A Unifying Framework for Modelling and Analysing Biochemical Pathways Using Petri Nets

David Gilbert[1], Monika Heiner[2], and Sebastian Lehrack[3]

[1] Bioinformatics Research Centre, University of Glasgow
Glasgow G12 8QQ, Scotland, UK
drg@brc.dcs.gla.ac.uk, on sabbatical leave from [3]
[2] INRIA Rocquencourt, Projet Contraintes
BP 105, 78153 Le Chesnay CEDEX - France
monika.heiner@inria.fr, on sabbatical leave from [3]
[3] Department of Computer Science, Brandenburg University of Technology
Postbox 10 13 44, 03013 Cottbus, Germany
sebastian.lehrack@informatik.tu-cottbus.de

Abstract. We give a description of a Petri net-based framework for modelling and analysing biochemical pathways, which unifies the qualitative, stochastic and continuous paradigms. Each perspective adds its contribution to the understanding of the system, thus the three approaches do not compete, but complement each other. We illustrate our approach by applying it to an extended model of the three stage cascade, which forms the core of the ERK signal transduction pathway. Consequently our focus is on transient behaviour analysis. We demonstrate how qualitative descriptions are abstractions over stochastic or continuous descriptions, and show that the stochastic and continuous models approximate each other. A key contribution of the paper consists in a precise definition of biochemically interpreted stochastic Petri nets. Although our framework is based on Petri nets, it can be applied more widely to other formalisms which are used to model and analyse biochemical networks.

1 Motivation

Biochemical systems are inherently governed by stochastic laws. However, due to the computational efforts required to analyse stochastic models, two abstractions are more popular: qualitative models, abstracting away from any time dependencies, and continuous models, commonly used to approximate stochastic behaviour by a deterministic one. The interrelationships between these three models are not always properly understood; for example, how the kinetics of a biochemical reaction, when described by a continuous model, is related to the stochastic nature of the underlying molecular mechanism.

In a previous paper [GH06] we developed an approach for modelling and analysing biochemical networks using discrete and continuous Petri nets. Our current work has taken this forward by considering stochastic Petri nets and

M. Calder and S. Gilmore (Eds.): CMSB 2007, LNBI 4695, pp. 200–216, 2007.

developing an overall framework to unify these three approaches, providing a family of related models with high analytical power.

A key contribution of this paper is the precise definition of biochemically interpreted stochastic Petri nets in a generic manner, and we demonstrate the benefit of their incorporation into the model development process. We show how the general definition can be tailored to very specific kinetic assumptions by appropriate adjustments of the general hazard function. Also we discuss the relation of the stochastic Petri net to its time-free, purely qualitative abstraction - the standard Petri net, as well as to its continuous approximation - the continuous Petri net (i.e., an ordinary differential equation system).

This paper is organised as follows. The following section provides an overview of the biochemical context and introduces our running example. Next we outline our framework, discussing the special contributions of the three individual analysis approaches with special emphasis on the transient behaviour analysis, and examining their interrelations. We then present the individual approaches and discuss mutually related properties in all three paradigms in the following order: we start off with the qualitative approach, which is conceptually the easiest, and does not rely on knowledge of kinetic information, but describes the network topology and presence of the species. The qualitative modelling and analysis basically adheres to the steps proposed in [GH06]. In addition, we show how to systematically derive and interpret the partial order run of the signal response behaviour. We then demonstrate how the validated qualitative model can be transformed into the stochastic representation by addition of stochastic firing rate information. Next, the continuous model is derived from the qualitative or stochastic model by considering only deterministic firing rates. Suitable sets of initial conditions for all three models are constructed by qualitative analysis. We conclude with a summary and outlook regarding further research directions.

2 Biochemical Context

We have chosen a model of the mitogen-activated protein kinase (MAPK) cascade published in [LBS00] as a running case study. This is the core of the ubiquitous ERK/MAPK pathway that can, for example, convey cell division and differentiation signals from the cell membrane to the nucleus. The model does not describe the receptor and the biochemical entities and actions immediately downstream from the receptor. Instead the description starts at the RasGTP complex which acts as a kinase to phosphorylate Raf, which phosphorylates MAPK/ERK Kinase (MEK), which in turn phosphorylates Extracellular signal Regulated Kinase (ERK). This cascade (RasGTP $\rightarrow$ Raf $\rightarrow$ MEK $\rightarrow$ ERK) of protein interactions controls cell differentiation, the effect being dependent upon the activity of ERK. We consider RasGTP as the input signal and ERKPP (activated ERK) as the output signal.

The bipartite graph in Figure 1 describes the typical modular structure for such a signalling cascade. Each layer corresponds to a distinct protein species. The protein Raf in the first layer is only singly phosphorylated. The proteins

