Probabilistic Model Checking of the PDGF Signaling Pathway[*]

Qixia Yuan[1,2], Panuwat Trairatphisan[3], Jun Pang[1**],
Sjouke Mauw[1], Monique Wiesinger[3], and Thomas Sauter[3]

[1] Computer Science and Communications, University of Luxembourg, Luxembourg
[2] School of Computer Science and Technology, Shandong University, China
[3] Life Sciences Research Unit, University of Luxembourg, Luxembourg

Abstract. In this paper, we apply the probabilistic symbolic model checker PRISM to the analysis of a biological system – the Platelet-Derived Growth Factor (PDGF) signaling pathway, demonstrating in detail how this pathway can be analyzed in PRISM. Moreover, we compare the results from verification and ODE simulation on the PDGF pathway and demonstrate by examples the influence of model structure, parameter values and pathway length on the two analysis methods.

1 Introduction

Biological systems consist of components, which interact to influence each other and therefore the whole system's behavior. The field of systems biology aims to understand such complex interactions. Due to the similarity between biological systems and complex distributed/reactive systems studied in computer science [2], modeling and analyzing techniques developed in the field of formal methods can be applied to biological systems as well [3]. Due to efficient verification techniques, formal methods can analyze large systems exhibiting complex behaviors – this process is typically supported by automatic computer tools. This potentially gives formal methods an advantage, as *in silico* experiments are much easier to perform than *in vitro* experiments for the aim of analyzing and understanding biological systems. During the last decade, there has been a rapid and successful development in applying formal methods to systems biology – new formalisms are developed for systems biology to create models for biological phenomena, new algorithms and tools are specially designed and tailored for the analysis of such models (e.g., see [4–6]).

In this paper, we explore the usage of model checking for biological systems. Model checking is referred to as the automatic process of checking whether a system model satisfies a given specification (expressed as a temporal logic formula), by exhaustively exploring all possible executions of the system. This differs from simulation-based techniques, which only study a subset of the executions. More specifically, we focus on the probabilistic (or stochastic) model checking approach [7, 8], first introduced by Hart,

---

[*] An extended abstract appears in the proceedings of CompMod 2011 [1]. The first two authors made equal contributions to this work.
[**] Corresponding author.

Sharir and Pnueli [9], as biological systems usually have complicated stochastic behaviors. This technique is well-established and widely used for ascertaining the correctness of real-life systems, including distributed systems and communication protocols. In probabilistic (or stochastic) model checking, systems are normally represented by Markov chains or Markov decision processes. Properties of the models are expressed in quantitative extensions of temporal logics. In the literature, depending on the models used, usually probabilistic model checking has its focus on discrete-time Markov chains (DTMCs), while stochastic model checking deals with continuous-time Markov chains (CTMCs). Stochastic verification, in particular, has gained notable success in analyzing probabilistic systems including biological signaling pathways (e.g., see [10, 11]).

The stochasticity which occurs in biological signaling pathways can considerably affect the changes of biological processes. For instance, the stochasticity of initial conditions of caspases enzymes in the separatrix region can influence the cells to escape or enter apoptotic process [12]. In this paper, we use the probabilistic model checker PRISM [11] to yield a better understanding of the Platelet-Derived Growth Factor (PDGF) signaling pathway. PDGF, described approximately 30 years ago as a major mitogenic component of whole blood [13], is a growth factor that regulates cell growth and division. It promotes angiogenesis and also preserves vascular integrity through the recruitment of pericytes to endothelial tubes. Clinical studies reveal that aberrant expression of PDGF and its receptor is often associated with a variety of disorders such as atherosclerosis, fibroproliferative diseases and most importantly, neoplasia [13]. Deregulation of the PDGF signaling pathway plays a critical role in the development of many types of human diseases such as gastrointestinal stromal tumor and hypereosinophilic syndrome [14, 15, 13, 16, 17]. Based on intensive literature review, we have built the PDGF signal transduction model in ODE (Ordinary Differential Equation) format. The essential part of the PDGF signaling pathway contains the coupling of PDGF ligand to its receptor `PDGFR`, the negative regulatory mechanism on `PDGFR` and the activation of two main downstream signaling pathways, i.e., MAPK (Mitogen-Activated Protein Kinase) and PI3K/Akt pathways. In addition, there also exist positive and negative crosstalk interactions between different downstream signaling pathways (more details on the PDGF signaling pathway can be found in Sect. 3). In our study of the PDGF pathway, there are three main goals: (1) to analyze the dynamics of PDGF induced signaling, (2) to analyze the influence of the crosstalk reactions and (3) to analyze the importance of individual reactions/molecules on downstream signaling molecules. The first two can be used to check whether the constructed signaling pathway is consistent with respect to biological data, while the last one can lead us to some prediction. We have achieved these goals by stochastic verification using PRISM.

Moreover, we present the differences of the results obtained from ODE simulation and stochastic verification on the PDGF pathway, and demonstrate by examples the influence of model structure, parameter values and pathway length on the two analysis methods. In particular, we show that the two methods can predict the results differently, especially when parameter values are small.

Related work. There exists a large body of work on applying formal techniques to the analysis of biological systems. We focus on the use of PRISM and probabilistic model

checking in the literature, and other studies on the modeling and analysis of the PDGF pathway.

Calder et al. [18] perform a case study on the the RKIP inhibited ERK pathway using PRISM. Interestingly, they present a result stating that with a small number of molecules simulation results of their stochastic CTMC model and the corresponding ODE model are comparable. In this paper, we show that this result holds for the PDGF signaling pathway even with only one instance for each molecule (see Sect. 6). In [19], PRISM is used to study the MAPK cascade where a small subset of the MAPK pathway was modeled. The authors explain how the biological pathway can be modeled in PRISM and how this enables the analysis of a set of quantitative properties. In principle, these studies are correlated to our work as both ERK pathway and MAPK cascade are among the main components in the PDGF signaling pathway. However, the work [18] focuses on the molecules in the ERK pathway, which is a part of the MAPK pathway, and the results from [19] only cover the analysis for a subset of the long MAPK pathway. In our study, we investigate a more general representation of the MAPK pathway in PDGF signaling. Thus, these make the direct comparison of these studies to the results from our PDGF signaling pathway's analysis infeasible.

Pronk et al. [20] apply PRISM to the biological problem of codon bias. They show that the results obtained from the quantitative analysis in PRISM agree with the biological literature. Ribosome kinetics and aa-tRNA competition are modeled as CTMCS analyzed in PRISM [21]. In [22], Kwiatkowska et al. use PRISM to analyze the FGF (Fibroblast Growth Factor) signaling pathway. Although only a model corresponding to a single instance of the pathway is built, it is still rich enough to explain the roles of the components in the pathway and how they interact. The tunable activation threshold hypothesis of T Cells is studied through computational modeling of T cell signaling pathways in PRISM [23], and the authors demonstrate tuning and synergy. Jha *et al.* [24] present the first algorithm for performing statistical model checking using Bayesian sequential hypothesis testing and test the algorithm on the FGF signaling pathway and several others. More recently, Liò et al. use PRISM to diagnose the emerging of bone pathologies [25].

In addition, there are a few papers which apply the traditional methods in Systems Biology to study the PDGF signaling pathway. Zhang et al. [26] model the survival signaling in large granular lymphocyte leukemia, which is partly related to the PDGF signaling, using a Boolean model of the network's dynamics. Wang et al. [27] model the crosstalk interaction between MAPK and PI3K/Akt pathways in ODE format. The experimental data and the evidence of crosstalk reaction from [27] also partly contribute to the model structure and the justification of the reactions in our work.

Outline of the paper. In Sect. 2, we give an overview of probabilistic model checking and the tool PRISM. Sect. 3 describes the PDGF signaling pathway. In Sect. 4, we build a model in PRISM for the PDGF signaling pathway and describe several properties of the model that we are interested in. Our verification results are given in Sect. 5. In Sect. 6, we compare stochastic verification and ODE simulation by investigating the influence of model structure, parameter values and pathway length. Finally, we draw the conclusions of this paper and discuss some future work in Sect. 7.

2 Probabilistic Model Checking and PRISM

We briefly introduce stochastic verification and the model checker – PRISM [11].

2.1 CTMC and CSL

Probabilistic model checking is a variant of model checking, which aims at analyzing the correctness of finite state systems with a focus on quantitative aspects. Model checking of a system requires two inputs: a formal description of the system, which is usually given in a high-level modeling formalism (e.g., Petri nets or process algebra) and a specification of the system properties, which is usually given as temporal logic (e.g., CTL or LTL) formulas. After accepting the two inputs, a model checking tool then can verify whether the system satisfies the desired properties and give counter-examples if the system does not satisfy a certain property, by exploring all possible behaviors of the system exhaustively. As the word "probabilistic" indicates, probabilistic model checking focuses on systems with stochastic behaviors. Instead of asking the model checker "will the molecule become active in the end?", we can ask "what is the probability of the molecule being active at the steady state?" or "what is the probability of the molecule being active at time instant t?". In probabilistic model checking, systems are normally represented by Markov chains or Markov decision processes. In this paper, we use continuous-time Markov chains (CTMCs) to build the signaling pathway models and stochastic verification for thesis anlyses.

A CTMC can model both (continuous) real time and probabilistic choice by assigning rates at transitions between states. The formal definition of a CTMC is given as follows.

Definition 1. *Let $\mathbb{R}_{\geq 0}$ denote the set of non-negative reals and AP be a fixed finite set of atomic propositions. A CTMC is a tuple (S, R, L) where:*

- *S is a finite set of states;*
- *$R : S \times S \to \mathbb{R}_{\geq 0}$ is a transition rate matrix;*
- *$L : S \to 2^{AP}$ is a labeling function which associates each state with a set of atomic propositions.*

The transition rate matrix R assigns rates to each pair of states, which are used as parameters of the exponential distribution. A transition can occur between two states s and s' if $R(s, s') > 0$, and the probability of the transition being triggered within t time-units equals to $1 - e^{R(s,s') \cdot t}$. If $R(s, s') > 0$ for more than one state s', the first transition to be triggered determines the next state. Therefore, the choice of the successor state of s is probabilistic. The time spent in state s before any such transition occurs is exponentially distributed with $E(s) = \sum_{s' \in S} R(s, s')$. Hence, the probability of moving to state s' is $\frac{R(s,s')}{E(s)}$, i.e., the probability that the delay of going from s to s' "finishes before" the delays of any other outgoing transition from s. A path in a CTMC is a sequence σ in the form of $s_0 t_0 s_1 t_1 \cdots$ with $R(s_i, s_{i+1}) > 0$ and $t_i \in \mathbb{R}_{\geq 0}$ for all $i \geq 0$. The amount of time spent in s_i is denoted by t_i.

