Modelling the influence of RKIP on the ERK signalling pathway using the stochastic process algebra PEPA

Muffy Calder[1], Stephen Gilmore[2] and Jane Hillston[2]

[1] Department of Computing Science, The University of Glasgow, Glasgow, Scotland. email muffy@dcs.gla.ac.uk

[2] Laboratory for Foundations of Computer Science, The University of Edinburgh, Scotland. email stg@inf.ed.ac.uk, jeh@inf.ed.ac.uk

Abstract. This paper examines the influence of the Raf Kinase Inhibitor Protein (RKIP) on the Extracellular signal Regulated Kinase (ERK) signalling pathway [2] through modelling in a Markovian process algebra, PEPA [7]. Two models of the system are presented, a reagent-centric view and a pathway-centric view. Each model affords a different perspective of the pathway and analysis. We demonstrate the two models to be formally equivalent using the timing-aware bisimulation defined over PEPA models and discuss the biological significance.

1 Introduction

In recent years several authors have investigated the use of Petri nets and process algebras – techniques originating in theoretical computer science – for representing the biochemical pathways within and between cells [9, 10, 6]. Largely, the previous work has focussed on capturing the appropriate functionality at the molecular level and analysis is through simulation. In this paper we present a preliminary exploration of the analytical application of a process algebra to a biochemical pathway with feedback. Our goal is to develop more than one representation, suitable for different forms of analysis. We prove the two representations to be equivalent (i.e. bisimilar).

The process algebra which we use is Hillston's PEPA [7], a Markovian process algebra which incorporates stochastic durations and probabilistic choices. The system which we consider is the Ras/Raf-1/MEK/ERK signalling pathway, as presented in [2]. We believe that our modelling is novel because we are able to combine performance and different modelling viewpoints.

We propose that process algebra models are appropriate in this domain for several reasons. First, an algebraic formulation of the model makes clear the interactions between the biochemical entities, or substrates. This is not always apparent in the classical, ordinary differential equation (ODE) models. Second, an algebraic approach permits comparison of high level descriptions. For example, when one is first building up a picture of a pathway from experimental evidence, it may be natural to describe the pathway in a fine-grained, distributed

fashion, e.g. each substrate (in this case a protein) is described in terms of its interactions. That is, each (collection of a) protein is a process and all processes run in parallel, synchronising accordingly. But later, we may prefer a higher level view of a pathway which describes how a pathway is composed of (perhaps already well known) sub-pathways. Indeed we may wish to derive the latter from the former, or vice-versa. Third, a stochastic process approach allows reasoning about livelocks, deadlocks, and the performance of the behaviour of the pathway in the long-run.

This paper is an extended version of the earlier paper [1]. As previously, we concentrate primarily on alternative approaches to constructing a representation of a pathway. We show that two contrasting representations can indeed be identified. Moreover they can be formally shown to be equivalent. The novelty of this paper lies in the systematic transformation between the alternative representations which are presented in algorithmic form. The analysis of the model has also been somewhat extended.

In the next section we give a brief overview of cell signalling and the Ras/Raf-1/MEK/ERK pathway. In section 3 we give two different PEPA formulations of the pathway: the first is reagent-based (i.e. distributed) and the second is pathway-based. In section 4 we compare the two models and show them to be bisimilar. Section 5 contains some analysis of the underlying continuous time Markov model. Transformation between the two styles of representation is presented in section 6. There follows a discussion of related work and our conclusions.

2 RKIP and the ERK Pathway

The most fundamental cellular processes are controlled by extracellular signalling [4]. This signalling, or communication between cells, is based upon the release of signalling molecules, which migrate to other cells and deliver stimuli to them (e.g. protein phosphorylation). Cell signalling is of special interest to cancer researchers because when cell signalling pathways operate abnormally, cells divide uncontrollably.

The Ras/Raf-1/MEK/ERK pathway (also called Ras/Raf, or ERK pathway) is a ubiquitous pathway that conveys mitogenic and differentiation signals from the cell membrane to the nucleus. Briefly, Ras is activated by an external stimulus, it then binds to and activates Raf-1 (to become Raf-1*, "activated" Raf) which in turn activates MEK and then ERK. This "cascade" of protein interaction controls cell differentiation, the effect being dependent upon the activity of ERK. A current area of experimental scientific investigation is the role the kinase inhibitor protein RKIP plays in the behaviour of this pathway: the hypothesis is that it inhibits activation of Raf and thus can "dampen" down the ERK pathway. Certainly there is much evidence that RKIP inhibits the malignant transformation by Ras and Raf oncogenes in cell cultures and it is reduced in tumours. Thus good models of these pathways are required to understand the role of RKIP and develop new therapies. Moreover, an understanding of

the functioning and structure of this pathway may lead to more general results applicable to other pathways.

Here, we consider how RKIP regulates the activity of the Raf-1/MEK/ERK module of the ERK pathway, as presented in [2]. This paper [2] presents a number of mathematical models in the form of nonlinear ODEs and difference equations representing the (enzyme) kinetic reactions, based on a graphical representation given in Figure 1. This figure is taken from [2], with some additions. Specifically, we have added MEK and an associated complex, following discussions with the authors[1].

We take Figure 1 as our starting point, and explain informally, its meaning. Each node is labelled by the protein (or substrate, we use the two interchangably) it denotes. For example, Raf-1, RKIP and Raf-1*/RKIP are proteins, the last being a complex built up from the first two. It is important to note that Raf-1*/RKIP is simply a *name*, following biochemical convention; the / symbol is not an operator (in this context). A sufffix -P or -PP denotes a phosyphorylated protein, for example MEK-PP and ERK-PP. Each protein has an associated concentration, denoted by $m1$, $m2$ etc. *Reactions* define how proteins are built up and broken down. We refer to the former as an association, or forward reaction, and the latter as a disassociation, or backward reaction. Associations are typically many to one, and disassociations one to many, relations. In the figure, bi-directional arrows denote both forward and backward reactions; unidirectional arrows denote disassociations. For example, Raf-1* and RKIP react (forwards) to form Raf-1*/RKIP, and Raf-1/RKIP disassociates (a backward reaction) into Raf-1* and RKIP. Reactions do not necessarily come in pairs; for example, Raf-1*/RKIP/ERK-PP disassociates into Raf-1*, ERK and RKIP-P. Each reaction has a rate denoted by the rate constants $k1$, $k2$, etc. These are given in the rectangles, with $kn/kn+1$ denoting that kn is the forward rate and $kn+1$ the backward rate. So for example, Raf-1* and RKIP react (forwards) with rate $k1$, and Raf-1/RKIP disassociates with rate $k2$.

Initially, all concentrations are unobservable, except for m_1, m_2, m_7, m_9, and m_{10} [2].

Figure 1 gives only a static, abstract view of the pathway; the dynamic behaviour is quite complex, particularly because some substrates are involved in more than one reaction. In the next section we develop two process algebraic models which capture that dynamic behaviour.

