# From Petri Nets to Differential Equations
# - an Integrative Approach
# for Biochemical Network Analysis

David Gilbert[1] and Monika Heiner[2]

[1] Bioinformatics Research Centre, University of Glasgow
Glasgow G12 8QQ, Scotland, UK
`drg@brc.dcs.gla.ac.uk`

[2] Department of Computer Science, Brandenburg University of
Technology
Postbox 10 13 44, 03013 Cottbus, Germany
`monika.heiner@informatik.tu-cottbus.de`

UNIVERSITY
*of*
GLASGOW

# From Petri Nets to Differential Equations
# - an Integrative Approach
# for Biochemical Network Analysis

David Gilbert[1] and Monika Heiner[2]

[1] Bioinformatics Research Centre, University of Glasgow
Glasgow G12 8QQ, Scotland, UK
`drg@brc.dcs.gla.ac.uk`
[2] Department of Computer Science, Brandenburg University of Technology
Postbox 10 13 44, 03013 Cottbus, Germany
`monika.heiner@informatik.tu-cottbus.de`

**Abstract.** We report on the results of an investigation into the integration of Petri nets and ordinary differential equations (ODEs) for the modelling and analysis of biochemical networks, and the application of our approach to the model of the influence of the Raf Kinase Inhibitor Protein (RKIP) on the Extracellular signal Regulated Kinase (ERK) signalling pathway. We show that analysis based on a discrete Petri net model of the system can be used to derive the sets of initial concentrations required by the corresponding continuous ordinary differential equation model, and no other initial concentrations produce meaningful steady states. Altogether, this paper represents a tutorial in step-wise modelling and analysis of larger models as well as in structured design of ODEs.

## 1 Motivation

Classical, i.e. time-less discrete Petri nets combine an intuitive modelling style with well-founded analysis techniques. It is for this reason that they are widely used in various application areas, where they have been proven to be useful for a qualitative verification of technical as well as "natural" systems, i.e. biochemical networks like metabolic networks, signal transduction networks, or gene regulatory networks.

However, any real system behaviour happens in time. Thus the next step following on from a qualitative analysis typically consists in quantitative analyses taking into account timing information. In the case of biochemical systems, all atomic actions take place continuously. Moreover, the rates of all the atomic actions typically depend on the continuous concentration of the involved substances. Hence systems of ordinary differential equations (ODEs) appear to be a natural choice for quantitative modelling of biochemical networks.

In this paper we bridge the gap between these two worlds, i.e. the (time-less) discrete and the (timed) continuous one, and demonstrate by means of

one of the standard examples used in the systems biology community – the core model of the influence of the Raf-1 Kinase Inhibitor Protein (RKIP) on the ERK signalling pathway – how both sides can play together by providing different, but complementary viewpoints on the same subject.

This paper can be considered as a tutorial in the step-wise modelling and analysis of larger models as well as in the structured design of ODEs. The discrete model is introduced as a supplementary intermediate step, at least from the viewpoint of the biochemist accustomed to ODE modelling only, and serves mainly for model validation since this cannot be performed on the continuous level. Having successfully validated the discrete model, the continuous model is derived from the discrete one by assigning rate equations to all of the atomic actions in the network. Thus the continuous model preserves the structure of the discrete one, and the continuous Petri net is nothing else than a structured description of ODEs.

The approach is presented by a small example, which is however sophisticated enough to highlight the main ideas — it is common sense to practice new techniques on small examples at first, before attempting larger ones, where the outcome to be expected tends to be less well-defined.

Moreover we demonstrate how the discrete model can be used to drive the continuous model by automatically generating sets of biochemically plausible values for the initial concentrations of protein species.

This paper is organized as follows. The next section provides an overview on the biochemical context on hand and introduces the running example. Afterwards, we demonstrate the step-wise modelling and analysis, where section 3 deals with the contributions by the discrete viewpoint, while section 4 is devoted to the continuous viewpoint. Having presented our own approach, we discuss some related work in section 5. We conclude with a summary and outlook on intended further research directions.

## 2   Biochemical Context

There are many networks of interacting components known to exist as part of the machinery of living organisms. Biochemical networks can be metabolic, regulatory or signal transduction networks. The role of metabolic networks is to synthesize essential biochemical compounds from basic components, or to degrade compounds. Regulatory networks are used to control the ways in which genes are expressed as RNAs or proteins, whereas signal transduction networks transmit biochemical signals between or within cells.

The two terms "pathway" and "network" tend to be used interchangeably in the literature, with "pathway" being (implicitly) taken to be a part of a more general network. In this paper we follow the generally accepted use of the term "pathway" to refer to the core of a biochemical network, comprising a sequence of activities, for example a kinase cascade. Thus, for example, we will describe the ERK pathway as being embedded in a more general signal transduction

network, and that the ERK pathway is a member of a large family of MAP Kinase pathways.

In this paper we focus on signal transduction, which is the mechanism which enables a cell to sense changes in its environment and to make appropriate responses. The basis of this mechanism is the conversion of one kind of signal into another. Extracellular signaling molecules are detected at the cell membrane by being bound to specific trans-membrane receptors that face outwards from the membrane and trigger intracellular events, which may eventually effect transcriptional activities in the nucleus. The eventual outcome is an alteration in cellular activity including changes in the gene expression profiles of the responding cells. These events, and the molecules that they involve, are referred to as (intracellular) "signalling pathways"; they contribute to the control of processes such as proliferation, cell growth, movement, apoptosis, and inter-cellular communication. Many signal transduction processes are "signalling cascades" which comprise a series of enzymatic reactions in which the product of one reaction acts as the catalytic enzyme for the next. The effect can be amplification of the original signal, although in some cases, for example the MAP kinase cascade, the signal gain is modest [1], suggesting that a main purpose is regulation [2] which may be achieved by positive and negative feedback loops.

The main factor which distinguishes signal transduction pathways from metabolic networks is that in the former the product of an enzymatic reaction becomes the enzyme for the next step in the pathway, whereas in the latter the product of one reaction becomes the substrate for the next, see Fig 1. In general, it is *dynamic* behaviour which is of interest in a signalling pathway, as opposed to the steady state in a metabolic network. In gene regulatory networks, on the other hand, the inputs are proteins such as transcription factors (produced from signal transduction or metabolic activity), which then influence the expression of genes – enzymatic activity plays no direct role here. However, the products of gene regulatory networks can play a part in the transcription of other proteins, or can act as enzymes in signalling or metabolic pathways.
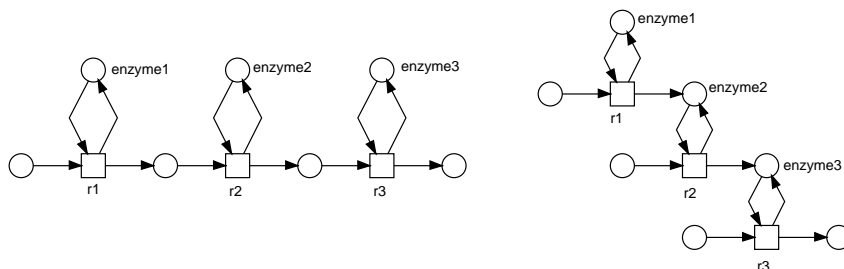


**Fig. 1.** The essential structural difference between metabolic networks (left) and signal transduction networks (right).