For a CTMC, we consider two types of state probabilities: *transient* probability is related to a state in the CTMC at a particular time instant, and *steady* probability

4

describes the state of the CTMC in a long run. If we denote the state of the CTMC at time t as $X(t)$, the transient probability at time t is the probability that the CTMC is in state s at time t, i.e., $p_s(t) = Pr\{X(t) = s\}$. Intuitively, the steady-state probability for being at state s is then defined as

$$p_s = \lim_{t \to \infty} p_s(t).$$

Corresponding to CTMC models, we use Continuous Stochastic Logic (CSL) to specify properties of built models. CSL, originally introduced by Aziz *et al.* [28], provides a powerful means to specify both path-based and traditional state-based performance measures on CTMCs.

Definition 2. *The syntax of CSL is given as follows:*

$$\phi ::= true \mid a \mid \neg\phi \mid \phi \wedge \phi \mid P_{\sim p}[\phi U^I \phi] \mid S_{\sim p}[\phi]$$

where a is an atomic proposition, $\sim \in \{<, \leq, \geq, >\}$, $p \in [0, 1]$, and I is an interval of $\mathbb{R}_{\geqslant 0}$.

CSL formulas are evaluated over the states of a CTMC. CSL includes the standard operators from propositional logic: $true$ (satisfied in all states); atomic propositions (a is true in states which are labelled with a); negation ($\neg\phi$ is true if ϕ is not); and conjunction ($\phi_1 \wedge \phi_2$ is true if both ϕ_1 and ϕ_2 are true). Other standard boolean operators such as disjunction ($\phi_1 \vee \phi_2 \equiv \neg(\neg\phi_1 \wedge \neg\phi_2)$) and implication ($\phi_1 \Rightarrow \phi_2 \equiv \neg\phi_1 \vee \phi_2$) can be derived from these in the usual way. CSL also includes two probabilistic operators, P and S, both of which include a probability bound $\sim p$. A formula $P_{\sim p}[\psi]$ is true in a state s if the probability of the path formula ψ being satisfied from state s meets the bound $\sim p$. A path formula evaluates to either true or false for a single path in a model. In this paper, we use a simple type of the path formula, $F^I\phi \equiv true\, U^I \phi$, called an *eventual* formula, which is true for a path σ if ϕ eventually becomes true for some time instant $t \in I$. Particularly, if the time interval is set to zero, e.g. $F^{[t,t]}\phi$, the formula is true for a path σ if ϕ becomes true at time instant t. The S operator is used to specify steady-state behavior of a CTMC, i.e., its behavior in the long-run or equilibrium. More precisely, $S_{\sim p}[\psi]$ asserts that the steady-state probability of being in a state satisfying ψ meets the bound $\sim p$. We refer the reader to the papers [29, 7, 8] for model checking algorithms for computing steady-state probabilities.

2.2 The model checker PRISM

PRISM [11] is a model checking tool developed at the universities of Birmingham and Oxford. It allows one to model and analyze systems containing stochastic behaviors. PRISM supports three kinds of models: discrete-time Markov chains (DTMCs), continuous-time Markov chains (CTMCs) and Markov decision processes (MDPs). Analysis is performed through model checking such systems against properties written in the probabilistic temporal logics PCTL if the model is a DTMC or an MDP, or CSL in the case of a CTMC, as well as their extensions for quantitative specifications and costs/rewards.

In PRISM a model consists of a number of modules that contain variables and can interact with each other. The values of the variables at any given time constitute the state of the module, and the local states of all modules decide the global state of the whole model. The behavior of a module, normally the changes in states which it can undergo, is specified by a set of guarded commands of the form:

$$[a] \, g \rightarrow r : u;$$

a is an action label in the style of process algebra, which introduces synchronization into the model. It can only be performed simultaneously by all modules that have an occurrence of action label a. If a transition does not have to synchronize with other transitions, then no action label needs to be provided for this transition. The symbol g is a predicate over all the variables in the system. A guarded command $g \rightarrow r : u$ means that if the guard g is true, the system is updated according to u with rate r, which is corresponding to the transition rate of CTMC. A transition updates the value of variables by giving their new value of the form $x' = \mathsf{expr}$, where x is a variable and its primed version x' refers to the value of x in the next state, expr is an expression built on the *unprimed* variables. If an update does not contain $x' = ...$, then the value of the variable x remains unchanged.

PRISM models can be augmented with information about rewards (or equivalently, costs). The tool can analyze properties which relate to the expected values of these rewards. A CTMC in PRISM can be augmented with two types of rewards: *state reward* associated with states which are accumulated in proportion to the time spent in the state, and *transition reward* associated with transitions which are accumulated each time the transition is taken. CSL is extended with quantitative costs/rewards as well, which is quite useful in analyzing the quantitative properties of a biological system, by introducing the R operator:

$$R \; ::= \; R_{\sim r}[I^{=t}] \mid R_{\sim r}[C^{\leq t}] \mid R_{\sim r}[F\phi] \mid R_{\sim r}[S]$$

where $\sim \in \{<, \leq, \geq, >\}$, $r, t \in \mathbb{R}_{\geq 0}$ and ϕ is a CSL formula. Intuitively, a state s satisfies $R_{\sim r}[I^{=t}]$ if from s the expected state reward at time instance t meets the bound $\sim r$; a state s satisfies $R_{\sim r}[C^{\leq t}]$ if the expected reward accumulated up until t time units past satisfies $\sim r$; a state s satisfies $R_{\sim r}[F\phi]$ if from s the expected reward accumulated before a state satisfying ϕ is reached meets the bound $\sim r$; and a state s satisfies $R_{\sim r}[S]$ if from s the long-run average expected reward satisfies $\sim r$.

It is often useful to take a quantitative approach – computing the actual probability that some behavior of a model is observed, rather than just verifying whether or not the probability is above or below a given bound. Hence, PRISM allows the P and S operators in CSL to take the following form: $P_{=?}[\psi]$ and $S_{=?}[\psi]$.

3 The PDGF Signaling Pathway

3.1 Biology of the PDGF signaling pathway

Cell signaling is part of a complex system in cellular communication. It allows the cells to activate a large number of signaling molecules and to regulate their activity. In

order to transfer a regulatory signal upon reception of a triggering stimulus, the signal is transformed into a chemical messenger within the signaling cell, e.g., via transfer of a phosphate group (phosphorylation) [30]. For further details on cell signaling see, for example, [31, 30].

Platelet-Derived Growth Factor (PDGF), described approximately 30 years ago as a major mitogenic component of whole blood [13], is a growth factor that regulates cell growth and division. By binding to its receptor (PDGFR), it regulates many biological processes such as migration, survival and proliferation [32]. PDGFR is a receptor tyrosine kinase, which in general transfers upstream signals to many downstream signaling pathways by phosphorylation. Up to now, five pairs of PDGF which can be formed as a molecule that can bind with its receptor to form a complex (or so called 'ligand') are known, PDGF-AA, -AB, -BB, -CC, -DD, interacting with three different types of PDGFR complexes, PDGFR-$\alpha\alpha$, -$\alpha\beta$ and -$\beta\beta$. Each of the PDGFR subtypes has a different affinity to the different PDGF ligands [33].

After PDGFRs couple with their respective ligands, phosphorylation of the receptor at specific tyrosine residues will occur, thus enabling binding of signaling enzymes including Src, phosphatidylinositol 3 kinase (PI3K), phospholipase Cγ (PLCγ) and SHP2 in the MAPK pathway at specific binding sites. The recruitment of these signaling enzymes to PDGFR is mediated via an intrinsic SH2 domain. The translocation of PI3K and PLCγ to the plasma membrane also increases their accessibility to their respective substrates. Moreover, recent findings suggest that PDGFR also has potential binding sites for CrkL [34], which will activate Rap1 to positively influence c-Raf in the MAPK pathway [35], for Signal Transducer and Activator of Transcription (STAT), which might regulate the signal in parallel to the JAK-STAT pathway [36] and also for cCbl, which promotes ubiquitination of PDGFR. cCbl is also considered to be one of the negative regulatory molecules in PDGF signal transduction [37].

3.2 Model structure of the PDGF signaling pathway

Based on intensive literature reviews, we have built a PDGF signal transduction model in ODE format which consists of 17 molecules (see Fig. 1). The model consists of the main parts from the PDGF signaling pathway including the coupling of PDGF ligand to PDGFR, the negative regulatory feedbacks on PDGFR and the activation of two main downstream signaling pathways, the MAPK and PI3K/Akt pathways, with the crosstalk interactions between the two pathways. The scope of our model with 17 nodes is sufficient to capture the main dynamic behavior of the PDGF signaling pathway which is further analyzed in PRISM.

Fig. 1 describes how signals are transduced in the PDGF pathway by activating or deactivating specific downstream pathways signaling molecules. In this model, there are three inputs, viz. PDGFL (PDGF ligand), bPTEN, and bPDK. PDGFL is the node which represents the upstream molecule activating the whole network. PPX, and bPTEN are nodes which represent phosphatase enzymes in the cytoplasm that negatively regulate their targets. Lastly, bPDK, standing for basal activity of PDK, is the node that constantly gives a basal additional input to PDK node. This node is always active in order to activate the survival pathway to counteract apoptotic signaling and keep the cell survive at a basal level. There are three different types of arrows in the network: blue
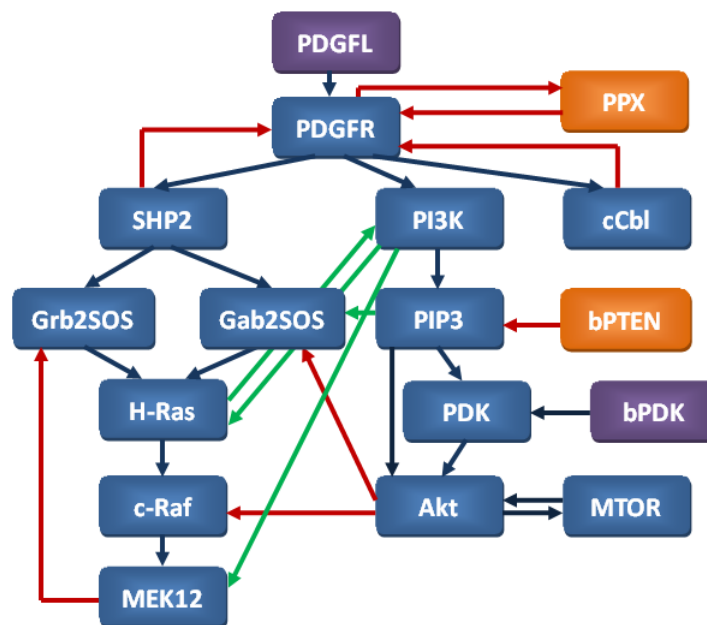
Fig. 1. The extracted PDGF signaling pathway (blue arrows: main pathway, green arrows: positive crosstalk, red arrows: negative regulatory)

arrows, green arrows and red arrows. The blue arrows represent the main activating interactions, indicating the two main downstream signaling pathways in the network. The MAPK pathway covers `SHP2`, `Grb2SOS`, `Gab2SOS`, `Ras`, `c-Raf` and `MEK12`. These molecules play a major role in the cellular proliferative circuit [38]. The PI3K/Akt pathway covers the molecules `PI3K`, `PIP3`, `PDK`, `bPTEN`, `bPDK` and `Akt`. In addition, this pathway closely interacts with the `MTOR` node which represents the mTOR pathway through positive feedback regulation. Both of these pathways play a major role together in the viability circuit of the cells [38]. The green arrows represent positive crosstalk interactions to the molecules the arrows point to. Lastly, the red arrows represent either negative crosstalk interactions or other negative regulatory interactions. The molecules will become active after they have been activated by either blue or green arrows. In contrary, the molecules will become inactive after they have been deactivated by the red arrows or the basal phosphatase activities in the cell (not shown in the figure).