3 Modelling the ERK signalling pathway in PEPA

In this section we present two stochastic process algebra models of the ERK signalling pathway.

The two models presented here encode different views of the underlying biochemistry. The first is a reagent-centric view, focussing on the variations in concentrations of the reagents, fluctuating with phosphorylation and product for-

[1] Analysis of our original model(s) indicated a problem with MEK and prompted us to contact an author of [2] who confirmed that there was an omission.

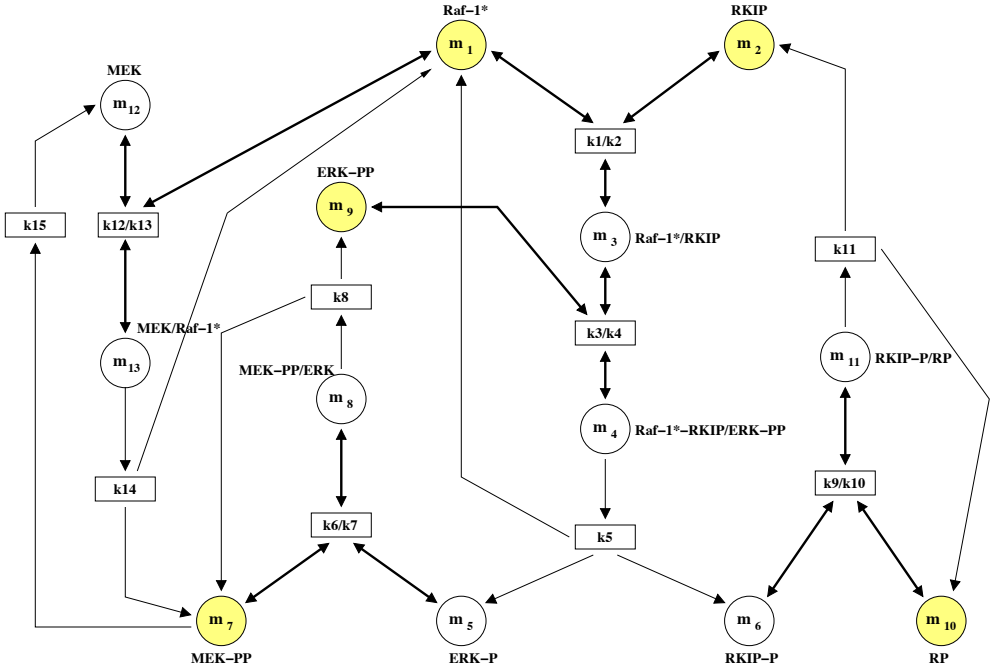

Fig. 1. RKIP inhibited ERK pathway

mation, i.e. with association and disassociation reactions. This model provides a fine-grained, distributed view of the system. The second is a pathway-centric view, tracking the legitimate serialisations of activities. This model provides a coarser grained, more abstract view of the same system.

For some purposes in biological study the former view provides the right conceptual tools and powers the programme of analysis. For other purposes the pathway-centric view brings to the fore the dynamics of greatest interest. A major contribution of this paper is the unification of both views.

We express both models in the PEPA stochastic process algebra [7]. We assume some familiarity with this process algebra; a brief introduction to PEPA is contained in Appendix A. All activities in PEPA are timed. Specifically, their durations are quantified using exponentially-distributed random variables. The PEPA algebra supports multi-way cooperations between components: the result of synchronising on an activity α is thus another α, available for further synchronisation. The multi-way synchronisation of PEPA makes this process algebra ideally suited to this domain.

Each reaction in the pathway is represented by a multi-way synchronisation – on the reagents of the reaction[2]. We refer to reagents as *producers* and *consumers*, depending upon their role within the reaction. Table 1 gives the pro-

[2] We agree with the authors of [9] – reactions are fundamentally synchronous.

ducers and consumers for reactions in the pathway. The first column names the reaction using the following convention. Reactions which are forward and backward are called *react*, with a prefix which is the associated rate constant. For example, $k1react$ is the name of the reaction between Raf-1* and RKIP, to produce Raf-1*/RKIP. Thus $k1react$ is a 3-way synchronisation. Reactions which are only disassociations are called *product* (because they produce *products*); again, the prefix denotes the associated rate constant. Table 1 gives only the forward reactions for the reactions which are both forward and backwards; to obtain the associated backward descriptions, replace Producer by Consumer and vice-versa.

Reaction	**Producer(s)**	**Consumer(s)**
k1react	{ Raf-1*, RKIP }	{ Raf-1*/RKIP }
k3react	{ ERK-PP, Raf-1*/RKIP }	{ Raf-1*/RKIP/ERK-PP }
k6react	{ MEK-PP, ERK-P }	{ MEK-PP/ERK }
k9react	{ RKIP-P, RP }	{ RKIP-P/RP }
k12react	{ MEK, Raf-1* }	{ MEK/Raf-1* }
k5product	{ Raf-1*/RKIP/ERK-PP }	{ ERK-P, RKIP-P, Raf-1* }
k8product	{ MEK-PP/ERK }	{ MEK-PP, ERK-PP }
k11product	{ RKIP-P/RP }	{ RKIP, RP }
k14product	{ MEK/Raf-1* }	{ Raf-1*, MEK-PP }
k15product	{ MEK-PP }	{ MEK }

Table 1. Reactions in the pathway

3.1 Modelling centred on reagents

The reagent-centred model is presented in Figures 2 and 3. In this view, we represent concentrations by discrete values. We distinguish between high (i.e. observable) and low (i.e. unobservable) concentrations of reagents. The former implies that a reagent *can* participate (as a producer) in a forward reaction; the latter implies that a reagent *can* participate (as a consumer) in a product, or (as a producer) in a backward reaction. Otherwise, the substrate is inert, with respect to a reaction. We define the behaviour of each substrate in turn, for each concentration. Thus there are $2n$ equations, where n is the number of proteins. We adopt the naming convention that high concentrations have a H subscript and low concentrations have a L subscript.

Most equations involve a choice between alternative behaviours (notated by +). For example, even in one of the simplest cases, RKIP, where there is a simple cycle between high and low concentrations, there is still a choice of how to return to a high concentration (by a backwards reaction, or through a product). Most behaviours are more complex.

The equations define the possible reactions within the pathway. All of the permissible interleavings of these reactions are obtained from the (synchronised) parallel composition of these components. Figure 3 shows how these are composed in the PEPA algebra. The composition operator ($\bowtie$) is indexed by an activity set (i.e. the events whose participants must be synchronised). The left and right operands must cooperate on these activities, introducing a synchronisation point. The degenerate case of this composition operator (where the set is empty) provides the expected unrestricted parallel composition of the components, allowing all possible interleavings without synchronisation. This case is denoted by $\|$ (there is one occurrence).

The initial state of the model has high concentrations of some reagents and low concentrations of the others, as described in the previous section. Therefore, in Figure 3, proteins with an initial concentration are initially high; all others are low.