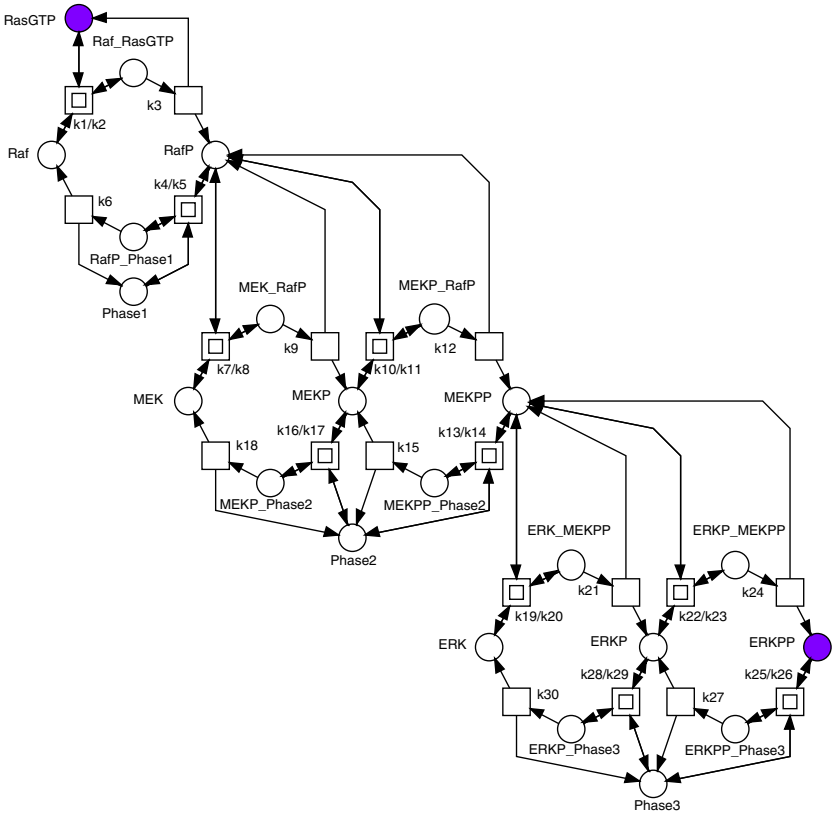

Fig. 1. The bipartite graph for the extended ERK pathway model. The graph has been derived by SBML import and automatic layout, manually improved, from the set of the ODEs in [LBS00]. Circles stand for species (proteins, protein complexes). Protein complexes are indicated by an underscore "_" between the constituent protein names. The suffixes P or PP indicate phosphorylated or doubly phosphorylated forms respectively. Squares stand for irreversible reactions, while two concentric squares specify reversible reactions. The species that are read as input/output signals are given in grey.

in the two other layers, MEK and ERK respectively, can be singly as well as doubly phosphorylated. In each layer, forward reactions are catalysed by kinases and reverse reactions by phosphatases (Phase1, Phase2, Phase3). The kinases in the MEK and ERK layers are the phosphorylated forms of the proteins in the previous layer, see also [CKS07].

3 Overview of the Framework

In the following we describe our overall framework, illustrated in Figure 2, that relates the three major ways of modelling and analysing biochemical networks described in this paper: qualitative, stochastic and continuous.

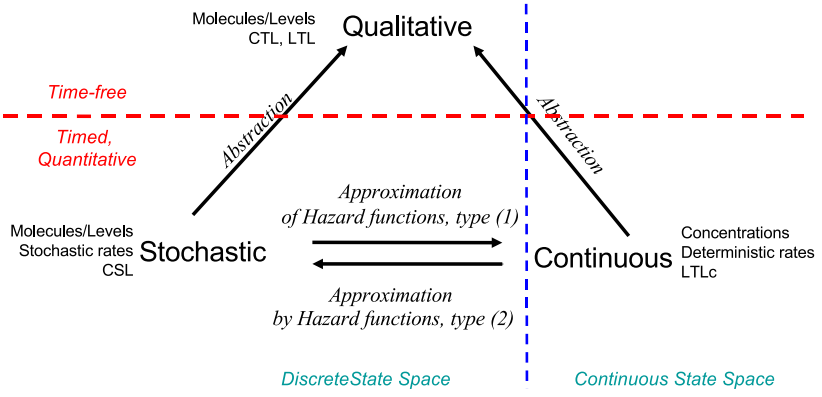

Fig. 2. Conceptual framework

The most abstract representation of a biochemical network is *qualitative* and is minimally described by its topology, usually as a bipartite directed graph with nodes representing biochemical entities or reactions, or in Petri net terminology *places* and *transitions* (see Figure 1). Arcs can be annotated with stoichiometric information, whereby the default stoichiometric value of 1 is usually omitted.

The qualitative description can be further enhanced by the abstract representation of discrete quantities of species, achieved in Petri nets by the use of tokens at places. These can represent the number of molecules, or the level of concentration, of a species, and a particular arrangement of tokens over a network is called a *marking*. The standard semantics for these qualitative Petri nets (QPN) does not associate a time with transitions or the sojourn of tokens at places, and thus these descriptions are time-free. The qualitative analysis considers however all possible behaviour of the system under any timing. The behaviour of such a net forms a discrete state space, which can be analysed in the bounded case, for example, by a branching time temporal logic, one instance of which is Computational Tree Logic (CTL), see [CGP01].

Timed information can be added to the qualitative description in two ways – stochastic and continuous. The stochastic Petri net (SPN) description preserves the discrete state description, but in addition associates a probabilistically distributed firing rate (waiting time) with each reaction. All reactions, which occur in the QPN, can still occur in the SPN, but their likelihood depends on the probability distribution of the associated firing rates (waiting times). Special behavioural properties can be expressed using e.g. Continuous Stochastic Logic (CSL), see [PNK06], a probabilistic counterpart of CTL. The QPN is an abstraction of the SPN, sharing the same state space and transition relation with the stochastic model, with the probabilistic information removed. All qualitative properties valid in the QPN are also valid in the SPN, and vice versa.

The continuous model replaces the discrete values of species with continuous values, and hence is not able to describe the behaviour of species at the level

of individual molecules, but only the overall behaviour via concentrations. We can regard the discrete description of concentration levels as abstracting over the continuous description of concentrations. Timed information is introduced by the association of a particular deterministic rate information with each transition, permitting the continuous model to be represented as a set of ordinary differential equations (ODEs). The concentration of a particular species in such a model will have the same value at each point of time for repeated experiments. The state space of such models is continuous and linear. It can be analysed by, for example, Linear Temporal Logic with constraints (LTLc) in the manner of [CCRFS06].

The stochastic and continuous models are mutually related by approximation. The stochastic description can be used as the basis for deriving a continuous Petri net (CPN) model by approximating rate information. Specifically, the probabilistically distributed reaction firing in the SPN is replaced by a particular average firing rate over the continuous token flow of the CPN. This is achieved by approximation over hazard functions of type (1), described in more detail in section 5.1. In turn, the stochastic model can be derived from the continuous model by approximation, reading the tokens as concentration levels, as introduced in [CVGO06]. Formally, this is achieved by a hazard function of type (2), see again section 5.1.