`PDGFR` can be activated by `PDGFL`. The active `PDGFR` in turn activates three downstream molecules which are `SHP2`, `PI3K` and `cCbl`. Both `SHP2` and `cCbl` assert a negative feedback to PDGF making it inactive. The three blue arrows connecting `PDGFR` to these downstream signaling molecules, so called *mutant arrows*, are the targets of system interventions both experimentally and computationally. The experimental intervention can be performed by introducing a point mutation from tyrosine to phenylalanine at the specific recruitment site for the downstream signaling enzyme, for

instance, Y720F for `SHP2` recruitment site, YY731/742FF for `PI3K` recruitment site, and Y1018F for `cCbl` recruitment site, leading to the loss of signal capacity of the respective signaling pathway [39–41]. Thus, the result of computational simulation such as the relative activities at the steady state of downstream signaling molecules from these respective mutants can be validated experimentally in biological laboratories.

After the upstream signaling molecules, `SHP2` and `PI3K`, have been activated, they transfer the signal via the phosphorylation process to downstream signaling molecules. In the MAPK pathway, the process starts from the transfer of the signal from `SHP2` to `Grb2SOS` and `Gab2SOS`. Then, the two molecules in turn transfer the signal to `H-Ras`, `c-Raf`, and `MEK12`, respectively. There is also a negative feedback regulation from `MEK12` to `Grb2SOS` to modulate the signal in this pathway. In the PI3K/Akt pathway, the signal transfer starts from `PI3K` to `PIP3` which then in turn activates `PDK` and `Akt`. `PIP3` can be deactivated by the phosphatase enzyme `PTEN` (represented as `bPTEN`) and the node `PDK` can get additional input from the basal activity of itself (represented as `bPDK`). At the downstream part of the PI3K/Akt pathway, the node `Akt` can be activated by either `PDK` or `PIP3` and it also forms a positive feedback loop with the mTOR pathway which was represented as the `MTOR` node.

In addition to the activation and regulation within each pathway, there are also several crosstalk reactions between the two pathways. These are the positive crosstalk regulations from `PI3K` to `MEK12`, from `PIP3` to `Gab2SOS`, and a positive feedback loop between `PI3K` and `H-Ras`. In parallel, there are also negative regulations from `Akt` to `Gab2SOS` and `c-Raf`. These crosstalk reactions modulate the signals between the two pathways to generate a robust network.

Fig. 2 contains the list of model reactions. Each molecule is simplified to be in two states, either *inactive* or *active* (indicated by the suffix _act in the figure). All the reactions except for reactions 9, 10 and 11 describe the molecules changing between the two states; while reactions 9, 10 and 11 indicate the basal production, the basal degradation and the internalization following activation of `PDGFR`. For instance, reaction 2 describes that an active `PPX` gives a negative feedback to `PDGFR`, making it inactive.

3.3 Conversion of the interaction graph to a reaction-based ODE model

Generally, we can generate a set of biochemical reactions to represent the interactions for each molecule based on the interaction graph presented in Fig. 1. Nevertheless, the strengths of both activation/phosphorylation and inhibition/dephosphorylation interactions for each molecule are different. Therefore, during the conversion process of an interaction graph to biochemical reactions, a consistent procedure is applied to determine the biochemical reactions as well as the parameter values. In this section we will explain this procedure in some detail.

In our study, we assigned the parameter set based on the knowledge derived from the literature (e.g., [14, 15, 13, 16, 17]) and our own experimental observation of PDGFRα mutant. We also obtained the information that the time constants of the MAPK and PI3K/Akt pathways are comparable as shown in [27]. The experimental data in [27] shows that the time for the MAPK pathway to activate `ERK12` (the molecule below `MEK12`) and the time for the PI3K/Akt pathway to activate `Akt` are highly similar.

| a.	PDGFR and PPX	
1)	PDGFL + PDGFR → PDGFR_act + PDGFL	kon
2)	PDGFR_act + PPX_act → PDGFR + PPX_act	koff1
3)	PDGFR_act + SHP2_act → PDGFR + SHP2_act	koff2
4)	PDGFR_act + cCbl_act → PDGFR + cCbl_act	kubi
5)	PDGFR_act + PPX_act → PPX + PDGFR_act	koffppx
6)	PDGFR_act + cCbl → cCbl_act + PDGFR_act	k1
7)	PDGFR_act + SHP2 → SHP2_act + PDGFR_act	k5
8)	PDGFR_act + PI3K → PI3K_act + PDGFR_act	k6
9)	→ PDGFR	kbasal
10)	PDGFR →	kbasal
11)	PDGFR_act →	kdeg
b.	SHP2, Grb2SOS and Gab2SOS	
12)	SHP2_act + Grb2SOS → Grb2SOS_act + SHP2_act	k52
13)	SHP2_act + Gab2SOS → Gab2SOS_act + SHP2_act	k522
14)	SHP2_act → SHP2	kp5
15)	Grb2SOS_act + H-Ras → H-Ras_act + Grb2SOS_act	k53
16)	Grb2SOS_act → Grb2SOS	kp52
17)	MEK12_act + Grb2SOS_act → Grb2SOS + MEK12_act	kcross1
18)	Gab2SOS_act + H-Ras → H-Ras_act + Gab2SOS_act	k532
19)	Gab2SOS_act → Gab2SOS	kp522
20)	PIP3_act + Gab2SOS → Gab2SOS_act + PIP3_act	kcross2
21)	Akt_act + Gab2SOS_act → Gab2SOS + Akt_act	kcross3
c.	H-Ras, c-Raf and MEK12	
22)	H-Ras_act + c-Raf → c-Raf_act + H-Ras_act	k54
23)	H-Ras_act → H-Ras	kp53
24)	PI3K_act + H-Ras → H-Ras_act + PI3K_act	kcross4
25)	H-Ras_act + PI3K → PI3K_act + H-Ras_act	kcross9
26)	c-Raf_act + MEK12 → MEK12_act + c-Raf_act	k55
27)	c-Raf_act → c-Raf	kp54
28)	Akt_act + c-Raf_act → c-Raf + Akt_act	kcross6
29)	MEK12_act → MEK12	kp55
30)	PI3K_act + MEK12 → MEK12_act + PI3K_act	kcross7
d.	PI3K, PIP3 and PDK	
31)	PI3K_act + PIP3 → PIP3_act + PI3K_act	k62
32)	PI3K_act → PI3K	kp6
33)	PIP3_act + PDK → PDK_act + PIP3_act	k63
34)	PIP3_act + Akt → Akt_act + PIP3_act	k64
35)	PIP3_act + bPTEN → PIP3 + bPTEN	kp62
36)	bPDK + PDK → PDK_act + bPDK	k632
37)	PDK_act + Akt → Akt_act + PDK_act	k642
38)	PDK_act → PDK	kp63
e.	Akt, cCbl, MTOR, bPTEN and bPDK	
39)	MTOR_act + Akt → Akt_act + MTOR_act	k643
40)	Akt_act + MTOR → MTOR_act + Akt_act	k644
41)	Akt_act → Akt	kp64
42)	cCbl_act → cCbl	kp1
43)	MTOR_act → MTOR	kp642

Fig. 2. List of biochemical reactions in the PDGF signaling pathway ODE model with respective parameters in tab-separated format

All parameters are assigned as relative values in the range of zero to one, compared to the sum of all positive reaction rates around the respective molecule. We believe that this assignment could still capture the real reaction rates on the same molecule. Moreover, these parameter values should be able to represent the rates which are comparable to another signaling pathway as the time constants between the two signaling pathways are relatively similar. In general, we follow Kwiatkowska et al.'s work [10, 11] on translating kinetic rates to stochastic rates. Namely, for first-order (non-binary) reactions they take the stochastic rate to be the kinetic rate; for binary reactions, assuming that the kinetic rate is given in terms of molar concentrations, the stochastic rate can be obtained by dividing by the product of the volume and Avogadro's number [42]. In the case of bimolecular reactions (second order) modeled by the standard law of mass action as used in our paper (e.g., reactions 2-8), the kinetic rate and the stochastic rate are equal if the states are normalized to the maximal value (as done in our paper), according to the conversion of mole in chemistry.