3.2 Modelling centred on pathways

A different view is afforded by the pathway-centric perspective. This de-emphasises reagents and emphasises sub-pathways within the signalling pathway. In this model, given in Figure 4, there are five (sub)pathways, one for each substrate with an initial concentration. Thus $Pathway_{10}$ corresponds to the pathway from RP (m_{10}), $Pathway_{20}$ to RKIP (m_2), $Pathway_{30}$ to ERK-PP (m_9), $Pathway_{40}$ to Raf-1* (m_1), and $Pathway_{50}$ to MEK-PP (m_7). Each (sub)pathway describes, in effect, how a substrate is consumed and then, eventually, replenished.

It is important to note that none of these (sub)pathways is *closed*, i.e. there are reactions with edges which are directed to/from outside of the (sub)pathway. Figure 6 gives a diagrammatic representation of the simplest pathway, $Pathway_{10}$. In this case, the pathway is not closed because there are two missing edges associated with *k9react* and *k11product*.

This presentation facilitates the direct verification of simple properties of the model such as "the first observable activity is event X". For example, an initial syntactic inspection of this model would lead to the conclusion that the first activity is one of *k1react*, *k3react*, *k9react* or *k15product*. Processing the model with the PEPA Workbench [5] confirms that the initial model configuration allows only *k15product* and *k1react*, the others are not permitted because some necessary participants are not initially ready to engage in these reactions.

4 Comparison of reagent and pathway-centric models

The pathway-centric model captures longer chains of behaviour flow within the system, leading to a smaller number of component definitions. Differentiating fewer components in the pathways model leads to a simpler composition of model components, presented in Figure 5. This is not only a matter of presentation. A larger state vector representation occupies more memory so the pathway-centric representation could potentially scale better to more detailed models of

$$
\begin{aligned}
\text{Raf-1}^*_{\text{H}} &\stackrel{def}{=} (\mathit{k1react}, k_1).\text{Raf-1}^*_{\text{L}} + (\mathit{k12react}, k_{12}).\text{Raf-1}^*_{\text{L}} \\
\text{Raf-1}^*_{\text{L}} &\stackrel{def}{=} (\mathit{k5product}, k_5).\text{Raf-1}^*_{\text{H}} + (\mathit{k2react}, k_2).\text{Raf-1}^*_{\text{H}} \\
&\quad + (\mathit{k13react}, k_{13}).\text{Raf-1}^*_{\text{H}} + (\mathit{k14product}, k_{14}).\text{Raf-1}^*_{\text{H}}
\end{aligned}
$$

$$
\begin{aligned}
\text{RKIP}_{\text{H}} &\stackrel{def}{=} (\mathit{k1react}, k_1).\text{RKIP}_{\text{L}} \\
\text{RKIP}_{\text{L}} &\stackrel{def}{=} (\mathit{k11product}, k_{11}).\text{RKIP}_{\text{H}} + (\mathit{k2react}, k_2).\text{RKIP}_{\text{H}}
\end{aligned}
$$

$$
\begin{aligned}
\text{MEK}_{\text{H}} &\stackrel{def}{=} (\mathit{k12react}, k_{12}).\text{MEK}_{\text{L}} \\
\text{MEK}_{\text{L}} &\stackrel{def}{=} (\mathit{k13react}, k_{13}).\text{MEK}_{\text{H}} + (\mathit{k15product}, k_{15}).\text{MEK}_{\text{H}}
\end{aligned}
$$

$$
\begin{aligned}
\text{MEK/Raf-1}^*_{\text{H}} &\stackrel{def}{=} (\mathit{k14product}, k_{14}).\text{MEK/Raf-1}^*_{\text{L}} + (\mathit{k13react}, k_{13}).\text{MEK/Raf-1}^*_{\text{L}} \\
\text{MEK/Raf-1}^*_{\text{L}} &\stackrel{def}{=} (\mathit{k12react}, k_{12}).\text{MEK/Raf-1}^*_{\text{H}}
\end{aligned}
$$

$$
\begin{aligned}
\text{MEK-PP}_{\text{H}} &\stackrel{def}{=} (\mathit{k6react}, k_6).\text{MEK-PP}_{\text{L}} + (\mathit{k15product}, k_{15}).\text{MEK-PP}_{\text{L}} \\
\text{MEK-PP}_{\text{L}} &\stackrel{def}{=} (\mathit{k8product}, k_8).\text{MEK-PP}_{\text{H}} + (\mathit{k7react}, k_7).\text{MEK-PP}_{\text{H}} \\
&\quad + (\mathit{k14product}, k_{14}).\text{MEK-PP}_{\text{H}}
\end{aligned}
$$

$$
\begin{aligned}
\text{ERK-PP}_{\text{H}} &\stackrel{def}{=} (\mathit{k3react}, k_3).\text{ERK-PP}_{\text{L}} \\
\text{ERK-PP}_{\text{L}} &\stackrel{def}{=} (\mathit{k8product}, k_8).\text{ERK-PP}_{\text{H}} + (\mathit{k4react}, k_4).\text{ERK-PP}_{\text{H}}
\end{aligned}
$$

$$
\begin{aligned}
\text{ERK-P}_{\text{H}} &\stackrel{def}{=} (\mathit{k6react}, k_6).\text{ERK-P}_{\text{L}} \\
\text{ERK-P}_{\text{L}} &\stackrel{def}{=} (\mathit{k5product}, k_5).\text{ERK-P}_{\text{H}} + (\mathit{k7react}, k_7).\text{ERK-P}_{\text{H}}
\end{aligned}
$$

$$
\begin{aligned}
\text{MEK-PP/ERK}_{\text{H}} &\stackrel{def}{=} (\mathit{k8product}, k_8).\text{MEK-PP/ERK}_{\text{L}} + (\mathit{k7react}, k_7).\text{MEK-PP/ERK}_{\text{L}} \\
\text{MEK-PP/ERK}_{\text{L}} &\stackrel{def}{=} (\mathit{k6react}, k_6).\text{MEK-PP/ERK}_{\text{H}}
\end{aligned}
$$

$$
\begin{aligned}
\text{Raf-1}^*\text{/RKIP}_{\text{H}} &\stackrel{def}{=} (\mathit{k3react}, k_3).\text{Raf-1}^*\text{/RKIP}_{\text{L}} + (\mathit{k2react}, k_2).\text{Raf-1}^*\text{/RKIP}_{\text{L}} \\
\text{Raf-1}^*\text{/RKIP}_{\text{L}} &\stackrel{def}{=} (\mathit{k1react}, k_1).\text{Raf-1}^*\text{/RKIP}_{\text{H}} + (\mathit{k4react}, k_4).\text{Raf-1}^*\text{/RKIP}_{\text{H}}
\end{aligned}
$$