It is well-known that time assumptions generally impose constraints on behaviour. The qualitative and stochastic models consider all possible behaviours under any timing, whereas the continuous model is constrained by its inherent determinism to consider a subset. This may be too restrictive when modelling biochemical systems, which by their very nature exhibit variability in their behaviour.

In the following the reader is assumed to be familiar with the standard Petri net terminology as well as foundations of temporal logics, for an introduction see, e.g., [Mur89] and [CGP01].

4 The Qualitative Approach

4.1 Qualitative Modelling

We interpret the graph given in Figure 1 as a place/transition Petri net, and call the circles *places*, and the rectangles *transitions*. Reversible reactions have to be modelled explicitly by two opposite transitions in the basic Petri net notation. However in order to retain the elegant graph structure of Figure 1, we use macro transitions, each of which stands here for a reversible reaction. The entire (flattened) place/transition Petri net consists of 22 places and 30 transitions, where $k1, k2, \ldots$ stand for reaction (transition) labels.

We associate a discrete concentration with each of the 22 species. In the simplest way, these concentrations can be thought of as being "high" or "low" (above or below a certain threshold), resulting in a two-level model where each species can be read as a Boolean variable. More generally, we could apply a multi-level approach by differentiating between a finite number of discrete levels,

each standing for an equivalence class of possibly infinitely many concentrations. Then species can be read as integer variables.

4.2 Qualitative Analysis

Analysis of general behavioural properties. The Petri net enjoys the three orthogonal general properties of a discrete Petri net: boundedness, liveness and reversibility. The decision about the first two can be made for our running example in a static way, while the last property requires dynamic analysis techniques. The necessary steps of the systematic analysis procedure follow basically those given in [GH06]. We restrict ourselves here to the most essential points.

The net is *strongly connected* and thus self-contained, i.e. a closed system. In order to bring the net to life, we construct an initial marking using *P-invariants*. These are non-trivial non-negative integer solutions of the homogeneous linear equation system $x \cdot \mathcal{C} = 0$, where $\mathcal{C}$ stands for the incidence matrix of the net. There are seven minimal P-invariants covering the net, and consequently the net is bounded for any initial marking. All these P-invariants x_i contain only entries of 0 and 1, permitting a short-hand specification by just giving the names of the places involved.

x_1 = (**RasGTP,** Raf_RasGTP)
x_2 = (**Raf,** Raf_RasGTP, RafP, RafP_Phase1, MEK_RafP, MEKP_RafP)
x_3 = (**MEK,** MEK_RafP, MEKP_RafP, MEKP_Phase2, MEKPP_Phase2, ERK_MEKPP, ERKP_MEKPP, MEKPP, MEKP)
x_4 = (**ERK,** ERK_MEKPP, ERKP_MEKPP, ERKP, ERKPP_Phase3, ERKP_Phase3, ERK_PP)
x_5 = (**Phase1,** RafP_Phase1)
x_6 = (**Phase2,** MEKP_Phase2, MEKPP_Phase2)
x_7 = (**Phase3,** ERKP_Phase3, ERKPP_Phase3)

Each P-invariant stands for a reasonable conservation rule, the species preserved being given by the first name in the invariant. In signal transduction networks a P-invariant typically comprises all the different states of one species. In a Boolean approach, each species can be only in one state at any time, thus each P-invariant gets exactly one token. Within a P-invariant, the species with the most *inactive* (i.e. non-phosphorylated) or the *monomeric* (non-complexed) state is chosen. Following these criteria, the initial marking is: one token on each of Raf, RasGTP, MEK, ERK, Phase1, Phase2 and Phase3, while all remaining places are empty. With this marking, the net is covered by 1-P-invariants (exactly one token in each P-invariant), and is therefore 1-bounded.

Generalising this reasoning to a multi-level concept, we could assign n tokens to each place, representing the most inactive state, in order to indicate the highest concentration level for them in the initial state. The "abstract" mass conservation within each P-invariant would then be n tokens, which could be distributed fairly freely over the P-invariant's places during the behaviour of the model. This results in a dramatic increase of the state space, cf. the discussion in Section 5.2, while not improving the qualitative reasoning.

Model validation should include a check of all *T-invariants* for their biological plausibility. T-invariants are non-trivial non-negative integer solutions of the homogeneous linear equation system $\mathcal{C} \cdot y = 0$. The entries of a T-invariant can be read as the specification of a multiset of transitions, which reproduce a given marking by their firing. If there are non-trivial solutions, then there are infinitely many ones. Therefore, the plausibility check is usually restricted to the consideration of all minimal solutions. The net representations of minimal T-invariants (their transitions plus their pre- and post-places and all arcs in between) characterise minimal self-contained subnetworks with an identifiable biological meaning.

The net under consideration is covered by T-invariants, a necessary condition for bounded nets to be live. Besides the expected ten trivial T-invariants for the ten reversible reactions, there are five non-trivial, but obvious minimal T-invariants, each corresponding to one of the five phosphorylation / dephosphorylation cycles in the network structure:

y_1 = (k1, k3, k4, k6), y_2 = (k7, k9, k16, k18), y_3 = (k10, k12, k13, k15),
y_4 = (k19, k21, k28, k30), y_5 = (k22, k24, k25, k27).

The interesting net behaviour, demonstrating how input signals cause finally output signals, is contained in a non-negative linear combination of all five non-trivial T-invariants, $y_{1-5} = y_1 + y_2 + y_3 + y_4 + y_5$, which is called an I/O T-invariant in the following. The I/O T-invariant is systematically constructed by starting with the two minimal T-invariants, involving the input and output signal, which define disconnected subnetworks. Then we add minimal sets of minimal T-invariants to get a connected subnet. For our running example, the solution is unique, which is not generally the case.