To assign the parameter set, the reaction rate of that respective reaction is equal to one if only one molecule which activates the respective node is present. For example, the node SHP2 is solely activated by PDGFR once PDGFR is active (*PDGFR_act*). Thus, the biochemical reaction for the activation of the node SHP2 is generated as follow:

$$PDGFR\_act + SHP2 \xrightarrow{k5} SHP2\_act + PDGFR\_act$$
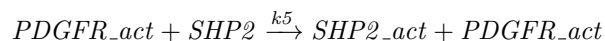
SHP2 will only be activated by PDGFR_act. Therefore, the respective parameter *k5* is assigned to 1.0. This rule is also applied to all nodes which have only one positive interaction on the interaction graph such as cCbl or MTOR. Apart from this, the activity of the node SHP2 is also under the influence of the activities of cytoplasmic phosphatase enzymes which dephosphorylate the active form of SHP2 (*SHP2_act*) resulting in the dephosphorylated/inactive state (*SHP2*). This is represented in the following biochemical reaction:

$$SHP2\_act \xrightarrow{kp5} SHP2$$
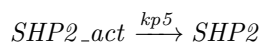
From biological observation, the reaction rate of the phosphorylation process is significantly higher than the rate of the dephosphorylation process [43]. In our study, it is assumed that the sum of the dephosphorylation strength for each molecule by cytoplasmic phosphatase enzymes in general is roughly 10% compared to the maximal phosphorylation strength. Therefore, parameter *kp5*, as well as all *kp*-parameters which represent the constant parameters for the dephosphorylation process, are all set to 0.1. In this case, we can derive an ODE for the *SHP2_act* node as follows:

$$d/dt(SHP2\_act) = k5 \times PDGFR\_act \times SHP2 - kp5 \times SHP2\_act$$

In another case, if there is more than one positive interaction on a single node in the interaction graph, we assume that the sum of all activation parameters is equal to one. In this situation, we additively consider the strength of each interaction according to their interaction strengths derived from the literature, without considering synergistic effects. These interaction strengths are then considered in term of relative values which have been assigned to each parameter accordingly. In general, the reaction rate of the

11

canonical pathway (the main pathway of activation) is always higher than the rate of other additional inputs such as crosstalk reactions or basal activities. For instance, node `PDK` can be activated by node `PIP3` with an interaction strength of 90%, while it can also be activated by node `bPDK` which represents the basal activity of `PDK` by 10%. In this case, two biochemical reactions of PDK node activation are generated as follows:
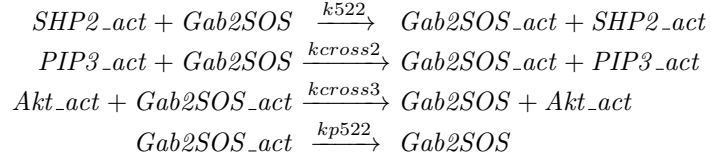
$$PIP3\_act + PDK \xrightarrow{k63} PDK\_act + PIP3\_act$$
$$bPDK + PDK \xrightarrow{k632} PDK\_act + bPDK$$
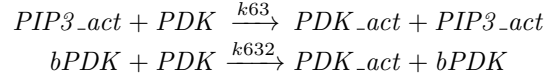
In this case, parameter *k63* has been assigned to 0.9 while parameter *k632* has been assigned to 0.1, according to their interaction strengths as mentioned. When also considering the dephosphorylation rate, we could derive an ODE for the *PDK_act* node as follows:

$$d/dt(PDK\_act) = k63 \times PIP3\_act \times PDK + k632 \times bPDK \times PDK$$
$$- kp63 \times PDK\_act$$

The same rule is also applied to nodes with the same type of interaction such as `H-Ras` or `Akt`.

In the most complex case, in the situation that there are both positive and negative interactions on a single node, special consideration and assignment are applied. First, we separate the positive and negative interactions and we consider only the strength of all positive interactions as a sum value of one. Then, we consider the proportion of negative interaction to deactivate/dephosphorylate the molecule in addition to the activities from cytoplasmic phosphatase enzymes with the parameter values according to their inhibition strengths. To give an example, node `Gab2SOS` is activated by two nodes which are `SHP2` with a strength of 80% and `PIP3` with strength 20%, according to the maximal activation strength. Node `Gab2SOS` is also deactivated by node `Akt` with strength 20% and by cytoplasmic phosphatase enzymes with strength 10%, according to the maximal activation. All biochemical reactions which are related to the changing states of the `Gab2SOS` node are shown as follows:

$$SHP2\_act + Gab2SOS \xrightarrow{k522} Gab2SOS\_act + SHP2\_act$$
$$PIP3\_act + Gab2SOS \xrightarrow{kcross2} Gab2SOS\_act + PIP3\_act$$
$$Akt\_act + Gab2SOS\_act \xrightarrow{kcross3} Gab2SOS + Akt\_act$$
$$Gab2SOS\_act \xrightarrow{kp522} Gab2SOS$$

As already mentioned, the parameters for the positive interactions from `SHP2` and `PIP3` are firstly considered and they have been assigned to each reaction with a sum of 1.0. In this case, parameter *k522* is assigned to 0.8 and parameter *kcross2* is assigned to 0.2. Then, the additional negative interaction from `Akt` is considered and the respective parameter is assigned. Thus, parameter *kcross3* is assigned to 0.2. Lastly, the basal dephosphorylation activity from cytoplasmic phosphatase enzymes which is represented by parameter *kp522* is assigned to 0.1. In this case, an ODE is derived for *Gab2SOS_act* node as follows:

$$d/dt(Gab2SOS\_act) = k522 \times SHP2\_act \times Gab2SOS - kp522 \times Gab2SOS\_act$$
$$+ kcross2 \times PIP3\_act \times Gab2SOS$$
$$- kcross3 \times Akt\_act \times Gab2SOS\_act$$

After we applied this procedure to convert the interaction graph to a biochemical reaction-based ODE model, we obtained a set of biochemical reactions as well as the respective parameter set for our modeling analysis in PRISM.

3.4 Sensitivity analysis of the derived ODE model

As mentioned in Sect. 3.3, the parameter set which has been derived from the interaction graph integrated with knowledge from the literature and our own experimental observations might not fully correlate to the actual biological system. Nevertheless, the dynamic behavior of the system is still conserved in the model structure. In this section, we present the results from the sensitivity analysis to confirm the validity of this statement.

After we obtained the ODE model from the interaction graph, we imported the ODE model into Matlab using Systems Biology Toolbox 2 [44] where a global parameter sensitivity analysis with the FAST method is integrated [45]. The analysis was performed by perturbing all parameter values to observe how the state values would differ based on the perturbation. The perturbation scale used in this analysis is one order of magnitude (for instance, from 1 to 10 or to 0.1).
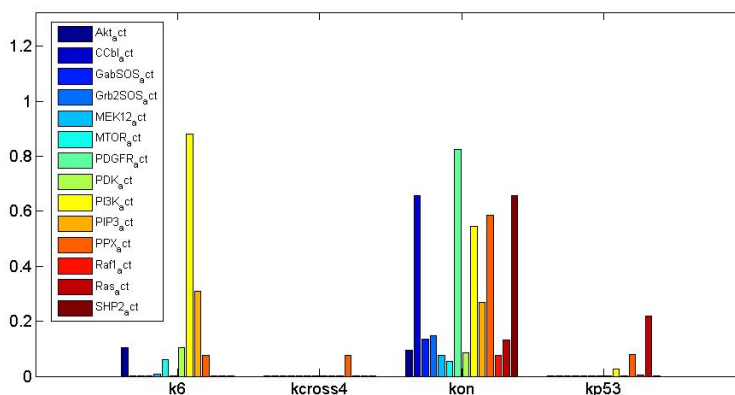

Fig. 3. Sensitivity analysis of the derived ODE model for some selected parameters

In Fig. 3, we see that some states are sensitive to the change of specific parameters. For instance, *PI3K_act* is sensitive to *k6* and *Ras_act* is sensitive to *kp53*. In parallel, we observe that there is no state, except for *PPX_act*, which is sensitive to the change of *kcross4*. In contrast, there are also many states which are sensitive to a change of a single parameter. For example, *kon*, which is the parameter involved with the activation of PDGFR by PDGF ligand and in turn activates the whole system, contributes to the changes of many activated form of molecules in the model when this parameter value is

13

perturbed. Based on this study, we identified the sensitive pairs of states and parameters such as *PI3K_act* which is sensitive to *k6*, *Ras_act* which is sensitive to *kp53* and most of the molecules in activated states which are all sensitive to *kon*. In addition, we also identified the insensitive pairs, such as *Ras_act* which is insensitive to *kcross4*.

Next, we challenged the system by increasing and decreasing selected parameter values from each pair up to 50%. For instance, the original *kon* parameter value of 1 was perturbed to 1.5 and 0.5, respectively. Then, the states trajectories of the respective molecules were plotted to observe the change of systems behavior according to the changes of the corresponding parameter values (see Fig. 4, 5 and 6).
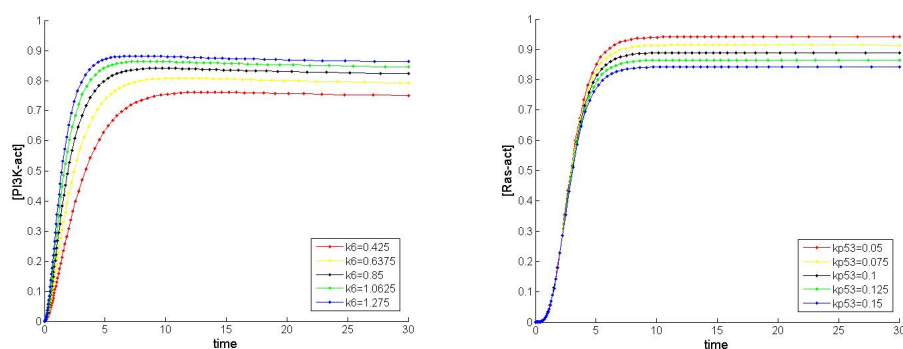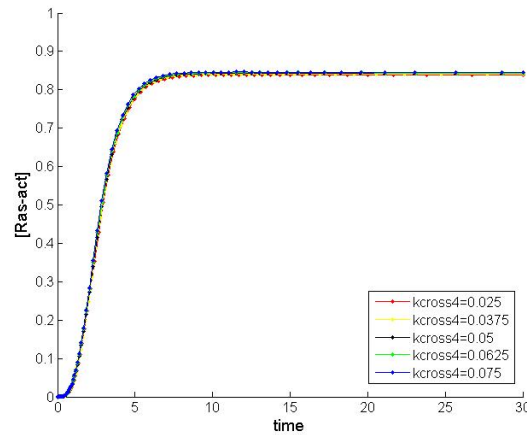

(a) *PI3K_act* concentration when *k6* changes from 25% to 50% (original *k6* = 0.85)

(b) *Ras_act* concentration when *k6* changes from 25% to 50% (original *kp53* = 0.1)

Fig. 4. Sensitivity analysis of *PI3K_act* concentration vs. *k6* and *Ras_act* concentration vs. *kp53*

From the results in Fig. 4, we observe the changes of the steady state values of *PI3K_act* and *Ras_act* which are related to changes of parameter values of *k6* and *kp53* accordingly. Nevertheless, we still see that the dynamic change of the molecules which are reflected by the shapes of the figures remained mostly the same. The same observation can be made for Fig. 5 where the *kon* parameter is perturbed and the state changes of *PDGFR_act*, *SHP2_act*, *PI3K_act*, *MEK12_act* and *Akt_act* are observed. Here, we see that *PDGFR_act* is activated slower and weaker (when $kon = 0.5$), or faster and stronger (when $kon = 1.5$), according to the *kon* values. Nevertheless, the dynamic changes (shape of the figures) as well as the steady state values of *PDGFR_act*, upstream molecules (*SHP2_act* and *PI3K_act*) and downstream molecules (*MEK12_act* and *Akt_act*) were not drastically changed. Moreover, the order of steady state values of these molecules is still preserved ($Akt\_act > MEK12\_act > PI3K\_act > SHP2\_act > PDGFR\_act$), even though the parameter value has been perturbed up to 50%. In parallel, when we plotted the state change of *Ras_act* after the perturbation on *kcross4*, little changes are observed as shown in Fig. 6, because *Ras_act* is insensitive to the change of this parameter. According to these results, we find that the decision to normalize the reaction rates in the interval between 0 and 1 is sensible to observe the dynamic changes of the systems which have been conserved within the model structure.