The ERK pathway (also called Ras/Raf, or Raf-1/MEK/ERK pathway) is a ubiquitous pathway that conveys cell division and differentiation signals from the cell membrane to the nucleus. Ras is activated by an external stimulus, via one of many growth factor receptors; it then binds to and activates Raf-1 to become Raf-1*, or activated Raf, which in turn activates MAPK/ERK Kinase (MEK) which in turn activates Extracellular signal Regulated Kinase (ERK). This cascade (Raf-1 → Raf-1* → MEK → ERK) of protein interaction controls cell differentiation, the effect being dependent upon the activity of ERK. An important area of experimental scientific investigation is the role that the Raf-1 Kinase Inhibitor Protein (RKIP) plays in the behaviour of this pathway. The hypothesis is that RKIP can inhibit activation of Raf-1 by binding to it, disrupting the interaction between Raf-1 and MEK, thus playing a part in regulating the activity of the ERK pathway.

## 3   The Discrete Approach

In this section we apply place/transition Petri nets to model the pathway of interest, and interpret them in the standard way. The reader is assumed to be familiar with the basic terms; for an introduction see e.g. [3], [4]. However, in order to make the paper readable for the biochemist, we will revise the most important key terms which are held to be crucial in the biochemistry context.

The software tools which have been used in this section are: for modelling – Snoopy [5], and for analysis – the Integrated Net Analyser (INA) [6], and the Model Checking Kit [7].

### 3.1   Qualitative Modelling

We apply the well-established modelling principles to represent biochemical networks by (various versions of) Petri nets, outlined e.g. in [8], [9].

**Definition 1 (place/transition Petri net).**
  A place/transition Petri net *is a quadruple* $\mathcal{N} = \langle P, T, f, m_0 \rangle$, *where*

 − $P$ *and* $T$ *are finite, non empty, and disjoint sets.* $P$ *is the set of* places *(in the figures represented by cycles).* $T$ *is the set of* transitions *(in the figures represented by rectangles).*
 − $f : (P \times T) \cup (T \times P) \rightarrow \mathbb{N}_0$ *defines the set of directed* arcs, *weighted by non-negative integer values.*
 − $m_0 : P \rightarrow \mathbb{N}_0$ *gives the* initial marking.

Moreover, we introduce the following notations: The preset of a node $x \in P \cup T$ is defined as $\bullet x := \{y \in P \cup T | f(y, x) \neq 0\}$, and its postset as $x^{\bullet} := \{y \in P \cup T | f(x, y) \neq 0\}$. A node $x \in P \cup T$ is called a boundary node, iff $\bullet x = \emptyset$ or $x^{\bullet} = \emptyset$.

According to definition 1 we create a place/transition Petri nets, see Figure 2, of the RKIP pathway, given in [10] in the style of a bichromatic graph. Circles
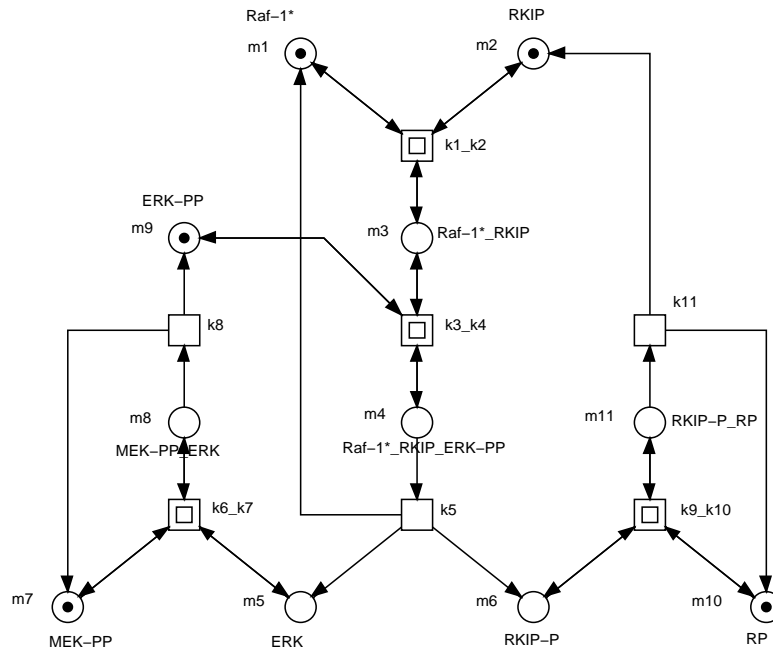
**Fig. 2.** The hierarchical Petri net for the core model of the RKIP pathway. Places stand for the states of the concentration of a protein; complexes are indicated by an underscore "_" between the protein names. Each macro transition (drawn by two centric squares) stands for a reversible reaction, modelled by two complementary transitions, which are hidden on the next lower net level. The layout follows the suggestions by the graphical notation used in [10]. At the bottom the two-lines result vector as produced by the Integrated Net Analyser [6] is given. The properties of interest in the given context of biochemical network analysis are explained in the text. The initial marking, not given in [10], is constructed systematically using standard Petri net analysis techniques.

(places) stand for the states of a protein or protein complex and are labelled with the corresponding name; complexes are indicated by an underscore "_" between the protein names. For example, Raf-1* and RKIP are proteins, and Raf-1*_RKIP is a protein complex formed from Raf-1* and RKIP. A suffix -P or -PP denotes a single or double phosphorylated protein, for example RKIP-P and ERK-PP. In the pathway under consideration there are 11 proteins or complexes; a discrete concentration $m1$, $m2$ etc. is associated with each protein or complex. In the case of the qualitative model, these concentrations can be thought of as being 'high' or 'low' (present or absent).

Rectangles (transitions) stand for reactions, with reversible reactions being indicated by a rectangle containing an inner rectangle (hierarchical transitions). In this pathway, reactions comprise protein complexation and decomplexation events, often accompanied by phosphorylation or dephosphorylation. Reactions can be one-way, indicated by single-headed arrows, or reversible, indicated by double-headed arrows on the top level of the net. For example, Raf-1* and RKIP combine in a forwards reaction to form Raf-1*_RKIP which can disassociate in a backwards reaction into Raf-1* and RKIP, or combine with ERK-PP to form the complex Raf1*_RKIP_ERK-PP. In this qualitative model, $k1, k2, \ldots$ stand for reaction labels.

### 3.2   Qualitative Analysis

The Petri net enjoys all the pleasant general properties a Petri net insider could dream of: boundedness, liveness, reversibility, which are three orthogonal basic behavioural net properties [4]. The decision about the first two properties can be made for our example in a static way, while the last property requires dynamic analysis techniques. The essential steps of the systematic analysis procedure for the example are given in more detail as follows.

**(1) Structural properties.** The following three structural properties reflect the modelling approach and can be read as preliminary consistency checks.

The net is *ordinary*, i.e. all arc weights equal to 1. This includes homogeneity, i.e. the outgoing arcs of each place have the same multiplicity, which is a necessary prerequisite for the Deadlock Trap Property (DTP), see step (4) below.

The net is *pure*, i.e. there are no side-conditions. So, the net structure is fully represented by the incidence matrix, which is used for the calculation of the P- and T-invariants, see next step.