$$
\begin{aligned}
\text{Raf-1}^*\text{/RKIP/ERK-PP}_{\text{H}} &\stackrel{def}{=} (\mathit{k5product}, k_5).\text{Raf-1}^*\text{/RKIP/ERK-PP}_{\text{L}} \\
&\quad + (\mathit{k4react}, k_4).\text{Raf-1}^*\text{/RKIP/ERK-PP}_{\text{L}} \\
\text{Raf-1}^*\text{/RKIP/ERK-PP}_{\text{L}} &\stackrel{def}{=} (\mathit{k3react}, k_3).\text{Raf-1}^*\text{/RKIP/ERK-PP}_{\text{H}}
\end{aligned}
$$

$$
\begin{aligned}
\text{RKIP-P}_{\text{H}} &\stackrel{def}{=} (\mathit{k9react}, k_9).\text{RKIP-P}_{\text{L}} \\
\text{RKIP-P}_{\text{L}} &\stackrel{def}{=} (\mathit{k5product}, k_5).\text{RKIP-P}_{\text{H}} + (\mathit{k10react}, k_{10}).\text{RKIP-P}_{\text{H}}
\end{aligned}
$$

$$
\begin{aligned}
\text{RP}_{\text{H}} &\stackrel{def}{=} (\mathit{k9react}, k_9).\text{RP}_{\text{L}} \\
\text{RP}_{\text{L}} &\stackrel{def}{=} (\mathit{k11product}, k_{11}).\text{RP}_{\text{H}} + (\mathit{k10react}, k_{10}).\text{RP}_{\text{H}}
\end{aligned}
$$

$$
\begin{aligned}
\text{RKIP-P/RP}_{\text{H}} &\stackrel{def}{=} (\mathit{k11product}, k_{11}).\text{RKIP-P/RP}_{\text{L}} + (\mathit{k10react}, k_{10}).\text{RKIP-P/RP}_{\text{L}} \\
\text{RKIP-P/RP}_{\text{L}} &\stackrel{def}{=} (\mathit{k9react}, k_9).\text{RKIP-P/RP}_{\text{H}}
\end{aligned}
$$

Fig. 2. PEPA model definitions for the reagent-centric model

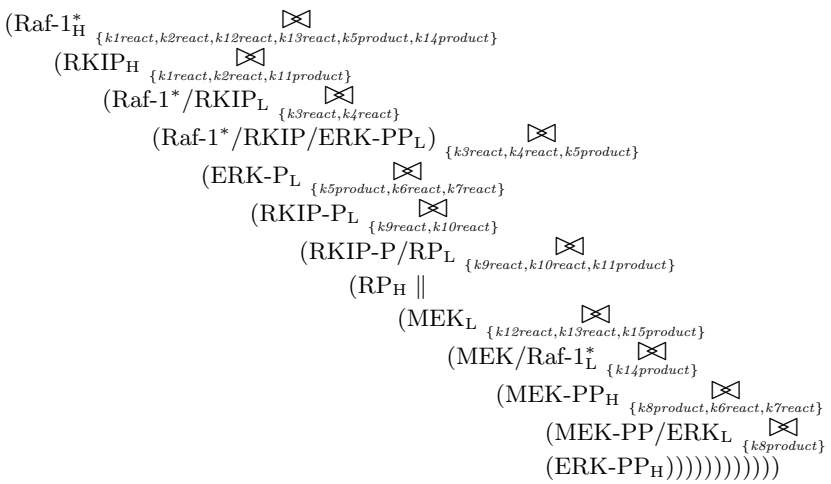

Fig. 3. PEPA model configuration for the reagent-centric model

the Ras/Raf-1/MEK/ERK signalling pathway than the reagent-centric representation. But, the disadvantage of the pathway-centric representation is that it is no longer possible to read off directly concentrations of components (i.e. there is no explicit high or low concentrations). These now have to be inferred from local observations of pathways. This is relatively easy for proteins which have initial concentrations, otherwise, the inference is non-trivial.

Fortunately, the two models are observationally equivalent, that is, the two models give rise to (timing aware) bisimilar—in fact *isomorphic*—labelled multi-transition systems. We demonstrate this relationship by plotting the statespace of the two systems, see Figure 7. There are 28 states, s_1 to s_{28}, thus it is not possible in Figure 7 to give meaningful labels. In Table 2 we enumerate a few of the states. We give the name from the reagent-centric model first, followed by the name of the equivalent state from the pathway-centric model. In all cases, the synchronisation operator $\bowtie$ is removed.

The consequence of this result is that the two models give rise to the same Markov chain representations which can be solved to find the steady-state distribution. The analysis is described in the following section.

5 Model analysis

We used the PEPA Workbench [5] to analyse our models. The Workbench implements the operational semantics (of PEPA) to generate Continuous-Time Markov Chain (CTMC) models of system descriptions, and it provides analysis tools. We used the Workbench to test for deadlocks in our models (initially, there were several, this is how we discovered the problem with MEK), then once

$$Pathway_{10} \stackrel{def}{=} (k9react, k_9).Pathway_{11}$$
$$Pathway_{11} \stackrel{def}{=} (k11product, k_{11}).Pathway_{10} + (k10react, k_{10}).Pathway_{10}$$

$$Pathway_{20} \stackrel{def}{=} (k1react, k_1).Pathway_{21}$$
$$Pathway_{21} \stackrel{def}{=} (k3react, k_3).Pathway_{22} + (k2react, k_2).Pathway_{20}$$
$$Pathway_{22} \stackrel{def}{=} (k5product, k_5).Pathway_{23} + (k4react, k_4).Pathway_{21}$$
$$Pathway_{23} \stackrel{def}{=} (k9react, k_9).Pathway_{24}$$
$$Pathway_{24} \stackrel{def}{=} (k11product, k_{11}).Pathway_{20} + (k10react, k_{10}).Pathway_{23}$$

$$Pathway_{30} \stackrel{def}{=} (k3react, k_3).Pathway_{31}$$
$$Pathway_{31} \stackrel{def}{=} (k5product, k_5).Pathway_{32} + (k4react, k_4).Pathway_{30}$$
$$Pathway_{32} \stackrel{def}{=} (k6react, k_6).Pathway_{33}$$
$$Pathway_{33} \stackrel{def}{=} (k8product, k_8).Pathway_{30} + (k7react, k_7).Pathway_{32}$$

$$Pathway_{40} \stackrel{def}{=} (k1react, k_1).Pathway_{41} + (k12react, k_{12}).Pathway_{43}$$
$$Pathway_{41} \stackrel{def}{=} (k2react, k_2).Pathway_{40} + (k3react, k_3).Pathway_{42}$$
$$Pathway_{42} \stackrel{def}{=} (k5product, k_5).Pathway_{40} + (k4react, k_4).Pathway_{41}$$
$$Pathway_{43} \stackrel{def}{=} (k13react, k_{13}).Pathway_{40} + (k14product, k_{14}).Pathway_{40}$$

$$Pathway_{50} \stackrel{def}{=} (k15product, k_{15}).Pathway_{51} + (k6react, k_6).Pathway_{53}$$
$$Pathway_{51} \stackrel{def}{=} (k12react, k_{12}).Pathway_{52}$$
$$Pathway_{52} \stackrel{def}{=} (k13react, k_{13}).Pathway_{51} + (k14product, k_{14}).Pathway_{50}$$
$$Pathway_{53} \stackrel{def}{=} (k8product, k_8).Pathway_{50} + (k7react, k_7).Pathway_{50}$$