14
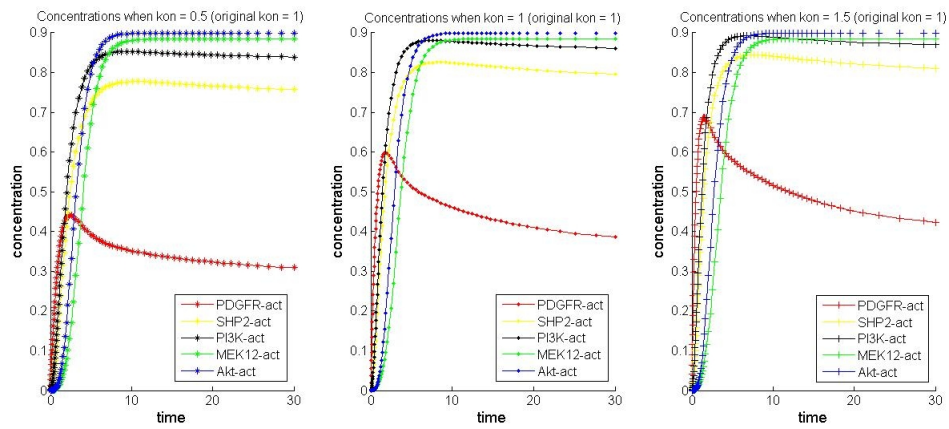
Fig. 5. Sensitivity analysis of the concentrations of *PDGFR_act*, *SHP2_act*, *PI3K_act*, *MEK12_act* and *Akt_act* vs. *kon*

Fig. 6. Sensitivity analysis of *Ras_act* concentration when *kcross4* changes from 25% to 50% (original *kcross4* = 0.05)

4 Modelling and Property Specifications in PRISM

4.1 PRISM model

We now describe in detail how to build a PRISM model for the PDGF signaling pathway as presented in the previous section. Though our model represents a single instance of the signaling pathway, meaning there can be at most one element of each molecule, we believe it is still rich enough to explain the roles of the molecules in the pathway

15

and how they interact with each other as shown in the literature [22]. In the single instance model, a molecule's steady state, which is expressed as a probability in PRISM, potentially corresponds to the molecule density in the biological experiments.

```
module PDGFR

    PDGFR : [0..2] init 0; //0 – inactive; 1 – active; 2 – degraded
    [] PDGFL & PDGFR=0 → kon : (PDGFR'=1);
    [bkoff1] PDGFR=1 → koff1 : (PDGFR'=0);
    [bkoff2] PDGFR=1 → koff2 : (PDGFR'=0);
    [bkubi] PDGFR=1 → kubi : (PDGFR'=0);
    [bkoffppx] PDGFR=1 → koffppx : (PDGFR'=1);
    [bk1] PDGFR=1 → k1 : (PDGFR'=1);
    [bk5] PDGFR=1 → k5 : (PDGFR'=1);
    [bk6] PDGFR=1 → k6 : (PDGFR'=1);
    [] PDGFR=0 → kbasal : (PDGFR'=2); //PDGFR degraded
    [] PDGFR=2 → kbasal : (PDGFR'=0);
    [] PDGFR=1 → kdeg : (PDGFR'=2); //PDGFR_act degraded

endmodule
```

(a) PRISM module for PDGFR

```
module PPX

    PPX : [0..1] init 1;
    [bkoff1] PPX=1 → (PPX'=1);
    [bkoffppx] PPX=1 → (PPX'=0);

endmodule
```

(b) PRISM model for PPX

```
rewards
"PDGFRactive"

    PDGFR=1:1;

endrewards
```

(c) PRISM rewards

Fig. 7. PRISM modules and rewards

Each of the nodes (or molecules) of the pathway in Fig. 1, except for the PDGFL, is represented by a separate PRISM module. Since PDGFL, bPTEN and bPDK remain the same after the reactions they are involved in, we set them as constant boolean values true. Fig. 7(a) and Fig. 7(b) show the modules for PDGFR and PPX. In each of the modules, the status of the molecule is represented by a variable with the same name as the module. The variables can have values of either 0 or 1 (PDGFR is an exception, because it can have value 2 since it can degrade), corresponding to the two states, inactive and active, of a molecule. Each command in the PRISM module represents a reaction in Fig. 2. Interactions of multiple molecules are implemented by the synchronization between modules. More precisely, the same label is given to the commands which require synchronization in PRISM modules. For example, in Fig. 7(a) and Fig. 7(b), there

16

are commands with label `bkoff1` in both of the modules `PDGFR` and `PPX`. The two commands are used to model the reaction (2) in Fig. 2, which involves both `PDGFR` and `PPX`. It guarantees that the two commands (corresponding to one reaction) can only occur when both guards are satisfied. The reaction rate is assigned by the command in module `PDGFR` and hence the reaction rate of the command in module `PPX` is omitted. We have modeled all the 17 molecules in 14 PRISM modules (`PDGFL`, `bPTEN` and `bPDK` are modeled as a constant).

As mentioned in Sect. 2.2, PRISM models can be augmented with information about rewards. We construct rewards to calculate the time for a molecule being active. Fig. 7(c) shows the rewards for calculating the active state of `PDGFR`. Each time `PDGFR` is in its active state, one is added to the total time of `PDGFR` being active. Similarly, we build rewards structures for other molecules as well, including `SHP2`, `Ras`, `MEK12`, `PIP3` and `Akt`.

4.2 Property specifications

As stated in the introduction, the three main goals for this study are: (1) to analyze the dynamics of PDGF induced signaling, (2) to analyze the influence of the crosstalk reactions as defined in Sect. 3, and (3) to analyze the importance of individual reactions on downstream signaling molecules. For the first goal, we study the signal transduction properties of each mutant by removing the mutant arrows one by one and examining how the states of each molecule change accordingly at different time instances. We also examine the total time for each molecule being active. Moreover, it is interesting to study the activities of each molecule at the steady state as well. For the second goal, we do the comparison of probabilities for molecules to be active between different mutants by removing each of the crosstalk reactions. For the last one, we study how the steady state probabilities of molecule `MEK12` and `Akt` change when a certain reaction is removed.

Below we list the properties of the PRISM model that we have analyzed to achieve our goals. Here, we use only the molecule `PDGFR` to illustrate the specification of the properties expressed as CSL formulas.

- $P_{=?}[\,F^{[t,t]}\mathtt{PDGFR} = 1\,]$
  The probability that the molecule `PDGFR` is active at time instant t.
- $R_{=?}^{\{\text{``}PDGFRactive\text{''}\}}[\,C <= t\,]$
  The expected time of `PDGFR` being active by time t. It refers to the reward structure "PDGFRactive" as defined in Fig. 7(c).
- $S_{=?}[\,\mathtt{PDGFR} = 1\,]$
  The long-run probability that `PDGFR` is active.

5 Verification Results

We use PRISM to construct the PDGF model described in Sect. 4.1 and analyze the set of properties listed in Sect. 4.2. We describe how to achieve our goals as explained in Sect. 1 in the following subsections.

5.1 Analyzing the dynamics of PDGF induced signaling

For the first goal to analyze the dynamics of PDGF induced signaling, we first develop a base model representing the system, in which all the reactions in Fig. 2 are included. Subsequently, we obtain the mutant models by removing the mutant arrows (as mentioned in Sect. 3) one by one. More precisely, we have developed four models in this experiment including the base model corresponding to the `WildType` condition. The second one, called `SHP2Mutant`, is obtained by removing the mutant arrow pointing to `SHP2`. The third and the fourth ones are `PI3KMutant` and `cCblMutant`. Following the same process, we remove the mutant arrows pointing to `PI3K` and `cCbl` separately in the last two models. The three removed mutant arrows correspond to the reactions (7), (8) and (6) in Fig. 2. For each model, we compute the probability of each molecule being active at time instance t, which is summarized in Fig. 8.
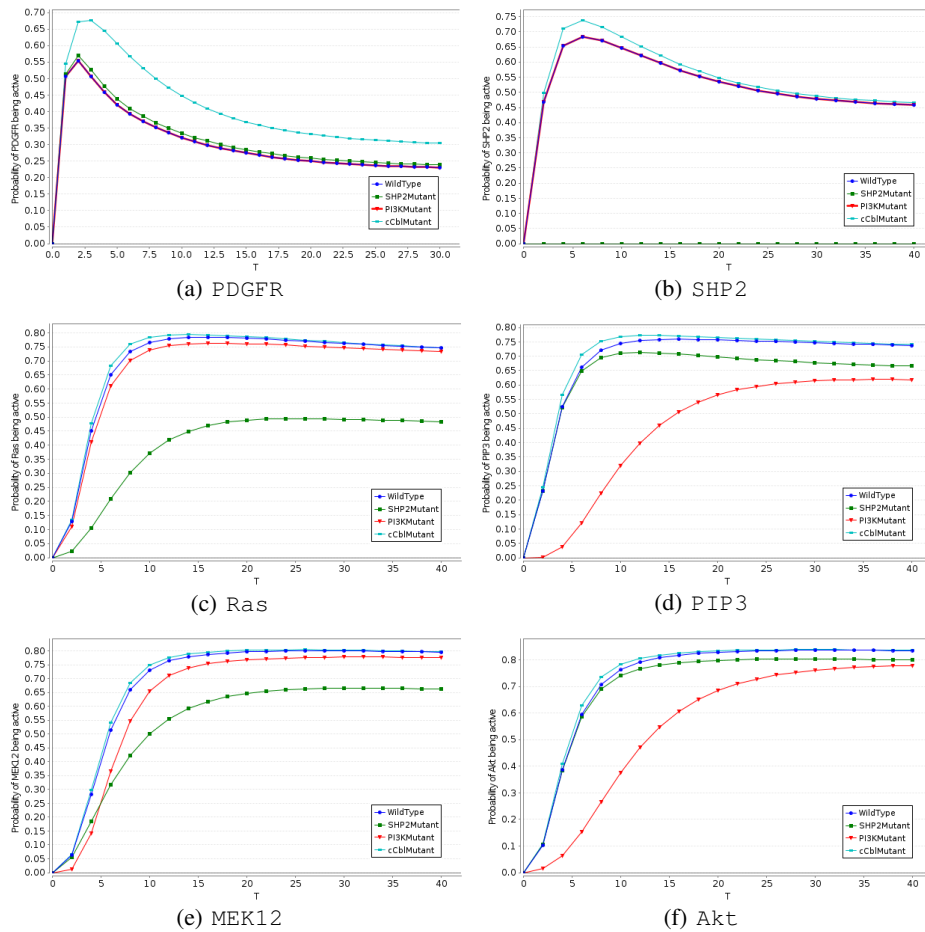
Fig. 8. PRISM verification results: Probabilities for molecules being active

18

Fig. 8 shows the probability of 6 out of the 17 molecules namely `PDGFR`, `SHP2`, `Ras`, `PIP3`, `MEK12`, and `Akt`. The 6 molecules are chosen according to their positions in the signaling pathway. We can see from Fig. 8(a) that `cCbl` affects `PDGFR` more than the other two molecules (`SHP2` and `PI3K`). This is due to strong negative feedback (red arrow) from `cCbl` to `PDGFR` in Fig. 1. There is also a negative feedback from `SHP2` to `PDGFR`. However, the reaction rate of the negative feedback from `cCbl` is 7 times as large as the one from `SHP2`, hence `cCbl` can affect `PDGFR` more than `SHP2`. In contrast, there would be no effect of `PI3KMutant` to `PDGFR` as there is neither a direct regulatory signal from `PI3K` on `PDGFR` nor indirect regulation through the molecule which can regulate `PDGFR` such as `SHP2` or `cCbl`. Apart from this, the red curve in Fig. 8(d) shows that not only `PIP3` in `PI3KMutant` is less active than in other conditions, but also the time of it becoming active is delayed. In biological experiments, this delay is *not* observed. This is due to the fact that there is another molecule `PLCg` giving a fast positive crosstalk to `PI3K`. This interaction is outside of the scope of the PDGF model we consider here.