We check the I/O T-invariant for feasibility in the constructed initial marking, which then involves the feasibility of all trivial T-invariants. We obtain an *infinite partial order run*, the beginning of which can be characterised in a shorthand notation by the following partially ordered word out of the alphabet of all transition labels (";" stands for "sequentiality", "‖" for "concurrency"):

(k1; k3; k7; k9; k10; k12;
((k4; k6) ‖ ((k19; k21; k22; k24); ((k13; k15; k16; k18) ‖ (k25; k27; k28; k30))))),

see [GHL07] for a graphical representation. This partial order run gives further insight into the dynamic behaviour of the network, which may not be apparent from the standard net representation, e.g. we are able to follow the (minimal) producing process of the proteins RafP, MEKP, MEKPP, ERKP and ERKPP, compare [GHL07], and we notice the clear independence of the dephosphorylation in all three levels.

The *reachability graph* of the net is finite because the net is bounded, and has in the Boolean token interpretation 118 states out of 2^{22} theoretically possible ones, forming one strongly connected component. 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. Moreover, from the

viewpoint of the discrete model, all of these 118 states are equivalent, and each could be taken as an initial state resulting in exactly the same total (discrete) system behaviour. This prediction will be confirmed by the observations gained during quantitative analyses, see Sections 5.2 and 6.2.

Model checking of special behavioural properties. Temporal logic is particularly helpful in expressing special behavioural properties of the expected transient behaviour, whose truth can be determined via model checking. We confine ourselves here to two CTL properties, checking the generalizability of the insights gained by the partial order run of the I/O T-invariant. In the following, places are interpreted as Boolean variables, in order to simplify notation.

Property Q1: The signal sequence predicted by the partial order run of the I/O T-invariant is the only possible one. In other words, starting at the initial state, it is necessary to pass through states RafP, MEKP, MEKPP and ERKP in order to reach ERKPP.

$$\neg\,[\,\mathbf{E}\,(\,\neg\,\mathrm{RafP}\;\;\mathbf{U}\;\mathrm{MEKP}\,)\;\vee\;\mathbf{E}\,(\,\neg\,\mathrm{MEKP}\;\;\mathbf{U}\;\mathrm{MEKPP}\,)\;\vee\;\mathbf{E}\,(\,\neg\,\mathrm{MEKPP}\;\;\mathbf{U}\;\mathrm{ERKP}\,)\;\vee\;\mathbf{E}\,(\,\neg\,\mathrm{ERKP}\;\;\mathbf{U}\;\mathrm{ERKPP}\,)\;\,]$$

Property Q2: Dephosphorylation takes place independently. E.g., the duration of the phosphorylated state of ERK is independent of the duration of the phosphorylated states of MEK and Raf.

$$(\,\mathbf{EF}\,[\,\mathrm{Raf}\wedge(\,\mathrm{ERKP}\vee\mathrm{ERKPP}\,)\,]\wedge\mathbf{EF}\,[\,\mathrm{RafP}\wedge(\,\mathrm{ERKP}\vee\mathrm{ERKPP}\,)\,]\wedge\mathbf{EF}\,[\,\mathrm{MEK}\wedge(\,\mathrm{ERKP}\vee\mathrm{ERKPP}\,)\,]\wedge\mathbf{EF}\,[\,(\,\mathrm{MEKP}\vee\mathrm{MEKPP}\,)\wedge(\,\mathrm{ERKP}\vee\mathrm{ERKPP}\,)\,]\,)$$

In subsequent sections we will use Q1 as a basis to illustrate how the stochastic and continuous approaches provide complementary views of the system behaviour.

5 The Stochastic Approach

5.1 Stochastic Modelling

As with a qualitative Petri net, a stochastic Petri net maintains a discrete number of tokens on its places. But contrary to the time-free case, a firing rate (waiting time) is associated with each transition t, which are random variables $X_t \in [0, \infty)$, defined by probability distributions. Therefore, all reaction times can theoretically still occur, but the likelihood depends on the probability distribution. Consequently, the system behaviour is described by the same discrete state space, and all the different execution runs of the underlying qualitative Petri net can still take place. This allows the use of the same powerful analysis techniques for stochastic Petri nets as already applied for qualitative Petri nets.

For a better understanding we describe the general procedure of a particular simulation run for a stochastic Petri net. Each transition gets its own local timer. When a particular transition becomes enabled, meaning that sufficient tokens arrive on its pre-places, then the local timer is set to an initial value,

which is computed at this time point by means of the corresponding probability distribution. In general, this value will be different for each simulation run. The local timer is then decremented at a constant speed, and the transition will fire when the timer reaches zero. If there is more than one enabled transition, a race for the next firing will take place.

Technically, various probability distributions can be chosen to determine the random values for the local timers. Biochemical systems are the prototype for exponentially distributed reactions. Thus, for our purposes, the firing rates of all transitions follow an exponential distribution, which can be described by a single parameter λ, and each transition needs only its particular, generally marking-dependent parameter λ to specify its local time behaviour.

Definition 1 (Stochastic Petri net, Syntax). *A* biochemically interpreted stochastic Petri net *is a quintuple* $\mathcal{SPN}_{Bio} = (P, T, f, v, m_0)$*, where*

- P *and* T *are finite, non empty, and disjoint sets.* P *is the set of* places*, and* T *is the set of* transitions*.*
- $f : ((P \times T) \cup (T \times P)) \to \mathbb{N}_0$ *defines the set of directed* arcs*, weighted by non-negative integer values.*
- $v : T \to H$ *is a function, which assigns a* stochastic hazard function h_t *to each transition* t*, whereby* $H := \bigcup_{t \in T} \left\{ h_t \,|\, h_t : \mathbb{N}_0^{|{}^\bullet t|} \to \mathbb{R}^+ \right\}$ *is the set of all stochastic hazard functions, and* $v(t) = h_t$ *for all transitions* $t \in T$*.*
- $m_0 : P \to \mathbb{N}_0$ *gives the* initial marking*.*

The stochastic hazard function h_t defines the marking-dependent transition rate $\lambda_t(m)$ for the transition t. The domain of h_t is restricted to the set of pre-places of t, i.e. ${}^\bullet t := \{p \in P | f\,(p,t) \neq 0\}$, to enforce a close relation between network structure and hazard functions. Therefore $\lambda_t(m)$ actually depends only on a sub-marking.