Fig. 4. PEPA model definitions for the pathway-centric model

$$((((Pathway_{50} \underset{\{k12react, k13react, k14product\}}{\bowtie} Pathway_{40}) \underset{\{k3react, k4react, k5product, k6react, k7react, k8product\}}{\bowtie} Pathway_{30}) \underset{\{k1react, k2react, k3react, k4react, k5product\}}{\bowtie} Pathway_{20}) \underset{\{k9react, k10react, k11product\}}{\bowtie} Pathway_{10})$$

Fig. 5. PEPA model configuration for the pathway-centric model

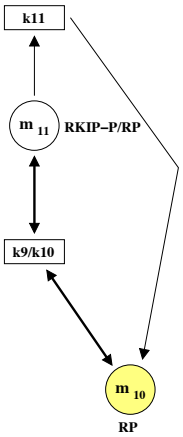

Fig. 6. $Pathway_{10}$

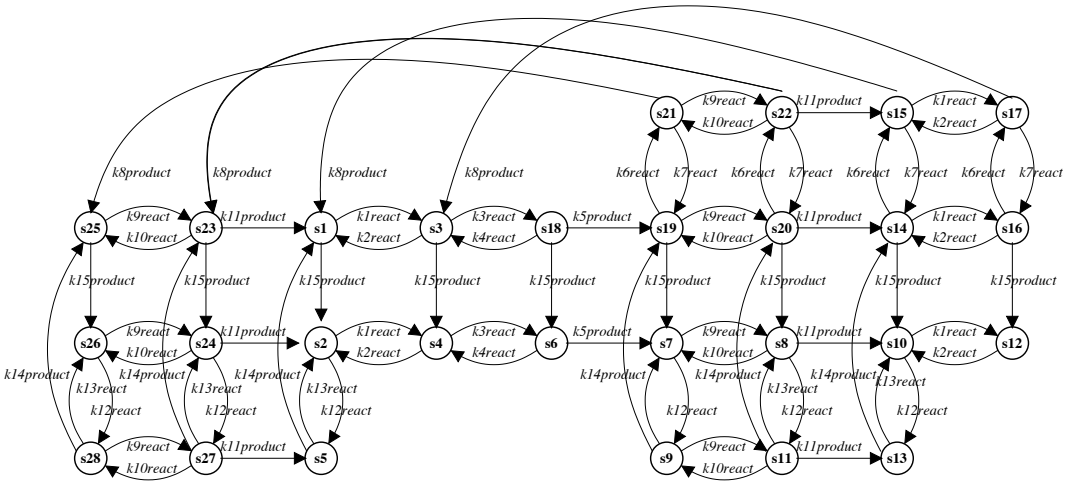

Fig. 7. The state space of the reagent and the pathway model

	Raf-1*	RKIP	Raf-1*/RKIP	Raf-1*/RKIP/ERK-PP	ERK-P	RKIP-P	RKIP-P/RP	RP	MEK	MEK/Raf-1*	MEK-PP	MEK-PP/ERK	ERK-PP	
s_1	$(H,$	$H,$	$L,$	$L,$	$L,$	$L,$	$L,$	$H,$	$L,$	$L,$	$H,$	$L,$	$H)$	$(Pwy_{50}, Pwy_{40}, Pwy_{30}, Pwy_{20}, Pwy_{10})$
s_2	$(H,$	$H,$	$L,$	$L,$	$L,$	$L,$	$L,$	$H,$	$H,$	$L,$	$L,$	$L,$	$H)$	$(Pwy_{51}, Pwy_{40}, Pwy_{30}, Pwy_{20}, Pwy_{10})$
s_3	$(L,$	$L,$	$H,$	$L,$	$L,$	$L,$	$L,$	$H,$	$L,$	$L,$	$H,$	$L,$	$H)$	$(Pwy_{50}, Pwy_{41}, Pwy_{30}, Pwy_{21}, Pwy_{10})$
s_4	$(L,$	$L,$	$H,$	$L,$	$L,$	$L,$	$L,$	$H,$	$H,$	$L,$	$L,$	$L,$	$H)$	$(Pwy_{51}, Pwy_{41}, Pwy_{30}, Pwy_{21}, Pwy_{10})$
s_5	$(L,$	$H,$	$L,$	$L,$	$L,$	$L,$	$L,$	$H,$	$L,$	$H,$	$L,$	$L,$	$H)$	$(Pwy_{52}, Pwy_{43}, Pwy_{30}, Pwy_{20}, Pwy_{10})$
s_6	$(L,$	$L,$	$L,$	$H,$	$L,$	$L,$	$L,$	$H,$	$H,$	$L,$	$L,$	$L,$	$L)$	$(Pwy_{51}, Pwy_{42}, Pwy_{31}, Pwy_{22}, Pwy_{10})$
s_7	$(H,$	$L,$	$L,$	$L,$	$H,$	$H,$	$L,$	$H,$	$H,$	$L,$	$L,$	$L,$	$L)$	$(Pwy_{51}, Pwy_{40}, Pwy_{32}, Pwy_{23}, Pwy_{10})$
s_8	$(H,$	$L,$	$L,$	$L,$	$H,$	$L,$	$H,$	$L,$	$H,$	$L,$	$L,$	$L,$	$L)$	$(Pwy_{51}, Pwy_{40}, Pwy_{32}, Pwy_{24}, Pwy_{11})$
s_9	$(L,$	$L,$	$L,$	$L,$	$H,$	$H,$	$L,$	$H,$	$L,$	$H,$	$L,$	$L,$	$L)$	$(Pwy_{52}, Pwy_{43}, Pwy_{32}, Pwy_{23}, Pwy_{10})$
s_{10}	$(H,$	$H,$	$L,$	$L,$	$H,$	$L,$	$L,$	$H,$	$H,$	$L,$	$L,$	$L,$	$L)$	$(Pwy_{51}, Pwy_{40}, Pwy_{32}, Pwy_{20}, Pwy_{10})$
$\vdots$							$\vdots$							$\vdots$
s_{28}	$(L,$	$L,$	$L,$	$L,$	$L,$	$H,$	$L,$	$H,$	$L,$	$H,$	$L,$	$L,$	$H)$	$(Pwy_{52}, Pwy_{43}, Pwy_{30}, Pwy_{23}, Pwy_{10})$

Table 2. Some bisimilar states

we had deadlock-free models, we used it to generate the CTMC and analyse its long-run probability distribution. This distribution varies as the rates associated with the activities of the PEPA model are varied so the solution of the model is relative to a particular assignment of the rates.