If we focus on the light blue curves in Fig. 8(a) – Fig. 8(f), we can see that `cCbl` has much impact on `PDGFR` but little impact on the other molecules. This is because after `PDGFR` activates `SHP2` and `PI3K`, the states of the other molecules are determined by `SHP2` and `PI3K`. Furthermore, from the green curve of Fig. 8(b), Fig. 8(c) and Fig. 8(e), we see that `SHP2` has great effect on the status of the molecules on the MAPK pathway. However, the more downstream the molecule is, the less prominent the effect becomes (the effect to `MEK12` is less than that to `Ras`). Due to the influence of the positive feedback from `PI3K` and `PIP3`, the molecules in the MAPK pathway can also become active without `SHP2`, which is the result from positive crosstalk interactions. Besides, the red curves of Fig. 8(d) and Fig. 8(f) show that the influence of `PI3K` to the molecules in the PI3K/Akt pathway is still present due to the positive feedback from `Ras`.

We analyze the long-run probability of the molecules being active after `PDGFL` stimulation (shown in Tab. 1). The result demonstrates the activities of each molecule in this model in `WildType` condition. In the table, we observe that the two downstream signaling molecules, `MEK12` and `Akt`, have high long-run probabilities, i.e., 0.77 and 0.82, respectively. This is sensible in the real biology as these two signaling molecules should be at sufficiently high concentrations after PDGF ligand activation in order to push the cell toward proliferation and to keep the cell survive from apoptotic process. In addition, we also observe high steady state probabilities on `PDK` and `MTOR` which might be a result from the accumulative effect from the basal activity and the positive feedback loop, in addition to the activation from PDGF ligand. In the mutants situations, we expect to observe different steady state probability of each molecule which would then affect the cell fate decision differently in each condition.

PRISM supports reward properties (see Sect. 4). Fig. 9 shows the expected time of six molecules being active by time instant t. These six molecules are the same as in Fig. 8. All the six curves tend to be linear after time instant 12, which shows that the 6 molecules start to be in a steady state after time instant 12.

19

| Molecule	Probability	Molecule	Probability
PDGFR	0.22	Grb2SOS	0.55
SHP2	0.45	Ras	0.72
Gab2SOS	0.53	MEK12	0.77
c-Raf	0.63	PIP3	0.72
PI3K	0.62	Akt	0.82
PDK	0.83	MTOR	0.84
cCbl	0.47	PPX	0

Table 1. PRISM verification results: Steady state probabilities of molecules

5.2 Analyzing the influence of the crosstalk reactions

For the second goal, to analyze the influence of the crosstalk reactions, the experiment is also performed as in silico genetics. After developing the base model, we get the model variants by removing one crosstalk arrow at a time. In total, there are four positive crosstalk reactions (green arrows) and two negative crosstalk reactions (red arrows pointing from Akt). In this analysis, we focus on the positive crosstalk from Ras to PI3K and the negative crosstalk from Akt to c-Raf. More precisely, we develop three models in this experiment. The first one is the base WildType model with all reactions included. The second one is the RasPI3KMutant model, in which the positive feedback from Ras to PI3K is removed. The last one is the Aktc-RafMutant model, in which the negative feedback from Akt to c-Raf is removed. After building the three models, we analyzed the influence of the two crosstalks by comparing the WildType model to the two mutant models, respectively.
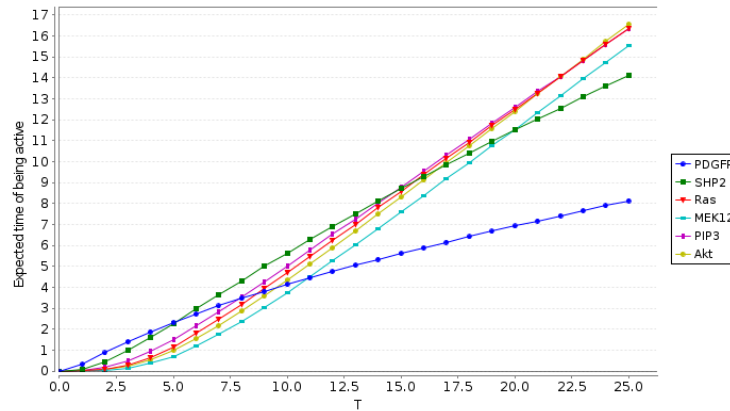
Fig. 9. PRISM verification results: Expected time of being active by time t

(a) Positive crosstalk from `Ras` to `PI3K`   (b) Negative crosstalk from `Akt` to `c-Raf`
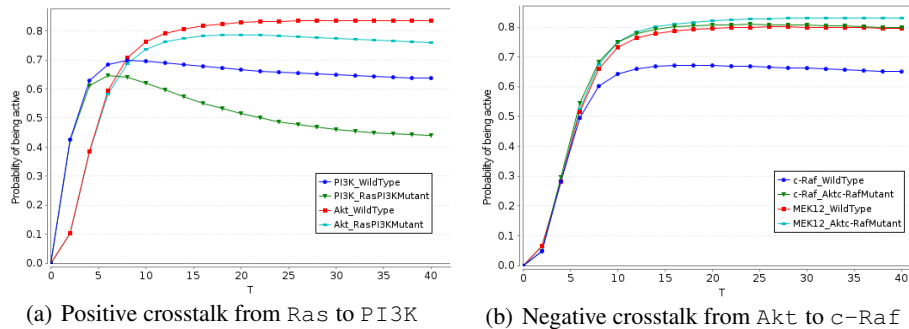
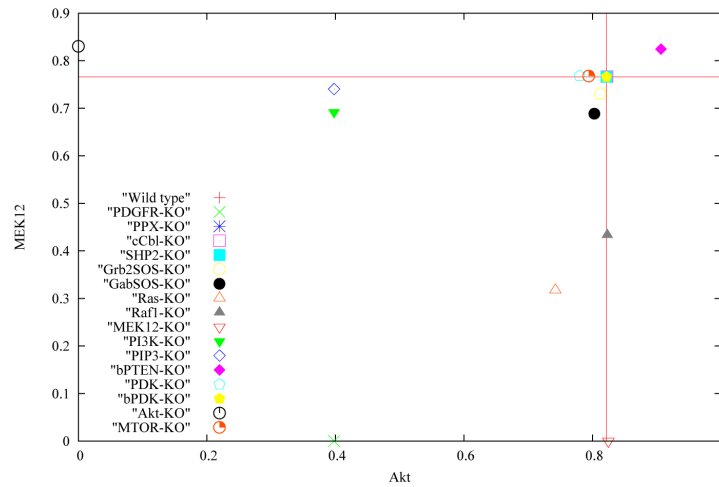**Fig. 10.** PRISM verification results: Influence of crosstalks

Fig. 10 shows the results of the comparison. In both sub-figures, we compare the molecules which are directly related to the crosstalk arrows and the molecules at the end of the related main signaling pathways. In Fig. 10(a), we compare in each model the probabilities for molecules `PI3K` and `Akt` being active, which is at the end of the PI3K/Akt pathway; while in Fig. 10(b) we have chosen the molecules `c-Raf` and `MEK12`, which is at the end of the MAPK pathway. The dark blue and green curves in Fig. 10(a) show how the green arrow influences the status of `PI3K`. If there is no positive crosstalk, the probability of `PI3K` being active (green curve) becomes smaller. The two curves are coincident for a time period because the molecule `Ras` needs some time to become active before it can activate `PI3K`. The red and light blue curves show that the influence of the positive feedback on `Akt` is smaller than that on `PI3K`. Just like the situation in Fig. 8(b), Fig. 8(c) and Fig. 8(e), the more downstream the molecules are, the smaller the influence becomes. The results in Fig. 10(b) are similar: the dark blue and green curves show that the influence the negative crosstalk gives to `Ras` is significant; while the positive crosstalk has little influence on `MEK12`.

5.3 Analyzing the importance of individual reactions

For the last goal to analyze the importance of individual reactions on downstream signaling molecules, we compute the steady state probabilities of `Akt` and `MEK12` in 31 different models (called an *edge knockout study*), each of which is obtained by removing one reaction from the `wildtype` model. For example, the 'PIP3-Gab2SOS' model is obtained by removing reaction No. 20 in Fig. 2). The results are shown in Fig. 11. In parallel to the edge knockout study, we have also computed the steady state probability of `Akt` and `MEK12` in 17 different models (called a *node knockout study*), each of which is obtained by removing all related interactions around each node. For instance, the interactions between `PDGFR`, `H-Ras`, `PIP3` and `MEK12` to `PI3K` are all omitted in the 'PI3K-KO' model. The results are shown in Fig. 12. By adding a horizontal and a vertical line through the dot of the wildtype model, we divide these figures into four areas. Apparently, the dots in different areas show that the removed reactions/nodes have different influences on the steady state probabilities of `Akt` and `MEK12`. For instance,
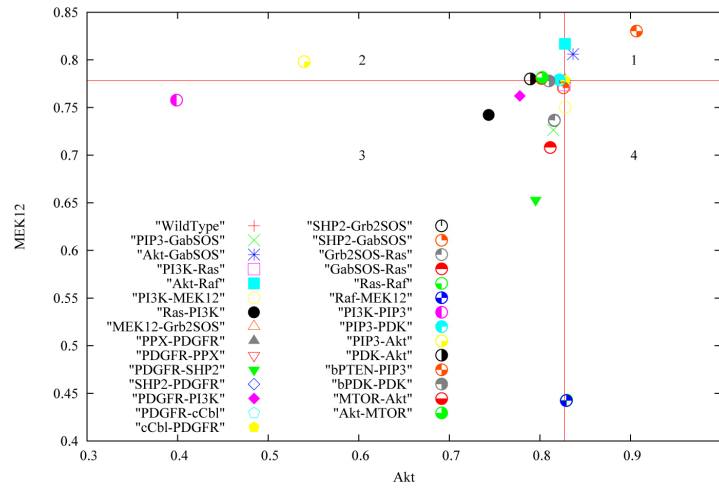
Fig. 11. PRISM verification results: Steady state probabilities of `Akt` and `MEK12` in different edge knockout models

Fig. 12. PRISM verification results: Steady state probabilities of `Akt` and `MEK12` in different node knockout models

the reactions (corresponding to the dots) in area 3 can decrease the steady state probabilities of both `Akt` and `MEK12`, while those in area 1 can increase both probabilities. The reactions in area 2 can decrease the steady state probability of `Akt` and increase the one of `MEK12`; while reactions grouped in area 4 lead to the opposite effect. For those dots lying on the horizontal line, the corresponding removed reactions have little

22

impact on the steady state probability of MEK12. Similarly, the reactions on the vertical line have little impact on Akt.