The net is *strongly connected*, which involves the absence of boundary nodes. So, the net is self-contained, i.e. closed system. Therefore, in order to make the net live, we have to construct an initial marking, compare step (3).

Moreover, the net belongs to the structural class "extended simple". Hence, we know that the net has the ability to be live independent of time, i.e. if it is live, then it remains live under any duration timing [11].

**(2) Static decision of marking-independent behavioural properties.**
Model validation should include a check of all invariants for their biological
plausibility. The following definition recalls the essential basic terms.

**Definition 2 (P-invariants, T-invariants).**
- *The* incidence matrix *of* $\mathcal{N}$ *is a matrix* $\mathcal{C} : P \times T \to \mathbb{Z}$, *indexed by* $P$ *and* $T$, *such that* $\mathcal{C}(p, t) = f(t, p) - f(p, t)$.
- *A* place vector (transition vector) *is a vector* $x : P \to \mathbb{Z}$, *indexed by* $P$ *(* $y : T \to \mathbb{Z}$, *indexed by* $T$).
- *A place vector (transition vector) is called a* P-invariant (T-invariant), *iff it is a non-trivial non-negative integer solution of the linear equation system* $x \cdot \mathcal{C} = 0$ *(* $\mathcal{C} \cdot y = 0$).
- *The nodes corresponding to an invariant's non-zero entries are called the* support *of this invariant* $x$, *written as* $supp(x)$.
- *An invariant* $x$ *is called* minimal, *iff it does not contain any other invariant* $z$, *i.e.* $\neg \exists$ *invariant* $z : supp(z) \subset supp(x)$, *and the greatest common divisor of all entries of* $x$ *is 1.*
- *A net is* covered by P-invariants, *shortly CPI, (*covered by T-invariants, *shortly CTI), iff every place (transition) participates in a P-invariant (T-invariant).*

A P-invariant stands for a set of places, over which the weighted sum of tokens
is constant and independent of any firing, i.e. for any markings $m_1$, $m_2$, which are
reachable by the firing of transitions, it holds that $y \cdot m_1 = y \cdot m_2$. So, P-invariants
represent token-preserving sets of places. In the context of metabolic networks, P-
invariants reflect substrate conservations, while in signal transduction networks
P-invariants often correspond to the several states of a given specie (protein or
protein complex). A place belonging to a P-invariant is obviously bounded.

In the net under consideration there are five minimal P-invariants covering
the net (CPI), consequently the net is bounded. All the P-invariants $x_i$ contain
only entries of 0 and 1, which allows a short-hand specification by just giving
the names of the places involved.

$x_1 = (\text{Raf-1*}, \text{Raf-1*\_RKIP}, \text{Raf-1*\_RKIP\_ERK-PP}),$
$x_2 = (\text{MEK-PP}, \text{MEK-PP\_ERK}),$
$x_3 = (\text{RP}, \text{RKIP-P\_RP}),$
$x_4 = (\text{ERK}, \text{ERK-PP}, \text{MEK-PP\_ERK}, \text{Raf-1*\_RKIP\_ERK-PP}),$
$x_5 = (\text{RKIP}, \text{Raf-1*\_RKIP}, \text{Raf-1*\_RKIP\_ERK-PP}, \text{RKIP-P\_RP}, \text{RKIP-P}).$

Each P-invariant $x_i$ stands obviously for a reasonable conservation rule. The
first name given indicates the specie preserved within each P-invariant. Due to
the chosen naming convention, this particular name also appears in all the other
place names of the same P-invariant.

A T-invariant has two interpretations in the given biochemical context. (1)
The entries of a T-invariant represent a multiset of transitions which by their fir-
ing reproduce a given marking in a partially manner, i.e. basically occurring one
after the other. This partial order sequence of the T-invariant's transition may

contribute to a deeper understanding of the net behaviour. (2) The entries of a T-invariant may also be read as the relative transition firing rates of transitions, all of them occurring permanently and concurrently. This activity level corresponds to the steady state behaviour. Independently of this interpretation, the net representation of minimal T-invariants (the T-invariant's transitions plus their pre- and post-places and all arcs in between) characterize minimal self-contained subnetworks with an enclosed biological meaning.

The net under consideration is covered by T-Invariants (CTI), which is a necessary condition for bounded nets to be live. Besides the expected four trivial T-invariants for the four reversible reactions, there is only one non-trivial minimal T-invariant $y = (k1, k3, k5, k6, k8, k9, k11)$. The net representation of this T-invariant describes the essential partial order behaviour of the modelled system, given in text style: (k1; k3; k5; (k6; k8), (k9; k11)), where ";" stands for "sequentiality" and "," for "concurrency". The automatic identification of non-trivial minimal T-invariants is in general useful as a method to highlight important parts of a network, and hence aid its comprehension by biochemists, especially when the entire network is too complex to easily comprehend.

All the properties above relate only to the structure, i.e. they are valid independently of the initial marking. In order to proceed we first need to generate an initial marking.

**(3) Initial marking construction.** For a systematic construction of the initial marking, the following criteria have to be taken into consideration.

- Each P-invariant needs at least one token.
- All (non-trivial) T-invariants should be realizable, meaning, the transitions, making up the T-invariant's multi-set can be fired in an appropriate order.
- Additionally, it is common sense to look for a minimal marking (as few tokens as possible), which guarantees the required behaviour.
- Within a P-invariant, choose the specie with the most *inactive* or the *monomeric* state.

Taking all these criteria together, the initial marking on hand is: Raf-1*, RKIP, ERK, MEK-PP, RP get each one token, while all remaining places are empty. With this initial marking, the net is covered by 1-P-invariants (exactly one token in each P-invariant), therefore the net is 1-bounded (also called safe). That is in perfect accordance with the understanding that in signal transduction networks a P-invariant comprises all the different states of one specie. Obviously, each specie can be only in one state at any time.