The steady-state probability distribution can be obtained using a number of routines from numerical linear algebra. In the case of the present model(s), we solved this using the implementation of the preconditioned biconjugate gradient method in the PEPA Workbench. This is an iterative procedure which solves systems of linear equations of moderate size very quickly.

Since both models are isomorphic, the underlying steady-state probability distributions are identical. However, it is possible to make different judgements about the two models using the PEPA state-finder which allows one to search for symbolic descriptions of states. For example, in the reagent-centric model, we used the PEPA state-finder to aggregate the probabilities of all states when ERK-PP is high, or low, for a given set of rates. That is, it aggregated the probabilities of states whose (symbolic) description has form $*\bowtie$ ERK-PP_H where * is a wildcard standing for any expression. We then repeated this with a different set of rates and compared results. In the reagent-centric model, we observed that the probability of being in a state with ERK-PP_H *decreases* as the rate $k1$ is increased, and the converse for ERK-PP_L *increases*. For example, with

$k1 = 1$ and $k1 = 100$, the probability of ERK-PP_H drops from .257 to .005. We can also plot throughput (rate × probability) against rate. Figures 8 and 9 shows two sub-plots which detail the effect of increasing the rate $k1$ on the $k14product$ and $k8product$ reactions – the production of (doubly) phosphorylated MEK and (doubly) phosphorylated ERK, respectively. These are obtained by solving the pathway model, taking each of the product and reaction rates to be unity and scaling $k1$ (keeping all other rates to be unity). The graphs show that increasing the rate of the binding of RKIP to Raf-1* dampens down the $k14product$ and $k8product$ reactions, and they quantify this information. The efficiency of the reduction is greater in the former case: the graph falls away more steeply. In the latter case the reduction is more gradual and the throughput of $k8product$ peaks at $k1 = 1$. Note that since $k5product$ is on the same pathway as $k8product$, both ERK-PP and ERK-P are similarily affected. Thus we conclude that the rate at which RKIP binds to Raf-1* (thus suppressing phosphorylation of MEK) affects the ERK pathway, as predicted (and observed); RKIP does indeed regulate the ERK pathway.

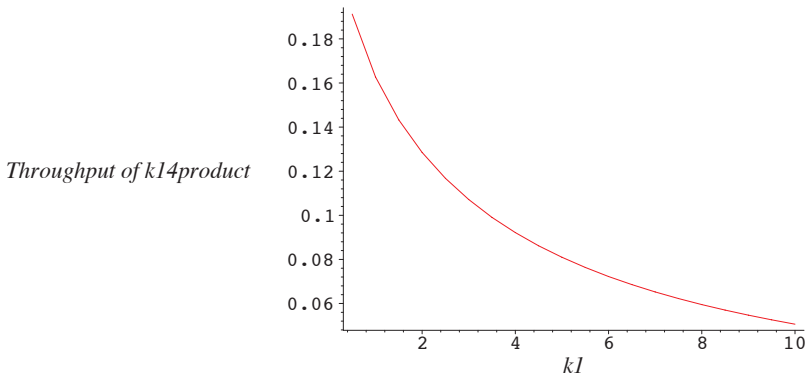

Fig. 8. Plotting the effect of $k1$ on $k14product$

6 Transformation

In this section we present a set of transformations between the two styles of representation, based on an intermediate matrix representation. Thus we define an *activity matrix* M_a which captures the relationship between reagents and reactions. This is clearly related to the stochiometry matrix of the chemical reaction. However, note that we take a more abstract view because we do not represent individual molecules. This has the consequence that the entries in the activity matrix will always be between -1 and +1.

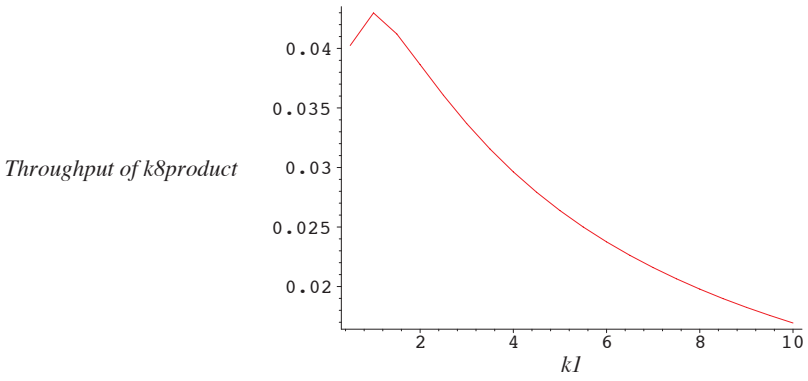

Fig. 9. Plotting the effect of $k1$ on $k8product$

In the remainder of this section we give the algorithms for each of the transformations, from the process algebra models to the matrix, and from the matrix to each form of process algebra model.

Definition 1 (Activity Matrix). *For a pathway with R reactions and S reagents, the* activity matrix M_a *is an $S \times R$ matrix, and the entries are defined as follows.*

$$(s_i, r_j) = \begin{cases} +1 & \textit{if } s_i \textit{ is a consumer of } r_j \\ -1 & \textit{if } s_i \textit{ is a producer of } r_j \\ 0 & \textit{otherwise.} \end{cases}$$

In the activity matrix each row corresponds to a single reactant, and each column to a single reaction. The relationship to the reagent-centric model is therefore straightforward. The relationship to the pathway model is somewhat more involved.

Reagent-centric model to activity matrix The algorithm for generating the activity matrix from a reagent-centric model is shown in Figure 10. The high and low annotations give a clear record of the impact of each reaction on each reagent.

Activity matrix to reagent-centric model The algorithm for generating a reagent-centric model from an activity matrix is shown in Figure 11. There are two stages to the algorithm. First, a pair of model components are formed corresponding to each row of the matrix, capturing the behaviour of that reagent in high and low concentrations. Second, the components much be configured with appropriate interactions between them. We exploit the knowledge that in this style of model each component must cooperate on all its activities. Thus the model configuration is built iteratively — as each component is added it is specified to cooperate with *the rest of the model* on all its activities.

```
// Construct a matrix of the appropriate size
Form a matrix with one row for each pair (H,L) of components
 and a column for each activity used in the process algebra definitions

// Populate the matrix
For each H component, on the appropriate row, make a -1 entry in the
 column corresponding to each activity it enables
For each L component, on the appropriate row, make a +1 entry in the
 column corresponding to each activity it enables
```