Comparing the two studies, the node knockout study is more relevant to biological observations, which could be observed under the kinase inhibitor treatment scheme. For instance, the 'PI3K-KO' model represents the cell system which has been treated with PI3K inhibitors such as Wortmannin or LY294002 drugs [46]. Moreover, these observations have biological implications as well. Akt and MEK12 are downstream molecules in the signal transduction process that regulate different cellular functions. The signal from Akt keeps the cells to survive from apoptosis and the signal from MEK12 regulates the cells' growth and proliferation. In cancer, both of these two main pathways (see Fig. 1) are more active so they drive the cells to keep growing and dividing in an uncontrolled manner. Therefore, if we could find the targets to control these two signaling molecules to be at a desirable level, it would be beneficial for cancer therapeutic development.

## 6 Comparing Stochastic Verification and ODE Simulation

In general, stochastic verification and ODE simulation produce comparable results, e.g., see [18]. To compare these approaches for the PDGF signaling pathway, we have conducted some ODE simulations as well, where the normalized initial concentrations of the molecules are set either to 0 or 1. Moreover, we have only one instance of each molecule in the PRISM model. Thus, we would expect that the steady state probabilities obtained from PRISM stochastic verification are comparable to the steady state values (concentrations) obtained from ODE simulation.

Two particular results are shown in Fig. 13(a) and Fig. 13(b), which correspond to Fig. 8(a) and Fig. 8(c). We can see that the dynamic behavior of the two molecules, PDGFR and Ras, are similar between the two methods. However, considerable quantitative differences have been observed between the two methods in the node knockout study as shown in Tab. 2. For instance, the difference of the steady state probabilities/values for Akt is roughly 0.18 in the PDGFR-KO and PI3K-KO scenarios.

Model	Akt (PRISM)	Akt (ODE)	MEK12 (PRISM)	MEK12 (ODE)
WildType	0.8217	0.8993	0.7660	0.8849
PDGFR-KO	0.3979	0.5748	0	0
PPX-KO	0.8217	0.8993	0.7660	0.8849
MEK12-KO	0.8245	0.8993	0	0
PI3K-KO	0.3979	0.5748	0.6918	0.8724

Table 2. Steady state probabilities (PRISM verification) and steady state values (ODE simulations) of Akt and MEK12

In order to better understand the source of these differences, we investigated the possible role of different model structures, parameter values, and pathway length. The

(a) `PDGFR`　　　　　　　　　　　　(b) `Ras`
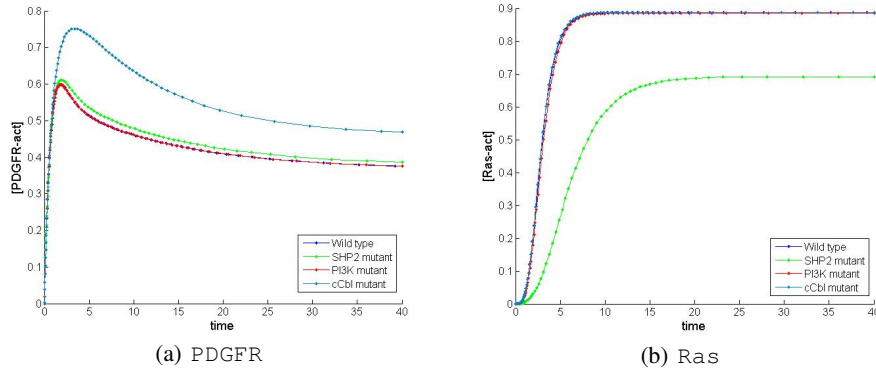
Fig. 13. ODE simulation results: Concentrations of molecules over time

prototypical model structures that we analyzed comprise a negative feedback loop as shown in Fig. 14(b) and Fig. 15(b), a positive feedback loop as shown in Fig. 14(c), and a linear chain (cascade) of molecules as shown in Fig. 15(a) and Fig. 15(b). Each molecule, except from the input node 'A', has two states active and inactive, to resemble the PDGF model. The impact of different parameter values was investigated by varying parameter 'k' in Fig. 14(c).

According to Fig. 14(d) and Fig. 15(c), we find that model structures with positive or negative feedbacks only have a minor contribution to the differences between the steady state probabilities from PRISM verification and steady state values from ODE simulation. Moreover, long-chain negative feedback loops which are generally present in biological systems, e.g., as shown in Fig. 15(b), do not cause such high differences either. On the other hand, when we consider the model with a positive feedback loop in Fig. 14(c), the differences between the two methods highly increase once the parameter values for the input node vary (see Fig. 14(e)). We observe that the decrease of parameter k in the model in Fig. 14(c) is correlated to the increase of the differences between the two methods. The steady state probabilities (PRISM verification) and steady state values (ODE simulation) for nodes B1 and B2 of the model in Fig. 14(c) with different k values are presented in Fig. 14(f) and Fig. 14(g). We can observe that the steady state probabilities from PRISM verification change drastically with different k values for both nodes, while the steady state values from ODE simulation do not change that much. This is essentially due to the different ways to compute steady state probabilities for CTMCs [47, 29] and to compute steady state values for ODE [48].

Thus, we see that in our setting the difference between the two methods mainly depends on parameter values and to some extent also on the model structure which correlates to the findings in the PDGF model. Here, we observed large differences e.g. for `Akt` in the `PDGFR-KO` conditions in the node knockout study as shown in Tab. 2. Under this condition `Akt` is activated only by `PDK` (from `bPDK`) and positive feedback from `MTOR` with a low parameter value of 0.1. This leads to a larger difference between the two methods, in contrast to the wild-type situation with larger parameter values.

24

This was in addition verified by analyzing model structures and parameters equal to the positive feedback loop in the PDGF model (which is not detailed in this paper).

In conclusion, we have demonstrated by examples the impact of specific parameter values and model structure on the differences between ODE simulation and stochastic verification. It has been stated that the reason for these differences might be that the approach based on verification leads to more accurate results when the number of molecules is small [49]. Nevertheless, the influence of different parameter values needs to be addressed in future research.

7   Conclusion and Future Work

In this paper, we have given a detailed description of using the probabilistic model checker PRISM to study the PDGF signaling pathway. Based on intensive literature reviews, we have built a PDGF signal transduction model in ODE format consisting of 17 molecule species and we converted it for the analysis in the probabilistic model checker PRISM. We have analyzed the dynamics of PDGF induced signaling, the influence of crosstalk reactions in the signaling pathway and the importance of individual reactions on downstream signaling molecules. Our experimental results show that stochastic verification can provide us with a good understanding of the PDGF signaling pathway, especially the result discussed in the end of Sect. 5.3 potentially can give rise to better behavior prediction of the pathway.[1]

Furthermore, we have also investigated the differences of the results of the two analysis methods, ODE simulation and stochastic verification (PRISM) as discussed in Sect. 6. We have demonstrated that the two methods can predict the results differently, especially when parameter values are small. Such results will bring more insights to determine which method is more suitable for the analysis of biological systems.
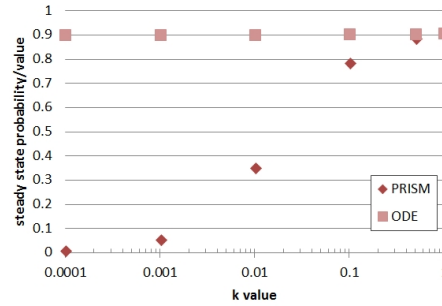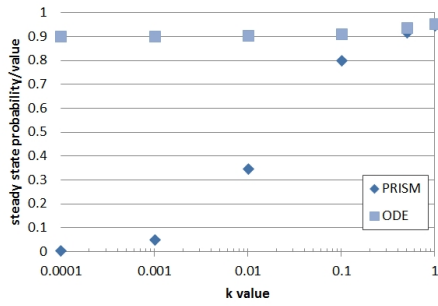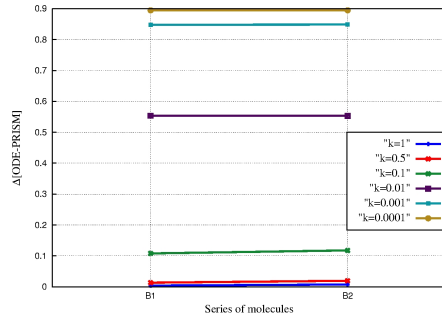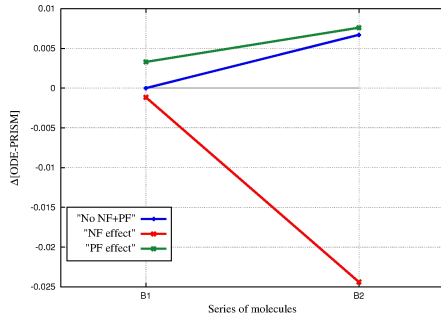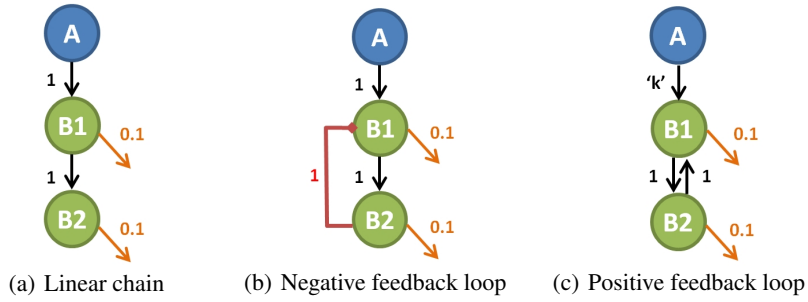
Nevertheless, as indicated at the end of Sect. 3, our model is not intended to be fully accurate. The reaction rates, especially the crosstalk reaction rates, are based on literature reviewing and are only relative values. Currently, experiments in a biological laboratory are performed to get more precise values for these reaction rates. We hope techniques like parametric verification for Markov chains (e.g., [50–52]) can help to synthesize values which are consistent with the lab experimental results.