In the following, however, we will use an initial marking derived from the initial concentrations used by Cho et al. [10] as part of their method to estimate rate parameters required for their ODE model of the RKIP pathway (see Table 1 in Section 4.2). This marking is represented in Figure 2. We use their initial marking because we focus the Cho et al. model throughout this paper for illustrative purposes. We will later in this paper demonstrate that the Cho et al. initial marking is equivalent the initial marking which we have constructed, and that in fact both markings are members of a larger equivalence class of markings.
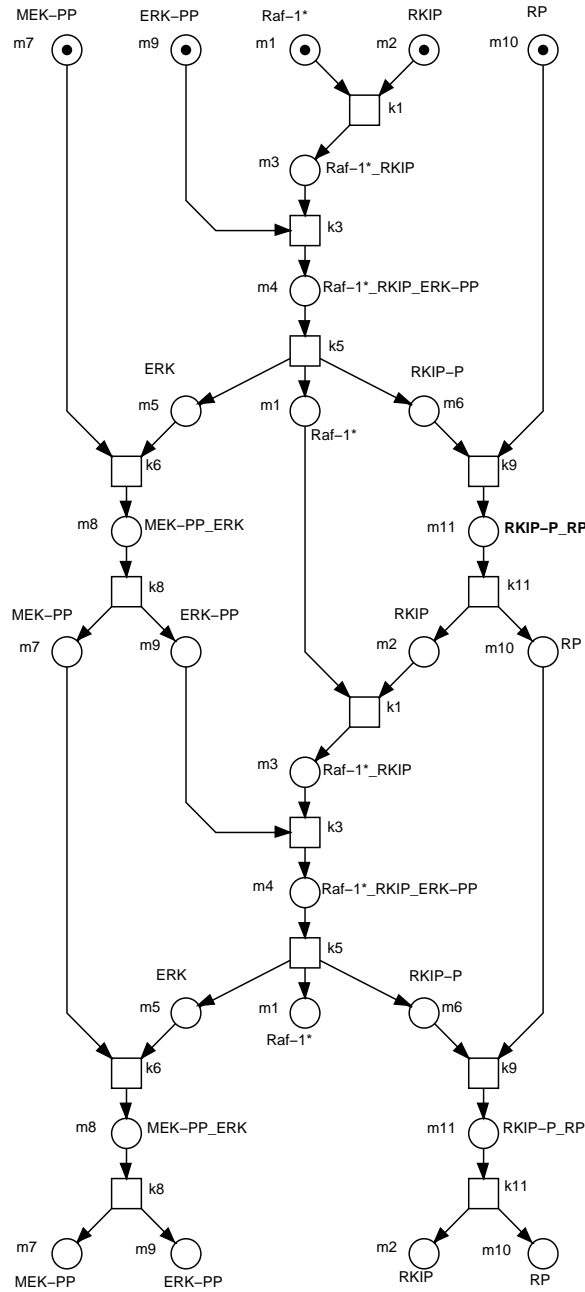
**Fig. 3.** The beginning of the infinite partial order run of the non-trivial minimal T-invariant of the place/transition Petri net given in Figure 2. Here, transitions represent events, labelled by the name of the reaction taking place, while places stand for binary conditions, labelled by the name of the species, set or reset by the event, respectively.

With the chosen marking we can check the non-trivial minimal T-invariant (see step (2)) for realizability, which then involves the realizability of the trivial T-invariants. We obtain an infinite run, the beginning of which is given as labelled condition/event net in Figure 3, and characterize this in a short-hand notation by the following set of partially ordered words out of the alphabet of all transition labels: (k1; k3; k5;[((k9; k11; k1), (k6; k8)); k3; k5]*). This partial order behaviour gives further insight into the dynamic behaviour of the network which may not be apparent from the standard net representation. In this example, the partial order representation illustrates the central role of Raf-1* and the fact that certain reactions take place sequentially, others can take place independently and that some reactions are constrained by the need for the presence of two or more precursors produced from prior reactions; e.g. the reaction labelled by k1 requires the participation of both proteins associated with concentration labels m1 and m2, which are produced by reactions k5 and k11 respectively.

Having established and justified our initial marking we proceed to the next steps of the analysis.

**(4) Static decision of marking-dependent behavioural properties.** The net belongs to the structural class "extended simple" and the Deadlock-Trap Property (DTP) holds (any structural deadlock contains a marked trap), therefore the net is live, see e.g. [4], [6]. However, most biochemical networks (as well as non-trivial technical networks) do not fulfill the DTP. Therefore we omit here the formal definition as well as a more detailed discussion to save space.

**(5) Dynamic decision of behavioural properties.** In order to decide reversibility we have to calculate the reachability graph. The nodes of a reachability graph represent all possible states (markings) of the net. The arcs in between are labelled by single transitions, the firing of which causes the related state change. Altogether, the reachability graph gives us a finite automaton representation of all possible single step firing sequences. Consequently, concurrent behaviour is described by enumerating all interleaving firing sequences.

Because we already know that the net is bounded, we also know that the reachability graph has to be finite. Here, the reachability graph has 13 states (out of $2048 = 2^{11}$ theoretically possible ones), forming one strongly connected component, see Figure 4. Therefore, the Petri net is reversible, i.e. the initial system state is always reachable again, or in other words - the system has the capability of self-reinitialization. Further, the liveness of the net has already been decided structurally, so we know that each transition (reaction) appears at least once in this strongly connected component.

Moreover, from the viewpoint of the discrete model, all these 13 states are equivalent, i.e. any of those 13 states could be taken as initial state resulting in exactly the same total (discrete) system behaviour. That is in perfect accordance with the observations gained during quantitative analyses, see Section 4.2.

For reasons of completeness we explored all other possible sensible initial states. Following our understanding that P-invariants in signal transduction net-

works reflect different states of a given specie, the net should be covered by 1-P-invariants. Therefore we had to consider only those initial markings which do not contradict this assumption. None of these potential initial markings results in a net whose behaviour is reversible and live, and only a few of them produce a terminal strongly connected component in the reachability graph (meaning that at least this part, consisting of mutually reversible reactions, is live).

This concludes the analysis of *general* behavioural net properties, i.e. of properties we can speak about in syntactic terms only, without any semantic knowledge. The next step consists in a closer look at *special* behavioural net properties, reflecting the expected special functionality of the network.
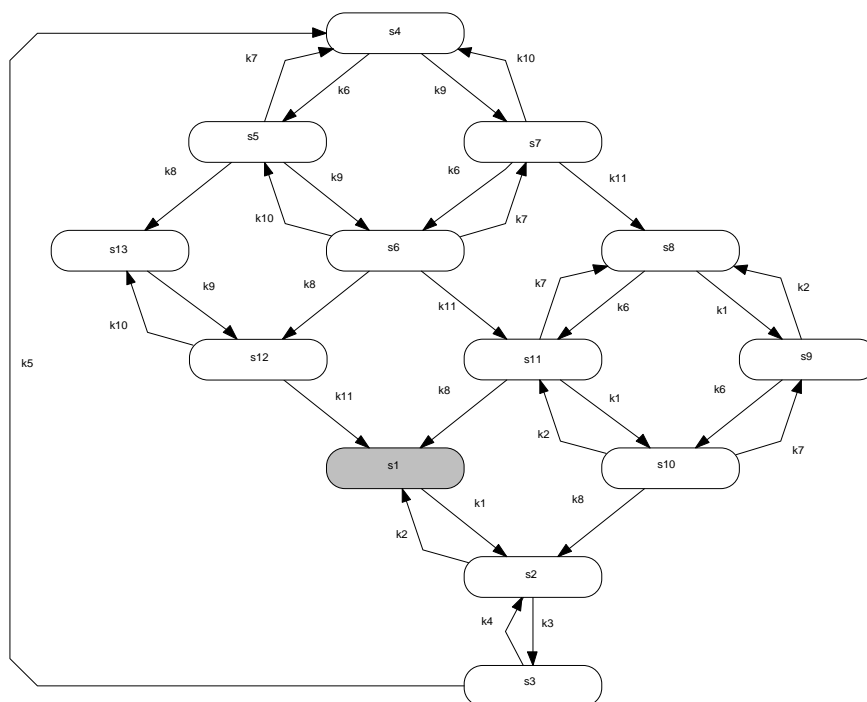


**Fig. 4.** The reachability graph of the place/transition Petri net given in Figure 2. Nodes represent states (markings), while the arcs are labelled with the transition responsible for a given change of states. For a list of the complete state descriptions see table 2. The initial state is highlighted in gray. In the reachability graph, the two concurrent transition sequences (k6; k8) and (k9; k11; k1), compare the partial order run of Figure 3, are represented by all possible totally ordered (interleaving) sequences, blowing up the state space.