Fig. 10. Pseudo-code for transforming a reagent-centric model to an activity matrix

```
// Form the model components
For each row of the matrix assign a reactant name.
For each reactant
    make a H subscripted component based on the reactant name
    define this component to be a choice of activities as follows:
        for each -1 in the corresponding row of the activity matrix
            make an activity of the type of the appropriate column
             which results in an L subscripted component of the same name
            add this activity to the choice for the H component
    make an L subscripted component based on the reactant name
    define this component to be a choice of activities as follows:
        for each +1 in the corresponding row of the activity matrix
            make an activity of the type of the appropriate column
             which results in an H subscripted component of the same name
            add this activity to the choice for the L component

// Form the model configuration
Create an empty stack S to store reagents
For each reactant
    if this reagent has high initial concentration
        push the H-subscripted component onto S
    if this reagent has low initial concentration
        push the L-subscripted component onto S

// build the appropriate cooperation sets
pop M from the top of the stack S
While S is not empty
    pop P from the top of the stack S
    set K to be (activities of M) intersect (activities of P)
    assign M to be M cooperating with P over K

Return M
```

Fig. 11. Pseudo-code for transforming an activity matrix to a reagent-centric model

Pathway-centric model to activity matrix The algorithm for generating the activity matrix from a pathway-centric model is shown in Figure 12. The construction of the matrix to capture the involvement of pathway model components in the reactions of the system is straightforward. However, this construction will result in some duplicate rows within the matrix because some compound reagents can be seen to be intermediate states of two or more pathways (e.g. RKIP-P/RP corresponds to both $Pathway_{11}$ and $Pathway_{12}$). Thus the duplicates must be removed in order to make a canonical representation suitable for generating the reagent model.

```
// Construct a matrix of the appropriate size
Form a matrix with one row for each of the components exhibited by the pathways
 and a column for each activity used in the process algebra definitions

// Populate the matrix
For each component, on the appropriate row, make a -1 entry in the
 column corresponding to each activity it enables and a +1 entry in
 the same column of the resulting component.

//Reduce the matrix
Detect and remove identical rows
```

Fig. 12. Pseudo-code for transforming a pathway-centric model to an activity matrix

Activity matrix to pathway-centric model The algorithm for generating a pathway-centric model from an activity matrix is shown in Figure 13. The activity matrix representation is canonical in the sense that each reagent corresponds to a single row within the matrix. In order to reconstruct the sub-pathways, we need to take into account that fact that some reagents may correspond to intermediate states in two or more pathways. Thus we introduce a notion of colouring, in which one colour is associated with each sub-pathway but a single row/reagent may have several colourings indicating which sub-pathways it participates in.

The next goal is to identify the sub-pathways. We note that for all reagents all the reactions that they participate in will be part of the same sub-pathway although the converse is not true. Consequently either all the entries in a row will be coloured with some colour C or none will. However, except for the rows corresponding to initial concentrations, which are taken as the roots of our sub-pathways, any row many have any number of colours associated with it.

In order to find the sub-pathways we need to find a consumer corresponding to each producer, and vice versa, within each colour. Once such an association is made we consider the coloured matrix entry to be *paired*. The pathway is complete when all entries of that colour have been paired. In some cases there may be several candidate matrix entries for forming a pair: the corresponding

```
// Colour assignment
Assign a unique colour to each reagent which has initial concentration
Identify the rows of the matrix corresponding to these reagents
Colour each row accordingly

// Find minimal pathways
For each colour C
    while there are unpaired C entries in the matrix
        for each -1(resp. +1) entry in row s and column r coloured C
            find all entries in column r
                if there are more than one +1(resp. -1) entry
                    if none are already coloured C
                        provisionally colour each corresponding entry
                        record them as a row set
                if there is only one +1(resp. -1) entry, in row s' say
                    if it is not already coloured C
                        colour row s' with colour C
                        if s' was previously provisionally coloured with C
                            remove the provisional colouring from all other
                             elements of the row set

// Form the model components
For each colour C
    make an initial Pathway component
    make a Pathway component for each other row with C coloured entries
    for each C coloured Pathway component/row
        define the pathway component with one activity corresponding
         to each -1 column in the row whose resulting component will
         be the C coloured +1 entry in the same column


// Form the model configuration
Create an empty stack S to store pathways
Push each of the initial Pathway components onto S
// build the appropriate cooperation sets
pop M from the top of the stack S
While S is not empty
    pop P from the top of the stack S
    set K to be (activities of M) intersect (activities of P)
    assign M to be M cooperating with P over K

Return M
```

Fig. 13. Pseudo-code for transforming an activity matrix to a pathway-centric model

	$k1$	$k2$	$k3$	$k4$	$k5$	$k6$	$k7$	$k8$	$k9$	$k10$	$k11$	$k12$	$k13$	$k14$	$k15$
Raf-1*	−1	+1	0	0	+1	0	0	0	0	0	0	−1	+1	+1	0
RKIP	−1	+1	0	0	0	0	0	0	0	0	+1	0	0	0	0
Raf-1*/RKIP	+1	−1	−1	+1	0	0	0	0	0	0	0	0	0	0	0
Raf-1*/RKIP/ERK-PP	0	0	+1	−1	−1	0	0	0	0	0	0	0	0	0	0
ERK-P	0	0	0	0	+1	−1	+1	0	0	0	0	0	0	0	0
RKIP-P	0	0	0	0	+1	0	0	0	−1	+1	0	0	0	0	0
MEK-PP	0	0	0	0	0	−1	+1	+1	0	0	0	0	0	+1	−1
MEK-PP/ERK	0	0	0	0	0	+1	−1	−1	0	0	0	0	0	0	0
ERK-PP	0	0	−1	+1	0	0	0	+1	0	0	0	0	0	0	0
RP	0	0	0	0	0	0	0	0	−1	+1	+1	0	0	0	0
RKIP-P/RP	0	0	0	0	0	0	0	0	+1	−1	−1	0	0	0	0
MEK	0	0	0	0	0	0	0	0	0	0	0	−1	+1	0	+1
MEK/Raf-1*	0	0	0	0	0	0	0	0	0	0	0	+1	−1	−1	0

Fig. 14. Activity matrix of the ERK pathway

rows are collected into a set of provisionally coloured rows until it becomes clear which entry completes a minimal cycle. The other rows are then discarded.

When, for each colour, all matrix entries are paired, the sub-pathway model components can be defined in a straightforward way. It remains to form the model component but this is simply uses the same procedure as in the reagent-centric model.

As an illustration we present the activity matrix corresponding to the example presented earlier in the paper in Figure 14. This can be derived from either the reagent- or the pathway-centric model.

7 Related Work

There are several approaches to modelling biological entities using computing formalisms, for brevity, we mention only two which refer to process algebras. Regev et al [10] use the Pi-calculus to model molecules by processes, and molecular interaction by communication. Priami et al [9] use the stochastic Pi-calculus, implementing Gillespie's algorithm to govern reactions. Both these approaches involve modelling at the molecular level, whereas we have abstracted to the substrate level (i.e. concentrations). It would be interesting to relate our model(s) to these lower level ones. We note that we have found no need for mobility yet (in this pathway), this may become relevant when we consider vesicles.

