has presented an investigation of Bio-PEPA with levels; first considering the structure of the transitions obtained for well-defined Bio-PEPA systems and then considering the equivalences defined for PEPA and applying them to Bio-PEPA. The conclusions that can be drawn here are that Bio-PEPA systems have limited branching behaviour and as a result of this, these equivalences are the same. Additionally the equivalences are very fine, and are too strict to be of interest in most biological settings.

A new parameterised bisimilarity, g-bisimilarity is introduced which allows the use of a function over the context, models and transition labels to identify aspects of the same behaviour. It is possible to express strong bisimulation using the new bisimilarity, as well as other equivalences that focus on biological aspects of behaviour. It is shown that g-bisimilarity is a congruence for all operators of Bio-PEPA under certain conditions on the function g and these conditions are not unreasonably strong. Congruence allows the exploitation of the inherent compositionality of the process algebra, since a model can be replaced by one with equivalent behaviour ensuring that the behaviour of the whole system remains unchanged. A weak variant of g-bisimilarity is also presented. This equivalence allows for abstraction from reactions. Congruence is proved for all operators except choice. A new operator that allows the addition of reaction capabilities is defined, and weak g-bisimilarity is shown to be a congruence with respect to this operator.

Two examples are presented. In the first, it is demonstrated how g-bisimilarity can be applied to a real biological modelling example, where two similar species are grouped as one species. It is possible to identify a function that will identify these systems as having the same behaviour using a renaming function and a rate-summing function. The state space can be quotiented into different cases and each case can be proved generally for n elements of the initial species, and none of the other species. Using congruence, it then becomes possible to use the system with the smaller state space (the one where species are grouped) in place of the larger one without changing the behaviour of a larger system containing them. The second example shows how weak g-bisimulation can be used to obtain a technique to compare alternative signalling pathways by taking into account the important species in the pathways.

To summarise, the contribution of this paper is to further our understanding of a stochastic process algebra for biological modelling and as a result of this to develop biologically-meaningful equivalences that can be used to capture when two different systems have the same behaviour.

Acknowledgements: The author thanks Jane Hillston for her comments. Vashti Galpin is supported by EPSRC Grants EP/E031439/1 and EP/C54370X/01.

References

[1] J. Fisher, T. A. Henzinger, Executable cell biology, Nature Biotechnology 25 (2007) 1239–1249.

[2] 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.

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

[4] 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.

[5] 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, 2004, pp. 20–33.

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

[7] F. Ciocchetta, J. Hillston, Bio-PEPA: a framework for the modelling and analysis of biological systems, Theoretical Computer Science, 410 (33-34) (2009) 3065–3084.

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

[9] 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.

[10] M. Calder, S. Gilmore, J. Hillston, Modelling the influence of RKIP on the ERK signalling pathway using the stochastic process algebra PEPA, in: C. Priami, A. Ingólfsdóttir, B. Mishra, H. R. Nielson (Eds.), Transactions on Computational Systems Biology VII, LNCS 4230, Springer, 2006, pp. 1–23.

[11] F. Ciocchetta, J. Hillston, Calculi for biological systems, in: Formal Methods for Computational Systems Biology, SFM08, LNCS 5016, 2008, pp. 265–312.

[12] 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.

[13] 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).

[14] F. Ciocchetta, A. Degasperi, J. Heath, J. Hillston, Modelling and analysis of the NF-κB pathway in Bio-PEPA, in: C. Priami, R. Breitling, D. Gilbert, M. Heiner, A. M. Uhrmacher (Eds.), Transactions on Computational Systems Biology XII, LNCS 5945, Springer, 2010, pp. 229–262.

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

[16] 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.Gorriero (Eds.), Proceedings of the 7th International Conference on Computational Models in Systems Biology (CMSB 2009), LNCS 5688, Springer, 2009, pp. 52–67.

[17] O. E. Akman, M. L. Guerriero, L. Loewe, C. Troein, Complementary approaches to understanding the plant circadian clock, Proceedings of FBTC'10, EPTCS 19, 2010, pp 1-19.

[18] 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.

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

[20] V. Galpin, J. Hillston, F. Ciocchetta, An semi-quantitative equivalence for abstracting from fast reactions, to be presented at CompMod 2011.

[21] V. Galpin, J. Hillston, Equivalence and discretisation in Bio-PEPA, in: P. Degano, R.Gorriero (Eds.), Proceedings of the 7th International Conference on Computational Models in Systems Biology (CMSB 2009), LNCS 5688, Springer, 2009, pp. 189–204.