**(6) Model checking of special behavioural properties.** Special properties are best expressed using temporal logics, e.g. Computational Tree Logic (CTL), which is a branching time logic with interleaving semantics. For an introduction into the specification of biologically relevant properties of biochemical networks using CTL see [12].

Because we are in the fortunate position of having a bounded model, these temporal-logic formulae can be checked using standard model checking techniques. Furthermore, the model under consideration is 1-bounded. Therefore, we can rely on a particularly rich choice of model checkers to solve this task [7]. In the case of our rather simple example the variety of model checkers is not important, and the properties could even be checked manually. However, the state space of more complex networks exceeds typically several millions.

We instantiate some of the generic property patterns provided in [12] and get the following samples of meaningful statements for our running example, whose truth can be determined via model checking:

– **property 1:** There are reachable states where ERK is phosphorylated and RKIP is not phosphorylated.

$$\textbf{EF} \ [ \ (\text{ERK-PP} \vee \text{Raf-1*\_RKIP\_ERK-PP}) \ \wedge \ \text{RKIP} \ ]$$

– **property 2:** The phosphorylation of ERK (to ERK-PP) is independent of the phosphorylated state of RKIP.

$$\textbf{E} \ [ \ \neg(\text{RKIP-P} \ \vee \ \text{RKIP-P\_RP}) \ \textbf{U} \ \text{ERK-PP} \ ]$$

– **property 3:** A cyclic behaviour w.r.t. the presence/absence of RKIP is possible forever.

$$\textbf{AG} \ [ \ ( \ \text{RKIP} \rightarrow \textbf{EF} \ (\neg\text{RKIP}) \ ) \wedge (\neg\text{RKIP} \rightarrow \textbf{EF} \ (\text{RKIP}) \ ) \ ]$$

### 3.3    Summary

To summarize the preceding validation steps, the model has passed the following validation criteria.

**validation criterion 0**

– All expected structural properties hold.
– All expected general behavioural properties hold.

**validation criterion 1**

– CPI.
– No minimal P-invariant without biological interpretation.

**validation criterion 2**

- CTI.
- No minimal T-invariant without biological interpretation.
- No known biological behaviour without corresponding, not necessarily minimal T-invariant.

**validation criterion 3**

- All expected special behavioural properties, expressed as temporal-logic formulae, hold.

It is worth noting that not all of the validation criteria outlined above are always feasible. E.g. it only makes sense to ask for CPI as well as CTI for self-contained (closed) systems, i.e. without boundary nodes. In the case of signal transduction networks it depends on the modelling style whether the essential system behaviour can be explained by the discussion of only T-invariants. Finally, validation criterion 3 relies on temporal logics as a flexible language to describe special properties. Thus it requires seasoned understanding of the network under investigation combined with the skill to correctly express the expected correct behaviour in temporal logics.

Therefore, the set of meaningful validation criteria has to be adjusted to the case study on hand, but it should become common practice to do some model validation and to make the criteria applied explicit.

Now we are ready for a more sophisticated quantitative analysis of our model.

## 4   The Continuous Approach

In this section we transform our validated time-less discrete model, given as place/transition Petri net, into a timed continuous one, specified as continuous Petri net. For an introduction into continuous Petri nets see e.g. [13], [14].

The software tools, which have been used in this section, are: an extended version of Snoopy [15], supporting modelling as well as analysis by some standard numerical integration algorithms. Additionally, we use Gepasi [16] and Matlab [17] for more detailed analyses.

### 4.1   Quantitative Modelling

In a continuous Petri net the marking of a place is no longer an integer, but a positive real number, called token value (which we are going to interpret as the concentration of a given specie). The instantaneous firing of a transition is carried out like a continuous flow, whereby the current firing rate depends generally on the current marking. To be precise we give the following definition.

**Definition 3 (Continuous Petri net ).** *A* continuous Petri net *is a quintuple* $\mathcal{CON} = \langle P, T, f, v, m_0 \rangle$, *where*

- *P and T are finite, non empty, and disjoint sets. P is the set of* continuous places *(in the figures represented by bold cycles).* T *is the set of* continuous transitions *(in the figures represented by bold rectangles).*
- $f : (P \times T) \cup (T \times P) \rightarrow \mathbb{R}_0^+$ *defines the set of directed* arcs, *weighted by non-negative real values.*
- $v : T \rightarrow H$ *assigns to each transition a* firing rate function, *whereby* $H := \bigcup_{t \in T} \left\{ h | h : \mathbb{R}^{|{}^\bullet t|} \rightarrow \mathbb{R} \right\}$ *is the set of all firing rate functions, and* $\mathrm{dom}\,(v\,(t)) = {}^\bullet t.$
- $m_0 : P \rightarrow \mathbb{R}_0^+$ *gives the* initial marking.

A continuous transition $t$ is enabled by $m$, iff $\forall p \in {}^\bullet t : m(p) > 0$. Due to the influence of time, a continuous transition is forced to fire as soon as possible. The firing rate of an atomic (re-) action depends typically on the current concentrations of the substances involved, i.e. of the token values of the transition's preplaces. So we get marking-dependent, i.e. variable firing rates. Please note, a firing rate may also be negative, in which case the reaction takes place in the reverse direction. This feature is commonly used to model reversible reactions by just one transition, where positive firing rates correspond to the forward direction, and negative ones to the backward direction.

Altogether, the semantics of a continuous Petri net is defined by a system of ordinary differential equations (ODEs), where one equation describes the continuous changes over time on the token value of a given place by the continuous increase of its pretransitions' flow and the continuous decrease of its posttransitions' flow:

$$\frac{m\,(p)}{dt} = \sum_{t \in {}^\bullet p} f\,(t, p)\,v\,(t) - \sum_{t \in p^\bullet} f\,(p, t)\,v\,(t)\,.$$

Each equation corresponds basically to a line in the incidence matrix, whereby now the matrix elements consist of the rate functions multiplied by the arc weight, if any. Moreover, as soon as there are transitions with more than one preplace, we get a non-linear system, which calls for a numerical treatment of the system on hand.