8 Conclusions

We have presented two alternative PEPA models of the Raf-1/MEK/ERK module of the ERK signalling pathway and shown them to be equivalent. The

reagent-based model has explicit concentrations whilst in the pathway model the concentrations are captured only implicitly via the possible activities of each sub-pathway. The pathway-based model can thus be regarded as less directly expressive, although it captures all the same behaviour. The congruence results of PEPA with respect to strong bisimulation mean that the two representations may be used interchangeably, for example within a large model. Thus we might envisage a model in which the key pathway is modelled using the reagent-style whilst peripheral pathways are modelled using the pathway-style. Or, we may have one style of model and hypothesise the other. We believe this ability to have different views is novel in the field of modelling pathways; informal discussions with biologists confirm their interest in it.

We note with interest that we have found we require only to distinguish between high and low concentrations, further granularity adds no analytic benefit. Rather we need only model the *direction* of change (i.e. an increase or decrease of concentration).

Furthermore, we have presented transformations between the two alternative styles of representation, via an intermediate, the *activity matrix*. This means that automatic translation between represenations is possible. The transformation from an activity matrix to the pathway model has some similarities with finding the minimal T-semiflows of a Petri net. Comparing with our algorithm with the algorithms for T-semiflows [3], or the more general mathematical programming problem of finding the extremal directions of a cone [8], are yet to be investigated.

We found the multi-way synchronisation of PEPA, and the performance aspects, to be ideally suited to modelling pathway behaviour. One strength of models of the kind which we have used here is that they give rise to compact Markov chain representations which can be efficiently solved for different assignments to the rate variables in a series of experiments. This delivers the benefit that a thorough series of experiments can be conducted at modest computational cost. Process algebra opens up a host of analysis possibilities, including reasoning with probabilistic logics using probabilistic model checking. We have conducted initial investigations with the logic CSL.

Several challenges remain. For example, we wish to derive the reagent-centric model from experimental data and model spatial aspects of pathways. We have some preliminary ideas, they are the topic of future research.

Acknowledgements

Stephen Gilmore and Jane Hillston are supported by the DEGAS (Design Environments for Global ApplicationS) project IST-2001-32072 funded by the FET Proactive Initiative on Global Computing. Muffy Calder is supported by the DTI Beacon Bioscience Projects programme and thanks Walter Kolch, Beatson Cancer Research Centre, and Rainer Breitling, David Gilbert, Richard Orton, and Oliver Sturm from the Bioinformatics Research Centre, University of Glasgow, for helpful discussions.

References

1. M. Calder, S. Gilmore, and J. Hillston. Modelling the influence of rkip on the erk signalling pathway using the stochastic process algebra pepa. In *Proc. of BioConcur 2004*, pages 26–41. Danmarks Tekniske Universitet, 2004. To appear in ENTCS.
2. K.-H. Cho, S.-Y. Shin, H.-W. Kim, O. Wolkenhauer, B. McFerran, and W. Kolch. Mathematical modeling of the influence of RKIP on the ERK signaling pathway. In C. Priami, editor, *Computational Methods in Systems Biology (CSMB'03)*, volume 2602 of *LNCS*, pages 127–141. Springer-Verlag, 2003.
3. J.M. Colom and M. Silva. Convex geometry and semiflows in P/T nets. a comparative study of algorithms for computation of minimal P-semiflows. In G. Rozenberg, editor, *Advances in Petri Nets 1990*, volume 483 of *LNCS*, pages 79–112. Springer-Verlag, 1991.
4. W.H. Elliot and D.C. Elliot. *Biochemistry and Molecular Biology, 2nd edition.* Oxford University Press, 2002.
5. S. Gilmore and J. Hillston. The PEPA Workbench: A Tool to Support a Process Algebra-based Approach to Performance Modelling. In *Proceedings of the Seventh International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, number 794 in Lecture Notes in Computer Science, pages 353–368, Vienna, May 1994. Springer-Verlag.
6. M. Heiner and I. Koch. Petri net based model validation in systems biology. In *25th International Conference on Application and Theory of Petri Nets, Bologna, Italy*, 2004.
7. J. Hillston. *A Compositional Approach to Performance Modelling.* Cambridge University Press, 1996.
8. T.H. Matheiss and D.S. Rubin. A survey and comparison of methods for finding all the vertices of conbex polyhedral sets. *Mathematics of Operational Research*, 5(2):167–185, 1980.
9. C. Priami, A. Regev, W. Silverman, and E. Shapiro. Application of a stochastic name passing calculus to representation and simulation of molecular processes. *Information Processing Letters*, 80:25–31, 2001.
10. A. Regev, W. Silverman, and E. Shapiro. Representation and simulation of biochemical processes using the pi-calculus process algebra. In *Proceedings of the Pacific Symposium of Biocomputing (PSB2001)*, pages 459–470, 2001.

A PEPA

This appendix provides a brief introduction to PEPA in order to make the paper self-contained. It can safely be skipped by anyone who already knows the PEPA language. For a full explanation which complements the brief description presented here the reader is referred to [7].

Prefix: The basic mechanism for describing the behaviour of a system with a PEPA model is to give a component a designated first action using the prefix combinator, denoted by a full stop. For example, $(\alpha, r).S$ carries out activity (α, r), which has action type α and an exponentially distributed duration with parameter r, and it subsequently behaves as S.

Choice: The component $P + Q$ represents a system which may behave either as P or as Q. The activities of both P and Q are enabled. The first activity to

complete distinguishes one of them: the other is discarded. The system will behave as the derivative resulting from the evolution of the chosen component.

Constant: It is convenient to be able to assign names to patterns of behaviour associated with components. Constants are components whose meaning is given by a defining equation. The notation for this is $X \stackrel{def}{=} E$. The name X is in scope in the expression on the right hand side meaning that, for example, $X \stackrel{def}{=} (\alpha, r).X$ performs α at rate r forever.

Hiding: The possibility to abstract away some aspects of a component's behaviour is provided by the hiding operator, denoted P/L. Here, the set L identifies those activities which are to be considered internal or private to the component and which will appear as the unknown type τ.

Cooperation: We write $P \underset{L}{\bowtie} Q$ to denote cooperation between P and Q over L. The set which is used as the subscript to the cooperation symbol, the *cooperation set* L, determines those activities on which the *cooperands* are forced to synchronise. For action types not in L, the components proceed independently and concurrently with their enabled activities. We write $P \parallel Q$ as an abbreviation for $P \underset{L}{\bowtie} Q$ when L is empty.

However, if a component enables an activity whose action type is in the cooperation set it will not be able to proceed with that activity until the other component also enables an activity of that type. The two components then proceed together to complete the *shared activity*. The rate of the shared activity may be altered to reflect the work carried out by both components to complete the activity (for details see [7]).

In some cases, when an activity is known to be carried out in cooperation with another component, a component may be *passive* with respect to that activity. This means that the rate of the activity is left unspecified (denoted $\top$) and is determined upon cooperation, by the rate of the activity in the other component. All passive actions must be synchronised in the final model.