[22] V. Galpin, J. Hillston, A semantic equivalence for Bio-PEPA based on discretisation of continuous values, Theoretical Computer Science 412(2011) 21422161.

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

[24] 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.

[25] K. Larsen, A. Skou, Bisimulation through probabilistic testing, Information and Computation 94 (1991) 1–28.

[26] B. D. Leinweber, P. C. Leavis, Z. Grabarek, C.-L. A. Wang, K. G. Morgan, Extracellular reulated kinase (ERK) interaction with actin and the calponin homology (CH) domain of actin-binding proteins, The Biochemical Journal 344 (1999) 117–123.

[27] K.-H. Cho, S.-Y. Shin, H. W. Kim, O. Wolkenhauer, B. McFerran, W. Kolch, Mathematical modeling of the influence of RKIP on the ERK signaling pathway, in: C. Priami (Ed.), Proceedings of the First International Workshop in Computational Methods in Systems Biology (CMSB 2003), LNCS 2602, Springer, 2003, pp. 127–141.

[28] A. C. Lloyd, Distinct functions for ERKs?, Journal of Biology 5 (2006) 13.

[29] B. Schoeberl, C. Eichler-Jonsson, E. D. Gilles, G. Muller, Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors, Nature Biotechnology 20 (2002) 270–375.

[30] Y. Gong, Z. Zhang, Alternative signaling pathways: When, where and why?, FEBS Letters 579 (2005) 5265 – 5274.

[31] Y. Gong, X. Zhao, Shc-dependent pathway is redundant but dominant in MAPK cascade activation by EGF receptors: a modeling inference, FEBS Letters 554 (2003) 467 – 472.

[32] A. Clark, S. Gilmore, M. L. Guerriero, P. Kemper, On verifying Bio-PEPA models, in: Proceedings of the 8th International Conference on Computational Models in Systems Biology (CMSB 2010), ACM Press, 2010, pp. 23–32.

[33] A. Clark, S. Gilmore, J. Hillston, P. Kemper, Verification and testing of biological models, in: Proceedings of the 2010 Winter Simulation Conference, 2010, pp. 620–630.

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

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

[36] 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.

[37] 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.

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

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

A. Example: grouping of species

Let $\mathcal{R} = \{(k - (j + l), j, l, n - k), (k - (j + l), j + l, n - k) \mid 0 \leq k \leq n, 0 \leq j + l \leq k\}$. The rates are considered symbolically in describing transitions after which the relationship necessary between the rates of the different models to obtain g-bisimilarity is considered. Compression bisimilarity [22] is a useful tool to identify the different cases to be considered in the proof. It quotients the state space into classes which have the same actions available, and it has been applied to the larger model M to determine these classes, and leads to the conclusion that there are eight distinct cases necessary to prove that $\mathcal{R}$ is a g-bisimulation. These are illustrated in the table below. Each row under a case heading represents a group of transitions, specified by the constraints in the case heading. Context and angle brackets are omitted. The left hand column is the source of the transition and has been written as simply as possible. The next column represents the label on the transition in the stochastic relation so that it is clear to see how g is calculated. The third column represents the target of the transition, and is written to make it immediate to check that the resulting pairs are in $\mathcal{R}$. The last column gives the value of g. The change/jump vectors for the actions are as follows.

$$\begin{array}{rcl@{\qquad}rcl}
\alpha_1 & : & (-1, +1, 0, 0) & \beta_1 & : & (0, -1, 0, +1) \\
\alpha_1 & : & (-1, 0, +1, 0) & \beta_2 & : & (0, 0, -1, +1) \\
\alpha & : & (-1, +1, 0) & \beta & : & (0, -1, +1)
\end{array}$$

Transitions with four-element vectors are from M and those with three-element vectors from M'. Hence the table makes it straightforward to check which transitions are potential matches and from this the relationships between the rates in the two models can be established.

From Cases A, B, C and D, it can be seen that for $\mathcal{R}$ to be a g-bisimulation, it is necessary that $r = r_1 + r_2$. From Cases B and G, $s_1 = s$ is required, and from C and H, $s_2 = s$. Furthermore, Cases A and E illustrate the requirement for $s_1 j + s_2 l = s(j + l)$. This is already implied by $s_1 = s_2 = s$ since $(s_1 j + s_2 l)/(j + l) = (sj + sl)/(j + l) = s$. Hence under these constraints, $\mathcal{R}$ is a g-bisimulation and $\langle \mathcal{T}, M' \rangle \sim_g \langle \mathcal{T}, M \rangle$.