Stochastic Petri net, Semantics. Transitions become enabled as usual, i.e. if all pre-places are sufficiently marked. However there is a time, which has to elapse, before an enabled transition $t \in T$ fires. The transition's waiting time is an exponentially distributed random variable X_t with the *probability density function*:

$$f_{X_t}(\tau) = \lambda_t(m) \cdot \mathrm{e}^{(-\lambda_t(m) \cdot \tau)}, \qquad \tau \geq 0.$$

The firing itself does not consume time and again follows the standard firing rule of qualitative Petri nets. The semantics of a stochastic Petri net (with exponentially distributed reaction times for all transitions) is described by a continuous time Markov chain (CTMC). The CTMC of a stochastic Petri net is isomorphic to the reachability graph of the underlying qualitative Petri net, while the arcs between the states are now labelled by the transition rates. For more details see [Mur89], [BK02].

Based on this general $\mathcal{SPN}_{Bio}$ definition, specialised biochemically interpreted stochastic Petri nets can be defined by specifying the required kind of

stochastic hazard function more precisely. We give two examples, reading the tokens as molecules or as concentration levels. The *stochastic mass-action hazard function* tailors the general $\mathcal{SPN}_{Bio}$ definition to biochemical mass-action networks, where tokens correspond to molecules:

$$h_t := c_t \cdot \prod_{p \in {}^\bullet t} \binom{m(p)}{f(p,t)}, \tag{1}$$

where c_t is the transition-specific stochastic rate constant, and $m(p)$ is the current number of tokens on the pre-place p of transition t. The binomial coefficient describes the number of non-ordered combinations of the $f(p,t)$ molecules, required for the reaction, out of the $m(p)$ available ones.

Tokens can also be read as concentration levels, as introduced in [CVGO06]. The current concentration of each species is given as an abstract level. We assume the maximum molar concentration is M, and the amount of different levels is $N+1$. Then the abstract values $0, \ldots, N$ represent the concentration intervals $0,\ (0,\ 1*M/N],\ (1*M/N,\ 2*M/N],\ \ldots,\ (N-1*M/N,\ N*M/N]$. Each of these (finite many) discrete levels stands for an equivalence class of (infinitely many) continuous states. The *stochastic level hazard function* tailors the general $\mathcal{SPN}_{Bio}$ definition to biochemical mass-action networks, where tokens correspond to concentration levels:

$$h_t := k_t \cdot N \cdot \prod_{p \in {}^\bullet t} \left(\frac{m(p)}{N}\right), \tag{2}$$

where k_t is the transition-specific deterministic rate constant, and N the number of the highest level. The transformation rules between the stochastic and deterministic rate constants are well-understood, see e.g. [Wil06]. In practice, kinetic rates are taken from literature, textbooks etc. or determined from biochemical experiments. Hazard function (2) is the means whereby the continuous model (see the framework in Figure 2 and Section 6) can be approximated by the stochastic model; this can generally be achieved by a limited number of levels – see Section 5.2.

5.2 Stochastic Analysis

Due to the isomorphy of the reachability graph and the CTMC, all qualitative analysis results obtained in Section 4 are still valid. The influence of time does not restrict the possible system behaviour. Specifically it holds that the CTMC of our case study is reversible, which ensures ergodicity; i.e. we could start the system in any of the reachable states, always resulting in the same CTMC with the same steady state probability distribution.

In the following our main focus is on the analytic model checking approach. In Section 4.2 we employed CTL to express behavioural properties. Since we have now a stochastic model, we apply Continuous Stochastic Logic (CSL) [PNK06], which replaces the path quantifiers (**E**, **A**) in CTL by the probability operator

$\mathbf{P}_{\bowtie p}$, whereby $\bowtie p$ specifies the probability of the given formula. For example, introducing in CSL the abbreviation $\mathbf{F}\,\phi$ for $true\,\mathbf{U}\,\phi$, the CTL formula $\mathbf{EF}\,\phi$ becomes the CSL formula $\mathbf{P}_{\geq 0}[\,\mathbf{F}\,\phi\,]$, and $\mathbf{AF}\,\phi$ becomes $\mathbf{P}_{\geq 1}[\,\mathbf{F}\,\phi\,]$.

In order to use the probabilistic model checker PRISM [PNK06], we encode the extended ERK pathway in its modelling language, as proposed in [DDS04]. This translation requires knowledge of the boundedness degree of all species involved, which we acquire by the structural analysis technique of P-invariants.

We only consider here the level semantics. Since the continuous concentrations of proteins in the ERK pathway are all in the same range (0.1...0.4 mMol in 0.1 steps), we employ a model with only 4, and a second version with 8 levels. The corresponding CTMCs (and reachability graphs) comprise 24,065 states for the 4 level version and 6,110,643 states for the 8 level version.

Equivalence check by transient analysis. We start with a transient analysis to prove the sufficient equivalence between the stochastic model in the level semantics and the corresponding continuous model, justifying the interpretation of the properties gained by the stochastic model also in terms of the continuous one. The probabilistic model checker PRISM permits the analysis of the transient behaviour of the stochastic model, e.g., the concentration of RafP at time t is given by:

$$\mathrm{C}_{RafP}(t) \;=\; \frac{0.1}{s} \cdot \underbrace{\sum_{i=1}^{4s} \big(i \cdot P(L_{RafP}(t) = i)\big)}_{expected\ value\ of\ L_{RafP}(t)} \;.$$

The random variable $L_{RafP}(t)$ stands for the level of RafP at time t. We set s to 1 for the 4 level version, and to 2 for the 8 level version. The factor $\frac{0.1}{s}$ calibrates the expected value for a given level to the concentration scale. In the 4 level version a single level represents 0.1 mMol and 0.05 mMol in the 8 level version. Figure 3 shows the simulation results for the species MEK and Ras-GTP in the time interval [0..100] according to the continuous and the stochastic models respectively. These results confirm that 4 levels are sufficiently adequate to approximate the continuous model, and that 8 levels are preferable if the computational expenses are acceptable.