With other words, the continuous Petri net becomes the structured description of the corresponding ODEs. Due to the explicit structure we expect to get descriptions which are less error prone compared to those ones created manually from the scratch. In fact, writing down a system of ODEs by designing continuous Petri nets instead of just using a text editor might be compared to high-level instead of assembler programming. In order to simulate the continuous Petri net, exactly the same algorithms are employed as for numerical differential equation solvers, see e.g. [15].
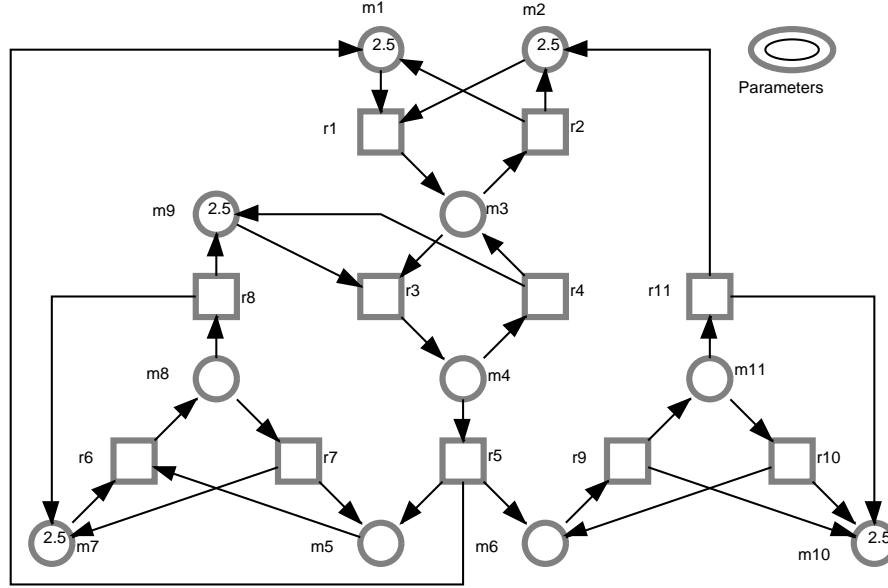
**Fig. 5.** The flattened continuous Petri net, corresponding to Figure 2. The system of ordinary differential equations defined by this net is given below in a structured notation (generated by our continuous Petri net tool). We use the $m_i$ as synonyms for the lengthy specie names. The initial concentrations reappear in Table 1.

$$\frac{dm_1}{dt} = r_2 + r_5 - r_1$$

$$\frac{dm_2}{dt} = r_2 + r_{11} - r_1$$

$$\frac{dm_3}{dt} = r_1 + r_4 - r_2 - r_3$$

$$\frac{dm_4}{dt} = r_3 - r_4 - r_5$$

$$\frac{dm_5}{dt} = r_5 + r_7 - r_6$$

$$\frac{dm_6}{dt} = r_5 + r_{10} - r_9$$

$$\frac{dm_7}{dt} = r_7 + r_8 - r_6$$

$$\frac{dm_8}{dt} = r_6 - r_7 - r_8$$

$$\frac{dm_9}{dt} = r_4 + r_8 - r_3$$

$$\frac{dm_{10}}{dt} = r_{10} + r_{11} - r_9$$

$$\frac{dm_{11}}{dt} = r_9 - r_{10} - r_{11}$$

$$r_1 = k_1 * m_1 * m_2$$

$$r_2 = k_2 * m_3$$

$$r_3 = k_3 * m_3 * m_9$$

$$r_4 = k_4 * m_4$$

$$r_5 = k_5 * m_4$$

$$r_6 = k_6 * m_5 * m_7$$

$$r_7 = k_7 * m_8$$

$$r_8 = k_8 * m_8$$

$$r_9 = k_9 * m_6 * m_{10}$$

$$r_{10} = k_{10} * m_{11}$$

$$r_{11} = k_{11} * m_{11}$$

$$k_1 = 0.53$$

$$k_2 = 0.0072$$

$$k_3 = 0.625$$

$$k_4 = 0.00245$$

$$k_5 = 0.0315$$

$$k_6 = 0.6$$

$$k_7 = 0.0075$$

$$k_8 = 0.071$$

$$k_9 = 0.92$$

$$k_{10} = 0.00122$$

$$k_{11} = 0.87$$

In the quantitative model of the biochemical pathway under consideration (see Figure 5), we associate a *rate* with each reaction, denoted by the rate constants $k1$, $k2$, etc.. These were reaction labels in the qualitative model, and are again placed in the figure next to the relevant reaction symbol (rectangle). In the case of reversible reactions, the rate constants are given by $kn$, $kn + 1$ denoting that $kn$ is the forward rate and $kn + 1$ the backward rate. E.g., the forward reaction rate for the combination of Raf-1* and RKIP is $k1$, and the disassociation rate for Raf-1_RKIP is $k2$.

## 4.2   Quantitative Analysis

In general, biochemists will wish to use ODE models of biochemical systems to explore in a general manner possible observable behaviours of the systems, for example the change in concentration of a component over time, or the steady-state properties of the system including oscillatory behaviour. Specifically in the case of signalling pathways the system components are proteins in both complexed and uncomplexed forms and in phosphorylated and unphosphorylated states. The kinds of experimental observations that can be made often result in very coarse data points — for example immuno-blotting will give quite inexact data on the *relative* concentrations of species at a few time-points, the data varying quite a lot between repeated experiments. In addition, experiments are often conducted *in-vivo* in cells, and immuno-blotting applied to the entire cell contents (after lysing, or breaking down the cell wall) — hence there is very little exactness possible in terms of concentrations since in reality these may vary through the cell, but local concentrations may not be measurable by this technique. Moreover it is often not possible to distinguish between the complexed and non-complexed form of proteins – thus for example the relative concentration of phosphorylated ERK (ERK-PP) will be given as a combination of ERK-PP alone ($m8$) plus the ERK-PP component of the Raf-1*_RKIP_ERK-PP complex ($m4$).

Given these inexactitudes, biochemists will want to know the answers to general questions, such as "Will the concentration of the phosphorylated form of protein X-PP rise for the first 10 minutes after a particular stimulus is given to the cell, and then remain constant?", and in the same experiment "Will the concentration of the unphosphorylated form of protein Y rise from the start of the experiment, peaking at 20 minutes at a concentration higher than that of X-PP, and then fall off during the remainder of the time, eventually becoming less than the concentration of X-PP?".

The ODE solvers which are normally used to interpret ODE models of biochemical networks rely on exact values of rate constants and initial concentrations in order for the computations to be performed. Thus the results produced by simulations of ODE models of networks may be over-exact with respect to the characteristics of the real data. For this reason, biochemists will often interpret the results of ODE-based simulations as indicators of the behaviour of the components of the network, rather than being concerned with the exact value of the concentration of a particular species at a particular point in time.

We have performed a quantitative analysis of the results of simulating the behaviour of the network using a system of ODEs. The aim of this analysis was to determine whether the 13 'good' initial states suggested by the qualitative Petri net analysis were indeed in some way equivalent (they all result in the same steady state), and that no other possible initial states can be used to give the same results.

The differential equation model of the pathway, taken from Cho et al. [10] and reproduced in Section 4 above, was coded in MatLab. Although these authors do not explicitly state the initial concentrations of the 11 species when computing the simulation of the network, we have deduced by inspection of Fig. 5 in their paper which presents their simulation results that they are as given in the $\mu M_{Cho}$ column of Table 1. For the purposes of our computations we have mapped any non-zero concentrations to 1, as in column $\mu M_{PN}$ of that table, hence our initial concentrations correspond to the marking in the Petri net in Figure 2.

**Table 1.** Initial concentrations

| Species | $\mu M_{Cho}$ | $\mu M_{PN}$ |
| --- | --- | --- |
| Raf1* | 2.5 | 1 |
| RKIP | 2.5 | 1 |
| Raf1_RKIP | 0 | 0 |
| RAF_RKIP_ERK | 0 | 0 |
| ERK | 0 | 0 |
| RKIP-P | 0 | 0 |
| MEK-PP | 2.5 | 1 |
| MEK-PP_ERK | 0 | 0 |
| ERK-PP | 2.5 | 1 |
| RP | 2.5 | 1 |
| RKIP-P_RP | 0 | 0 |

Since there are eleven species, there are $2^{11}$ i.e. 2048 possible initial states, including that given in the original paper. Of these, 13 were identified by the reachability graph analysis (Section 3.2) to form one strongly connected component, and thus to be be "good" initial states (see Table 2).