Source	**Label**	**Target**	$\mathbf{g}(\theta, \mathbf{P}, \mathbf{P}')$
Case A: $2 \leq k \leq n, 2 \leq j{+}l \leq k{-}1$			
$(k-(j+l), j, l, n-k)$	$(\alpha_1, r_1(k-j+l))$	$(k-((j+1)+l), j+1, l, n-k)$	$(\alpha, (r_1+r_2)k)$
$(k-(j+l), j, l, n-k)$	$(\alpha_2, r_2(k-j+l))$	$(k-(j+(l+1)), j, l+1, n-k)$	$(\alpha, (r_1+r_2)k)$
$(k-(j+l), j, l, n-k)$	$(\beta_1, s_1 j)$	$((k-1)-((j-1)+l), j-1, l, n-(k-1))$	$(\beta, s_1 j)$
$(k-(j+l), j, l, n-k)$	$(\beta_2, s_2 l)$	$((k-1)-(j+(l-1)), j, l-1, n-(k-1))$	$(\beta, s_2 l)$
$(k-(j+l), j+l, n-k)$	$(\alpha, r(k-j+l))$	$(k-(j+l+1), j+l+1, n-k)$	(α, rk)
$(k-(j+l), j+l, n-k)$	$(\beta, s(j+l))$	$((k-1)-(j+l-1), j+l-1, n-(k-1))$	$(\beta, s(j+l))$
Case B: $2 \leq k \leq n, 1 \leq j \leq k{-}1, l = 0$			
$(k-j, j, 0, n-k)$	$(\alpha_1, r_1(k-j))$	$(k-(j+1), j+1, 0, n-k)$	$(\alpha, (r_1+r_2)(k-j))$
$(k-j, j, 0, n-k)$	$(\alpha_2, r_2(k-j))$	$(k-(j+1), j, 1, n-k)$	$(\alpha, (r_1+r_2)(k-j))$
$(k-j, j, 0, n-k)$	$(\beta_1, s_1 j)$	$((k-1)-(j-1), j-1, 0, n-(k-1))$	$(\beta, s_1 j)$
$(k-j, j, n-k)$	$(\alpha, r(k-j))$	$(k-(j+1), j+1, n-k)$	$(\alpha, r(k-j))$
$(k-j, j, n-k)$	(β, sj)	$((k-1)-(j-1), j-1, n-(k-1))$	(β, sj)
Case C: $2 \leq k \leq n, 1 \leq l \leq k{-}1, j = 0$			
$(k-l, 0, l, n-k)$	$(\alpha_1, r_1(k-l))$	$(k-(l+1), 1, l, n-k)$	$(\alpha, (r_1+r_2)(k-l))$
$(k-l, 0, l, n-k)$	$(\alpha_2, r_2(k-l))$	$(k-(l+1), 0, l+1, n-k)$	$(\alpha, (r_1+r_2)(k-l))$
$(k-l, 0, l, n-k)$	$(\beta_2, s_2 l)$	$((k-1)-(l-1), 0, l-1, n-(k-1))$	$(\beta, s_2 l)$
$(k-l, l, n-k)$	$(\alpha, r(k-l))$	$(k-(l+1), l+1, n-k)$	$(\alpha, r(k-l))$
$(k-l, l, n-k)$	(β, sl)	$((k-1)-(l-1), l-1, n-(k-1))$	(β, sl)
Case D: $1 \leq k \leq n, j = 0, l = 0$			
$(k, 0, 0, n-k)$	$(\alpha_1, r_1 k)$	$(k-1, 1, 0, n-k)$	$(\alpha, (r_1+r_2)k)$
$(k, 0, 0, n-k)$	$(\alpha_1, r_2 k)$	$(k-1, 0, 1, n-k)$	$(\alpha, (r_1+r_2)k)$
$(k, 0, n-k)$	(α, rk)	$(k-1, 1, n-k)$	(α, rk)
Case E: $2 \leq k \leq n, j{+}l = k, j \geq 1, l \geq 1$			
$(0, j, l, n-k)$	$(\beta_1, s_1 j)$	$((k-1)-j-1)+l), j-1, l, n-k-1))$	$(\beta, s_1 j)$
$(0, j, l, n-k)$	$(\beta_2, s_2 l)$	$((k-1)-j+(l-1)), j, l-1, n-k-1))$	$(\beta, s_2 l)$
$(0, j+l, n-k)$	$(\beta, s(j+l))$	$((k-1)-j+l-1), j+l-1, n-k-1))$	$(\beta, s(j+l))$
Case F: $2 \leq k \leq n, j = k, l = 0$			
$(0, j, 0, n-k)$	$(\beta_1, s_1 j)$	$((k-1)-j-1), j-1, 0, n-k-1))$	$(\beta, s_1 j)$
$(0, j, n-k)$	(β, sj)	$((k-1)-j-1), j-1, n-k-1))$	(β, sj)
Case G: $2 \leq k \leq n, l = k, j = 0$			
$(0, 0, l, n-k)$	$(\beta_2, s_2 l)$	$((k-1)-l-1), 0, l-1, n-k-1))$	$(\beta, s_2 l)$
$(0, l, n-k)$	(β, sl)	$((k-1)-l-1), l-1, n-k-1))$	(β, sl)
Case H: $k = 0, j = 0, l = 0$			
$(0, 0, 0, n)$	–	–	–
$(0, 0, n)$	–	–	–