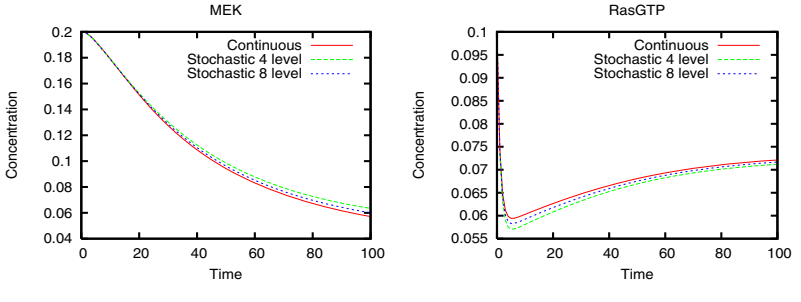

Fig. 3. Comparison of the concentration traces

Probabilistic model checking of special behavioural properties. We give two properties related to the partial order run of the I/O T-invariant, see Section 4.2 and qualitative property Q1 therein, from which we expect a consecutive increase of RafP, MEKPP and ERKPP. Both properties are expressed as so-called experiments, which are analysed varying the parameter L over all levels, i.e. 0 to N. For the sake of efficiency, we restrict the **U** operator to 100 time steps. Note that places are read as integer variables in the following.

Property S1: What is the probability of the concentration of RafP increasing, when starting in a state where the level is already at L (the latter side condition is specified by the filter given in braces)?

$$\mathbf{P}_{=?}\ [\ (\ \mathrm{RafP}\ =\ \mathrm{L}\ )\ \mathbf{U}^{<=100}\ (\ \mathrm{RafP} > \mathrm{L}\ )\ \{\ \mathrm{RafP} = \mathrm{L}\ \}\ ]$$

The results indicate, see Figure 4(a), that it is absolutely certain that the concentration of RafP increases from level 0 and likewise there is no increase from level N; this behaviour has already been determined by the qualitative analysis. Furthermore, an increase in RafP is very likely in the lower levels, increase and decrease are almost equally likely in the intermediate levels, while in the higher levels, but obviously not in the highest, an increase is rather unlikely (but not impossible). In summary this means that the total mass, circulating within the first layer of the signalling cascade, is unlikely to be accumulated in the activated form. We need this understanding to interpret the results for the next property.

Property S2: What is the probability that, given the initial concentrations of RafP, MEKPP and ERKPP being zero, the concentration of RafP rises above some level L while the concentrations of MEKPP and ERKPP remain at zero, i.e. RafP is the first species to react?

$$\mathbf{P}_{=?}\ [\ (\ (\ \mathrm{MEKPP}\ =\ 0\ )\ \wedge\ (\ \mathrm{ERKPP}\ =\ 0\ )\ )\ \mathbf{U}^{<=100}\ (\ \mathrm{RafP} >\ \mathrm{L}\ )$$
$$\{\ (\ \mathrm{MEKPP}\ =\ 0\ )\ \wedge\ (\ \mathrm{ERKPP}\ =\ 0\ )\ \ \wedge\ (\ \mathrm{RafP} =\ 0\ )\ \}\ ]$$

The results indicate, see Figure 4(b), that the likelihood of the concentration of RafP rising, while those of MEKPP and ERKPP are zero, is very high in the bottom half of the levels, and quite high in the lower levels of the upper half. The decrease of the likelihood in the higher levels is explained by property S1. Property S2 is related to the qualitative property Q1 (Section 4.2), and the

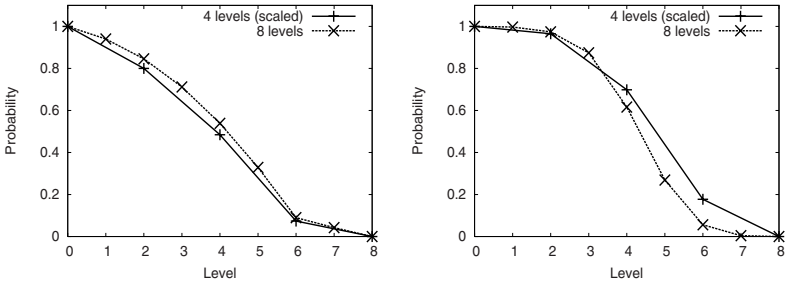

Fig. 4. Probability of the accumulation of RafP. (a) property S1. (b) property S2.

continuous property C1 (Section 6.2) – the concentration of RafP rises before those of MEKPP and ERKPP.

Due to the computational efforts of probabilistic model checking we are only able to treat properties over a stochastic model with 4 or at most 8 levels. This restricts the kind of properties that we can prove; e.g., in order to check increases of MEKPP and ERKPP – as suggested by the qualitative property Q1 and done above for RafP in the stochastic properties S1 and S2 – we would need 50 or 200 levels respectively.

Analytic probabilistic model checking becomes more and more impracticable with increasing size of the state space. Hence, the computation time of a probabilistic experiment, which typically consists of a series of probabilistic queries, can easily exceed several hours on a standard workstation. In order to avoid the enormous computational power required for larger state spaces, the time-dependent stochastic behaviour can be simulated by dedicated algorithms, e.g. [Gil77], or approximated by a continuous one, see next section.