These are 'sensible' initial states from the point of view of biochemistry, in that in all these 13 cases, and in none of the other 2035 states, each protein specie is in a high initial concentration in only one of the following states: uncomplexed, complexed, unphosphorylated or phosphorylated. These conditions relate exactly to the 1-P-invariant interpretation given in our initial marking construction procedure in Section 3.2.

We then computed the final steady state of the set of species for each possible initial state, using the MatLab ODE solver ode45, which is based on an explicit Runge-Kutta formula, the Dormand-Prince pair [18], with 100 time steps.

We found that all of the 13 'good' initial states resulted in the same final state, within the bounds of computational error of the ODE solver. These results

**Table 2.** Initial 13 'good' state configurations

| Species | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 | S10 | S11 | S12 | S13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Raf-1* | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| RKIP | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| Raf-1*_RKIP | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| Raf-1*_RKIP_ERK-PP | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ERK | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| RKIP-P | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| MEK-PP | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 |
| MEK-PP_ERK | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| ERK-PP | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| RP | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 |
| RKIP-P_RP | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |

are summarized in Table 3 which reports the mean steady state concentration and standard deviation for each of the 11 species.

In the Appendix we reproduce two simulations of the model: State 1 (Figure 7) corresponding to the initial marking suggested by Cho et al [10] where the initial concentration of ERK-PP is high and ERK is low, and State 8 (Figure 8) corresponding to the initial marking suggested by our approach described above in Section 3.2 with ERK-PP low and ERK high. State 8 has been confirmed by an expert signal transduction researcher as the most sensible starting state [19]. The equivalence of the final states, compared with the difference in some intermediate states is clearly illustrated in these figures. For example, the concentration of Raf-1*_RKIP behaves overall in a similar manner in both State 1 and State 8, peaking before 10 minutes although the peak is greater when ERK is not phosphorylated at the start of the experiment. In Figure 9 we reproduce the computed behaviour of ERK-PP for all 13 good initial states, showing that despite differences in the concentrations at early time-points, the steady state concentration is the same in all 13 states.

We computed the Euclidean distances between the vector of mean values of the final steady states of the 13 states in the reachability graph and each of the final steady states for the states not identified by the reachability graph. These distances ranged from 0.7736 to 6.0889, and we summarize these results in Figure 6. None of the initial states which is not identified by the reachability analysis resulted in a final steady state which was near that of the set of 13 'good' states.

### 4.3  Summary

In summary, our results show that

1. All of the 13 states identified by the reachability graph of the discrete Petri net result in the same set of steady state values for the 11 species in the pathway.

**Table 3.** Mean values for steady states for the 13 'good' initial states

| Species | Mean steady state concentration | Standard Deviation |
|---|---|---|
| Raf-1* | 0.2133 | 0.1225 * 1.0e-04 |
| RKIP | 0.1727 | 0.0854 * 1.0e-04 |
| Raf-1*_RKIP | 0.2163 | 0.5546 * 1.0e-04 |
| Raf-1*_RKIP_ERK-PP | 0.5704 | 0.4346 * 1.0e-04 |
| ERK | 0.0332 | 0.0135 * 1.0e-04 |
| RKIP-P | 0.0200 | 0.0169 * 1.0e-04 |
| MEK-PP | 0.7469 | 0.6020 * 1.0e-04 |
| MEK-PP_ERK | 0.2531 | 0.6020 * 1.0e-04 |
| ERK-PP | 0.1433 | 0.1846 * 1.0e-04 |
| RP | 0.9793 | 0.0471 * 1.0e-04 |
| RKIP-P_RP | 0.0207 | 0.0473 * 1.0e-04 |
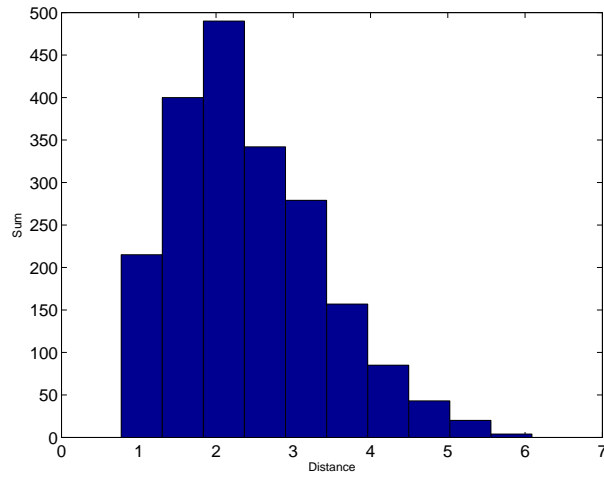


**Fig. 6.** Distribution of 'bad' steady states as Euclidean distances from the 'good' final steady state

2. None of the remaining 2035 possible initial states of the discrete Petri net results in a final steady state close to that generated by the 13 markings in the reachability graph.

3. The transient behaviour — the crucial point of interest in signal transduction networks — of the continuous model is sensible for the 13 states identified by our method.

## 5    Related Work

There are several research groups, applying various kinds of Petri nets to model and analyse biochemical networks. However, most of them are devoted to hybrid Petri nets, see e.g. [8], [20]; for a bibliography of related papers see [21]. Hybrid Petri nets comprise the discrete as well as the continuous case. Thus, they have to be treated by dedicated simulation techniques, instead by standard ODE solvers.

An approach combining qualitative and quantitative analysis techniques is proposed in [22]. In this paper time Petri nets are used to describe the steady state behaviour of a given biochemical network, whereby the time intervals are derived from the time-less model by help of the T-invariants, which are interpreted as firing count vectors. Interval-timed Petri nets provide a continuous time scale, but keep the discrete firing behaviour. Therefore they can still be treated in a discrete way, but they do not help in investigating continuous firing behaviour, e.g. in the transient state of a given network.

An approach describing the automatic derivation of ODEs from stochastic process algebra models of signalling pathways is presented in [23]. Consequently, the authors employ different analysis methods, which might be complementary to our ones, but they do not generate initial good markings.

Investigations on the relation between the properties of discrete and continuous Petri nets are fairly recent, as mentioned in [24]. In contrast to our approach, these authors focus on technical applications, which are inherently discrete. Then, they use the continuous model as a relaxation/approximation of the discrete one to get a better efficiency of analysis. In any case, they find themselves confronted with exactly the same questions we face: how do the properties of the one model relate to the properties in the other one. In particular the paper mentions some open questions of great interest, the solution of which coincide perfectly with our requirements.

## 6    Summary

We have created a discrete Petri net model of the influence of the Raf Kinase Inhibitor Protein (RKIP) on the Extracellular signal Regulated Kinase (ERK) signalling pathway, based on the ODE model presented by Cho et al [10]. We have then analysed the discrete model using a set of Petri net based tools and shown that the model enjoys several nice properties, among them boundedness, liveness, and reversibility. Moreover, the net is covered by P-invariants and T-invariants, all of them having sensible biological interpretation, and it fulfills

several special functional properties, which have been expressed in temporal logic. Reachability graph analysis identifies 13 strongly connected states out of 2048 theoretically possible ones, which permit self-reinitialization of the Petri net. From the viewpoint of the discrete model, all these 13 states are equivalent, and could be taken as an initial state resulting in exactly the same total (discrete) system behaviour.