B. Example: alternative signalling pathways

These are presented using the style of component description that is used in the Bio-PEPA Eclipse Plug-in [24]. As before, *EG* is an abbreviation for $(EGF{-}EGFR^{*})2{-}GAP$. Additionally, p is used instead of a superscript asterisk to indicate phosphorylation, <*> indicates cooperation on all shared reactions, << indicates the role of reactant and >> the role of product.

B.1. Shc-independent pathway

```
EG                  = v16 <<                  + v34 >>;
Grb2                = v16 <<                  + v35 >>;
Sos                 = v17 <<                  + v35 >>;
Ras_GTPp            = v20 <<;
Ras_GDP             = v18 <<                  + v21 >>;
Ras_GTP             =                           v19 >>;
EG_Grb2             = v17 <<                  + v16 >>;
EG_Grb2_Sos         = v18 << + v34 << + v20 << + v17 >> + v19 >>
                                                       + v21 >>;
Grb2_Sos            = v35 <<                  + v34 >>;
EG_Grb2_Sos_Ras_GTP = v21 <<                  + v20 >>;
EG_Grb2_Sos_Ras_GDP = v19 <<                  + v18 >>;

EG[0]                  <*> Grb2[0]        <*> Sos[0]      <*>
Ras_GTPp[0]            <*> Ras_GDP[0]     <*> Ras_GTP[0]  <*>
EG_Grb2[0]             <*> EG_Grb2_Sos[0] <*> Grb2_Sos[0] <*>
EG_Grb2_Sos_Ras_GTP[0] <*> EG_Grb2_Sos_Ras_GDP[0]
```

B.2. Shc-dependent pathway

```
EG                        = v22 <<           + v32 >> + v37 >> + v39 >>;
Shc                       = v22 <<           + v36 >>;
Shcp                      = v36 << + v38 <<  + v33 >> + v37 >>;
Grb2                      = v38 << + v24 <<  + v35 >>;
Sos                       = v25 << + v40 <<  + v35 >>;
Ras_GTPp                  = v30 <<;
Ras_GDP                   = v26 <<           + v31 >>;
Ras_GTP                   =                    v27 >>;
Grb2_Sos                  = v35 << + v41 <<  + v33 >>;
EG_Shc                    = v23 <<           + v22 >>;
EG_Shcp                   = v24 << + v37 <<  + v23 >>
                                   + v41 <<;
Shcp_Grb2                 = v40 <<           + v38 >> + v39 >>;
Shcp_Grb2_Sos             = v33 <<           + v32 >> + v40 >>;
EG_Shcp_Grb2              = v25 << + v39 <<  + v24 >>;
EG_Shcp_Grb2_Sos          = v26 << + v30 <<  + v25 >> + v27 >>
                                   + v32 <<  + v31 >> + v41 >>;
EG_Shcp_Grb2_Sos_Ras_GTP = v31 <<           + v30 >>;
EG_Shcp_Grb2_Sos_Ras_GDP = v27 <<           + v26 >>;
```

```
EG[0]       <*> Shc[0]      <*> Shcp[0]                   <*>
Grb2[0]     <*> Sos[0]                                    <*>
Ras_GTPp[0] <*> Ras_GDP[0]  <*> Ras_GTP[0]                <*>
Grb2_Sos[0] <*> EG_Shc[0]   <*> EG_Shcp[0]                <*>
Shcp_Grb2[0]                <*> Shcp_Grb2_Sos[0]          <*>
EG_Shcp_Grb2[0]             <*> EG_Shcp_Grb2_Sos[0]       <*>
EG_Shcp_Grb2_Sos_Ras_GTP[0] <*> EG_Shcp_Grb2_Sos_Ras_GDP[0]
```