6 The Continuous Approach

6.1 Continuous Modelling

In a continuous Petri net the marking of a place is no longer an integer, but a positive real number, which can be read as the concentration of the species modelled by the place. Transitions fire continuously, whereby the current deterministic firing rate generally depends on the current marking of the pre-places, i.e. of the current concentrations of the reactants. For our running case study, we derive the continuous model from the qualitative Petri net by associating a *mass action rate* with each transition in the network, i.e., the reaction labels are now read as the deterministic rate constants. We can likewise derive the continuous Petri net from the stochastic Petri net by approximating over the hazard function of type (1), see for instance [Wil06]. In both cases, we obtain a *continuous Petri net*, preserving the structure of the discrete one, see Figure 2.

The semantics of a continuous Petri net is defined by a system of ODEs, whereby one equation describes the continuous change over time on the token value of a given place by the continuous increase of its pre-transitions' flow and the continuous decrease of its post-transitions' flow, i.e., each place subject to changes gets its own equation. See [GH06] for more details.

The initial concentrations as suggested by the qualitative analysis correspond to those given in [LBS00], when mapping non-zero values to 1. For reasons of better comparability we have also considered more precise initial concentrations, where the presence of a species is encoded by biologically motivated real values varying between 0.1 and 0.4 in steps of 0.1. The complete system of non-linear ODEs generated from the continuous Petri net is given in [GHL07].

6.2 Continuous Analysis

Steady state analysis. Since there are 22 species, there are 2^{22} possible initial states in the qualitative Petri net (Boolean token interpretation). Of these,

118 were identified by the reachability graph analysis (Section 4.2) to form one strongly connected component, and thus to be "good" initial states. We then computed the steady state of the set of species for each possible initial state. In summary, our results show that all of the 'good' 118 states result in the same set of steady state values for the 22 species in the pathway, within the bounds of computational error of the ODE solver. None of the remaining possible initial states results in a steady state close to that generated by the 118 markings in the reachability graph; for details see [GHL07]. This is an interesting result, because the net considered here is not covered by the class of net structures discussed in [ADLS06] with the unique steady state property.

In Figure 5 (a) we reproduce the computed behaviour of MEK for all 118 good initial states, showing that despite differences in the concentrations at early time points, the steady state concentration is the same in all 118 states.

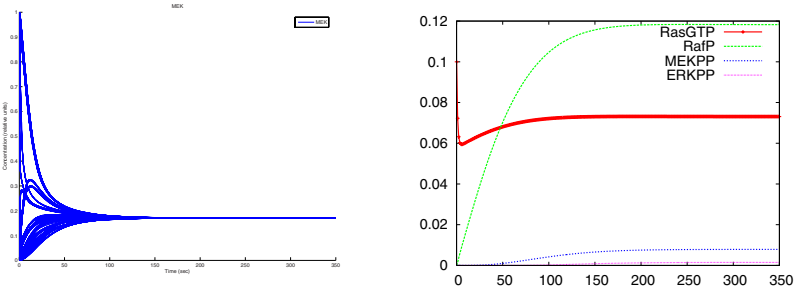

Fig. 5. (a) Steady state analysis of MEK for all 118 'good' states. (b) Continuous transient analysis of the phosphorylated species RasP, MEKPP, ERKPP, triggered by RasGTP.

Continuous model checking of the transient behaviour. Corresponding to the partial order run of the I/O T-invariant, see Section 4.2, we expect a consecutive increase of RafP, MEKPP, ERKPP, which we get confirmed by the transient behaviour analysis, compare Figure 5 (b). To formalise the visual evaluation of the diagram we use the continuous linear logic LTLc [CCRFS06], which is interpreted over the continuous simulation trace of ODEs.

The following three queries confirm together the claim of the expected propagation sequence. In the queries we have to refer to absolute values. The steady state values are obtained from the steady state analysis in the previous section; these are 0.12 mMol for RafP, 0.008 mMol for MEKPP and 0.002 mMol for ERKPP, all of them being zero in the initial state. If a species' concentration is above half of its steady state value, we call this concentration level significant. Note that in order to simplify the notation, places are interpreted as real variables in the following.

Property C1: The concentration of RafP rises to a significant level, while the concentrations of MEKPP and ERKPP remain close to zero; i.e. RafP is really the first species to react.

$$((\text{MEKPP} < 0.001) \wedge (\text{ERKPP} < 0.0002)) \ \mathbf{U} \ (\text{RafP} > 0.06)$$

Property C2: if the concentration of RafP is at a significant concentration level and that of ERKPP is close to zero, then both species remain in these states until the concentration of MEKPP becomes significant; i.e. MEKPP is the second species to react.

$$((\text{RafP} > 0.06) \wedge (\text{ERKPP} < 0.0002)) \Rightarrow ((\text{RafP} > 0.06) \wedge (\text{ERKPP} < 0.0002)) \mathbf{U} (\text{MEKPP} > 0.004)$$

Property C3: if the concentrations of RafP and MEKPP are significant, they remain so, until the concentration of ERKPP becomes significant; i.e. ERKPP is the third species to react.

$$((\text{RafP} > 0.06) \wedge (\text{MEKPP} > 0.004)) \Rightarrow ((\text{RafP} > 0.06) \wedge (\text{MEKPP} > 0.004)) \mathbf{U} (\text{ERKPP} > 0.0005)$$

Note that properties C1, C2 and C3 correspond to the qualitative property Q1, and that S2 is the stochastic counterpart of C1.

7 Tools

The bipartite graph in Figure 1 and its interpretation as the three Petri net models have been done using Snoopy [Sno], a tool to design and animate hierarchical graphs, including SBML import.

The qualitative analyses have been made using the Integrated Net Analyser INA [SR99] and the Model Checking Kit [SSE04]. We employed PRISM [PNK06] for probabilistic model checking, and Biocham [CCRFS06] for LTLc-based continuous model checking.

MATLAB [SR97] was used to produce the steady state analysis of all initial states in the continuous model, and the transient analysis was done using BioNessie [Bio], an SBML-based simulation and analysis tool for biochemical networks.