We have then transformed the discrete Petri net into a continuous Petri net, defining ODEs. We have shown empirically that in the ODE model the 13 initial states, derived from the discrete model, result in the same steady state. This analysis was performed by numerically solving the system of ordinary differential equations. Moreover, none of the other 2035 possible states result in a steady state close to that derived using those identified by reachability analysis.

Altogether we advocate a two-step technology for the modelling and analysis of biochemical networks in a systematic manner: (1) qualitative, i.e. (timeless) discrete modelling and analysis, esp. for the beneficial effect of confidence-increasing model validation, and (2) quantitative, i.e. (timed) continuous modelling and analysis, esp. with the hope of the reliable prediction of behaviour. For both steps we favour the deployment of both discrete as well as continuous Petri nets, sharing the same net structures for a given case. The quantitative model is derived from the qualitative one only by the addition of the quantitative parameters. Hence both models are likely to share some behavioural properties. However, the meticulous rules of this approach are the subject of further ongoing investigations by the authors.

## Acknowledgements

## References

1. Schoeberl, B., Eichler-Jonsson, C., Gilles, E., Muller, G.: Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. Nature Biotechnology **20** (2002) 370–375
2. Kolch, W., Calder, M., Gilbert, D.: When kinases meet mathematics: the systems biology of MAPK signalling. FEBS Letters **579** (2005) 1891–5
3. Reisig, W.: Petri nets; An introduction. Springer (1982)
4. Murata, T.: Petri nets: Properties, analysis and applications. Proc.of the IEEE 77 **4** (1989) 541–580
5. Fieber, M.: Design und implementation of a generic and adaptive tool for graph manipulation, (in German). Master thesis, BTU Cottbus, Dep. of CS (2004)
6. Starke, P., Roch, S.: INA - The Intergrated Net Analyzer, Humboldt University Berlin, http://www.informatik.hu-berlin.de/∼starke/ina.html. (1999)

7. Schröter, C., Schwoon, S., Esparza, J.: The Model Checking Kit. In: Proc. ICATPN, LNCS 2697, Springer (2004) 463–472
8. Matsuno, H., Fujita, S., Doi, A., Nagasaki, M., Miyano, S.: Towards pathway modelling and simulation. In: Proc. 24th ICATPN, LNCS 2679. (2003) 3–22
9. Heiner, M., Koch, I.: Petri net based model validation in systems biology. In: Proc. 25th ICATPN 2004, LNCS 3099, Springer (2004) 216–237
10. Cho, K.H., Shin, S.Y., Kim, H.W., Wolkenhauer, O., McFerran, B., Kolch, W.: Mathematical modeling of the influence of RKIP on the ERK signaling pathway. Lecture Notes in Computer Science **2602** (2003) 127–141
11. Starke, P.H.: Some properties of timed nets under the earliest firing rule. Lecture Notes in Computer Science; Advances in Petri Nets **424** (1989) 418–432
12. Chabrier-Rivier, N., Chiaverini, M., Vincent Danos, F.F., Schächter, V.: Modeling and querying biomolecular interaction networks. Theoretical Computer Science **325** (2004) 25–44
13. David, R., Alla, H.: Discrete, Continuous, and Hybrid Petri Nets. Springer (2005)
14. Lonitz, K.: 'Hybrid Systems Modelling in Engineering and Life Sciences'. Master thesis, Universität Koblenz-Landau (2005)
15. Scheibler, D.: A software tool for design and simulation of continuous Petri nets, (in German). Master thesis, BTU Cottbus, Dep. of CS (2005, to appear)
16. Mendes, P.: GEPASI: A software package for modelling the dynamics, steady states and control of biochemical and other systems. Comput. Applic. Biosci. **9** (1993) 563–571
17. Shampine, L.F., Reichelt, M.W.: The MATLAB ODE Suite. SIAM Journal on Scientific Computing **18** (1997) 1–22
18. Dormand, J.R., Prince, P.J.: A family of embedded runge-kutta formulae. J. Comp. Appl. Math. **6** (1980) 1–22
19. Kolch, W. (personal communication)
20. Chen, M., Hofestädt, R.: A medical bioinformatics approach for metabolic disorders: Biomedical data prediction, modeling, and systematic analysis. J Biomedical Informatics **in press** (2005) 13 p.
21. Will, J., Heiner, M.: Petri nets in biology, chemistry, and medicine - bibliography. Technical Report 04/2002, BTU Cottbus, Computer Science (2002)
22. Popova-Zeugmann, L., Heiner, M., Koch, I.: Time petri nets for modelling and analysis of biochemical networks. Fundamenta Informaticae **67** (2005) 149162
23. Calder, M., Gilmore, S., Hillston, J.: Automatically deriving ODEs from process algebra models of signalling pathways. In: Proc. Computational Methods in Systems Biology (CSMB05), LFCS, University of Edinburgh (2005) 204–215
24. Silva, M., Recalde, L.: Continuization of timed Petri nets: From performance evaluation to observation and control. In: Proc. 26th ICATPN 2005, LNCS 3536, Springer (2005) 26–47

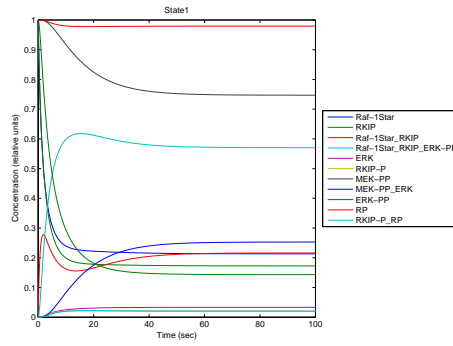## Appendix: Dynamic behaviour of the pathway

**Fig. 7.** Dynamic behaviour for state 1



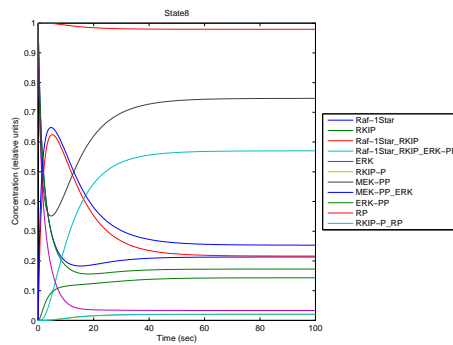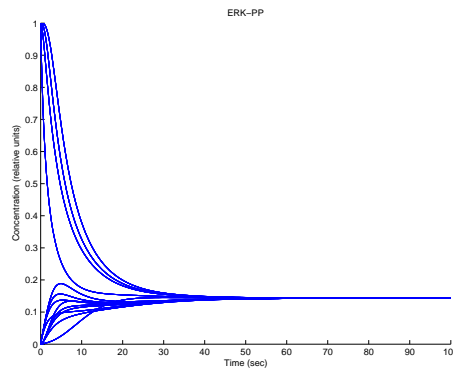**Fig. 8.** Dynamic behaviour for state 8



**Fig. 9.** Dynamic behaviour of ERK-PP for all 13 'good' states