8 Summary and Outlook

In this paper we have described an overall framework that relates the three major ways of modelling biochemical networks – qualitative, stochastic and continuous – and illustrated this in the context of Petri nets. In doing so we have given a precise definition of biochemically interpreted stochastic Petri nets. We have shown that the qualitative time-free description is the most basic, with discrete values representing numbers of molecules or levels of concentrations. The qualitative description abstracts over two timed, quantitative models. In the stochastic description, discrete values for the amounts of species are retained, but a stochastic rate is associated with each reaction. The continuous model describes amounts of species using continuous values and associates a deterministic rate with each reaction. These two time-dependent models can be mutually approximated by hazard functions belonging to the stochastic world.

We have illustrated our framework by considering qualitative, stochastic and continuous Petri net descriptions of the ERK signalling pathway, based on the

model from Levchenko et al [LBS00]. We have focussed on analysis techniques available in each of these three paradigms, in order to illustrate their complementarity. Our special emphasis has been on model checking, which is especially useful for transient behaviour analysis, and we have demonstrated this by discussing related properties in the qualitative, stochastic and continuous paradigms. Although our framework is based on Petri nets, it can be applied more widely to other formalisms which are used to model and analyse biochemical networks.

We are now working on the incorporation of deterministic time into stochastic models, as well as the integration of continuous and stochastic aspects into one model.

Acknowledgements. The running case study has been partly carried out by Sebastian Lehrack during his study stay at the Bioinformatic Research Centre of the University of Glasgow. This stay was supported by the Max Gruenebaum Foundation [MGF] and the UK Department of Trade and Industry Beacon Bioscience Programme.

We would like to thank Richard Orton and Xu Gu for the constructive discussions as well as Vladislav Vyshemirsky for his support in the computational experiments.

References

[ADLS06] Angeli, D., De Leenheer, P., Sontag, E.D.: On the structural monotonicity of chemical reaction networks. In: ICATPN 2003, pp. 7–12. IEEE Computer Society Press, Los Alamitos (2006)

[Bio] BioNessie. A biochemical pathway simulation and analysis tool. University of Glasgow, `http://www.bionessie.org`

[BK02] Bause, F., Kritzinger, P.S.: Stochastic Petri Nets. Vieweg (2002)

[CCRFS06] Calzone, L., Chabrier-Rivier, N., Fages, F., Soliman, S.: Machine learning biochemical networks from temporal logic properties. In: Priami, C., Plotkin, G. (eds.) Transactions on Computational Systems Biology VI. LNCS (LNBI), vol. 4220, pp. 68–94. Springer, Heidelberg (2006)

[CGP01] Clarke, E.M., Grumberg, O., Peled, D.A.: Model checking. MIT Press, Cambridge (2001)

[CKS07] Chickarmane, V., Kholodenko, B.N., Sauro, H.M.: Oscillatory dynamics arising from competitive inhibition and multisite phosphorylation. Journal of Theoretical Biology 244(1), 68–76 (2007)

[CVGO06] Calder, M., Vyshemirsky, V., Gilbert, D., Orton, R.: Analysis of signalling pathways using continuous time Markov chains. In: Priami, C., Plotkin, G. (eds.) Transactions on Computational Systems Biology VI. LNCS (LNBI), vol. 4220, pp. 44–67. Springer, Heidelberg (2006)

[DDS04] D'Aprile, D., Donatelli, S., Sproston, J.: CSL model checking for the GreatSPN tool. In: Aykanat, C., Dayar, T., Körpeoğlu, İ. (eds.) ISCIS 2004. LNCS, vol. 3280, pp. 543–552. Springer, Heidelberg (2004)

[GH06] Gilbert, D., Heiner, M.: From Petri nets to differential equations - an integrative approach for biochemical network analysis. In: Donatelli, S., Thiagarajan, P.S. (eds.) ICATPN 2006. LNCS, vol. 4024, pp. 181–200. Springer, Heidelberg (2006)

[GHL07] Gilbert, D., Heiner, M., Lehrack, S.: A unifying framework for modelling and analysing biochemical pathways using Petri nets. TR I-02, CS Dep., BTU Cottbus (2007)

[Gil77] Gillespie, D.T.: Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry 81(25), 2340–2361 (1977)

[LBS00] Levchenko, A., Bruck, J., Sternberg, P.W.: Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. Proc. Natl. Acad. Sci. USA 97(11), 5818–5823 (2000)

[MGF] Max-Gruenebaum-Foundation, `http://www.max-gruenebaum-stiftung.de`

[Mur89] Murata, T.: Petri nets: Properties, analysis and applications. Proc.of the IEEE 77 4, 541–580 (1989)

[PNK06] Parker, D., Norman, G., Kwiatkowska, M.: PRISM 3.0.beta1 Users' Guide (2006)

[Sno] Snoopy. A tool to design and animate hierarchical graphs. BTU Cottbus, CS Dep., `http://www-dssz.informatik.tu-cottbus.de`

[SR97] Shampine, L.F., Reichelt, M.W.: The MATLAB ODE Suite. SIAM Journal on Scientific Computing 18, 1–22 (1997)

[SR99] Starke, P.H., Roch, S.: INA - The Intergrated Net Analyzer. Humboldt University, Berlin (1999), `www.informatik.hu-berlin.de/~starke/ina.html`

[SSE04] Schröter, C., Schwoon, S., Esparza, J.: The Model Checking Kit. In: van der Aalst, W.M.P., Best, E. (eds.) ICATPN 2003. LNCS, vol. 2679, pp. 463–472. Springer, Heidelberg (2003)

[Wil06] Wilkinson, D.J.: Stochastic Modelling for System Biology, 1st edn. CRC Press, New York (2006)

Appendix

The data files of the model in its three versions and the analysis results are available at www-dssz.informatik.tu-cottbus.de/examples/levchenko. A self-contained documentation of the case study as well as related work is given in [GHL07].