References

1. Yuan, Q., Pang, J., Mauw, S., Trairatphisan, P., Wiesinger, M., Sauter, T.: A study of the PDGF signaling pathway with PRISM. In: Proc. 3rd Workshop on Computational Models for Cell Processes. Volume 67 of EPTCS. (2011) 65–81
2. Regev, A., Shapiro, E.: Cellular abstractions: Cells as computation. Nature **419** (2002) 343
3. Bonzanni, N., Anton Feenstra, K., Fokkink, W.J., Krepska, E.: What can formal methods bring to systems biology? In: Proc. 2nd World Congress on Formal Methods. Volume 5850 of LNCS., Springer (2009) 16–22

---

[1] The PRISM model and property specifications can be found at `http://satoss.uni.lu/jun/models/PDGF.zip`.

(a) Linear chain   (b) Negative feedback loop   (c) Positive feedback loop

(d) Differences between the steady state probabilities (PRISM verification) and steady state values (ODE simulation) for the models in Fig. 14(a)-14(c) with $k=1$

(e) Differences between the steady state probabilities (PRISM verification) and steady state values (ODE simulation) for the model in Fig. 14(c) with different k values

(f) The steady state probabilities (PRISM verification) and steady state values (ODE simulation) for node B1 in the model of Fig. 14(c) with different k values

(g) The steady state probabilities (PRISM verification) and steady state values (ODE simulation) for node B2 in the model of Fig. 14(c) with different k values

Fig. 14. Influence of model structures and parameter values on differences between PRISM verification and ODE simulation (PF: positive feedback; NF: negative feedback)

26

(a) Linear chain of 10 molecules (b) Linear chain with negative feedback loop
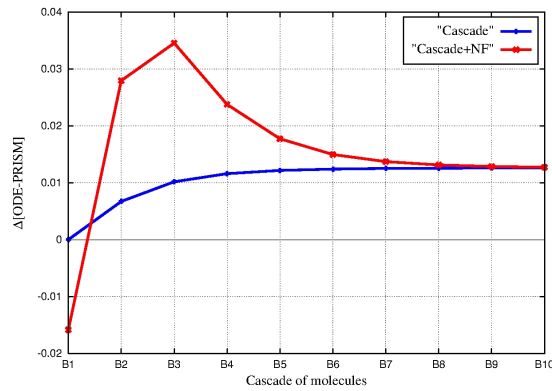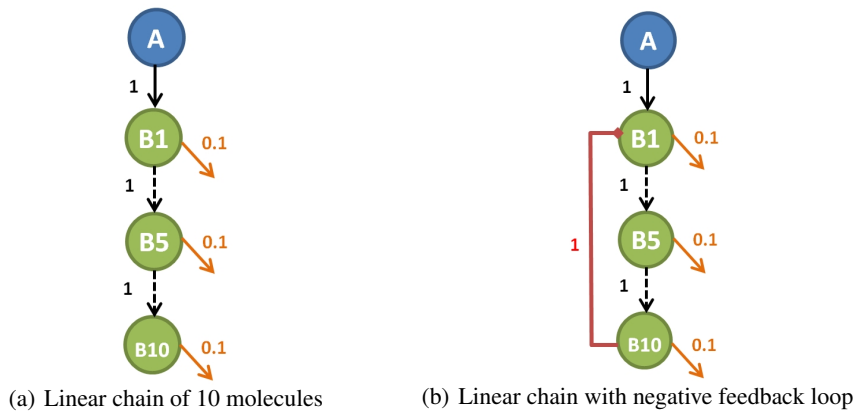
(c) Differences between the steady state probabilities (PRISM verification) and steady state values (ODE simulation) for the models in Fig. 15(a)-15(b)

Fig. 15. Influence of pathway length on differences between PRISM verification and ODE simulation (NF: negative feedback)

4. Fisher, J., Henzinger, T.A.: Executable cell biology. Nature Biotechnology **25**(11) (2007) 1239–1249
5. Sadot, A., Fisher, J., Barak, D., Admanit, Y., Stern, M.J., Jane Albert Hubbard, E., Harel, D.: Toward verified biological models. IEEE/ACM Transactions on Computational Biology Bioinformatics **5**(2) (2008) 223–234
6. Fisher, J., Piterman, N.: The executable pathway to biological networks. Briefings in Functional Genomics and Proteomics **9**(1) (2010) 79–92
7. Kwiatkowska, M.Z., Norman, G., Parker, D.: Stochastic model checking. In: Formal Methods for the Design of Computer, Communication and Software Systems: Performance Evaluation. Volume 4486 of LNCS., Springer (2007) 220–270
8. Baier, C., Katoen, J.P.: Principles in Model Checking. MIT Press (2008)
9. Hart, S., Sharir, M., Pnueli, A.: Termination of probabilistic concurrent programs. ACM Transactions on Programming Languages and Systems **5**(3) (1983) 356–380
10. Kwiatkowska, M.Z., Norman, G., Parker, D.: Quantitative Verification Techniques for Biological Processes. In: Algorithmic Bioprocesses. Springer (2009) 391–409

11. Kwiatkowska, M.Z., Norman, G., Parker, D.: Probabilistic Model Checking for Systems Biology. In: Symbolic Systems Biology. Jones and Bartlett (2010) 31–59
12. Aldridge, B.B., Haller, G., Sorger, P.K., Lauffenburger, D.A.: Direct lyapunov exponent analysis enables parametric study of transient signalling governing cell behaviour. Systems Biology **153**(6) (2006) 425–432
13. Yu, J., Ustach, C., Kim, H.R.: Platelet-derived growth factor signaling and human cancer. Journal of Biochemistry and Molecular Biology **36**(1) (2003) 49–59
14. Golub, T.R., Barker, G.F., Lovett, M., Gilliland, D.G.: Fusion of PDGF receptor beta to a novel ets-like gene, tel, in chronic myelomonocytic leukemia with t(5;12) chromosomal translocation. Cell **77**(2) (1994) 307–316
15. Anan, K., Morisaki, T., Katano, M., Ikubo, A., Kitsuki, H., Uchiyama, A., Kuroki, S., Tanaka, M., Torisu, M.: Vascular endothelial growth factor and platelet-derived growth factor are potential angiogenic and metastatic factors in human breast cancer. Surgery **119**(3) (1996) 333–339
16. Cools, J., DeAngelo, D.J., Gotlib, J., Stover, E.H., Legare, R.D., Cortes, J., Kutok, J., Clark, J., Galinsky, I., Griffin, J.D., Cross, N.C., Tefferi, A., Malone, J., Alam, R., Schrier, S.L., Schmid, J., Rose, M., Vandenberghe, P., Verhoef, G., Boogaerts, M., Wlodarska, I., Kantarjian, H., Marynen, P., Coutre, S.E., Stone, R., Gilliland, D.G.: A tyrosine kinase created by fusion of the PDGFRA and FIP1L1 genes as a therapeutic target of imatinib in idiopathic hypereosinophilic syndrome. The New England journal of medicine **348**(13) (2003) 1201–1214
17. Miettinen, M., Lasota, J.: Gastrointestinal stromal tumors: Pathology and prognosis at different sites. Seminars in Diagnostic Pathology **23**(2) (2006) 70–83
18. Calder, M., Vyshemirsky, V., Gilbert, D., Orton, R.: Analysis of signalling pathways using continuous time Markov chains. Transactions on Computational Systems Biology VI **4220** (2006) 44–67
19. Kwiatkowska, M.Z., Norman, G., Parker, D.: Using probabilistic model checking in systems biology. SIGMETRICS Performance Evaluation Review **35** (2008) 14–21
20. Pronk, T., de Vink, E., Bosnacki, D., Breit, T.: Stochastic modeling of codon bias with PRISM. In: Proc. 3rd Workshop on Methods and Tools for Coordinating Concurrent, Distributed and Mobile Systems, Computer Science Department, University of Cyprus, Nicosia (2007)
21. Bosnacki, D., Pronk, T., de Vink, E.: In silico modelling and analysis of ribosome kinetics and aa-tRNA competition. Transactions on Computational Systems Biology XI **5750** (2009) 69–89
22. Heath, J., Kwiatkowska, M., Norman, G., Parker, D., Tymchyshyn, O.: Probabilistic model checking of complex biological pathways. Theoretical Computer Science **319**(3) (2008) 239–257
23. Owens, N., Timmis, J., Greensted, A., Tyrrell, A.: Modelling the Tunability of early T cell signalling events. In: Proc. 7th Conference on Artificial Immune Systems. Volume 5132 of LNCS. (2008) 12–23
24. Jha, S.K., Clarke, E.M., Langmead, C.J., Legay, A., Platzer, A., Zuliani, P.: A bayesian approach to model checking biological systems. In: Proc. 7th Conference on Computational Methods in Systems Biology. Volume 5688 of LNCS., Springer (2009) 218–234
25. Liò, P., Merelli, E., Paoletti, N.: Multiple verification in computational modeling of bone pathologies. In: Proc. 3rd International Workshop on Computational Models for Cell Processes. Volume 68 of EPTCS. (2011) 82–96
26. Zhang, R., Shah, M.V., Yang, J., Nyland, S.B., Liu, X., Yun, J.K., Albert, R., Loughran, T.P.: Network model of survival signaling in large granular lymphocyte leukemia. Proceedings of the National Academy of Sciences of the United States of America **105**(42) (2008) 16308–16313

27. Wang, C.C., Cirit, M., Haugh, J.M.: PI3K-dependent cross-talk interactions converge with Ras as quantifiable inputs integrated by Erk. Molecular Systems Biology **5** (2009) 246

28. Aziz, A., Sanwal, K., Singhal, V., Brayton, R.: Model checking continuous time markov chains. ACM Transactions on Computational Logic **1**(1) (2000) 162–170

29. Baier, C., Hermanns, H., Haverkort, B.R., Katoen, J.P.: Model-checking algorithms for continuous-time markov chains. IEEE Transactions on Software Engineering **29**(6) (2003) 524–541

30. Krauss, G.: Biochemistry of Signal Transduction and Regulation. Wiley-VCH, Weinheim (2008)

31. Bhalla, U.: Understanding complex signaling networks through models and metaphor. Progress in Biophysics & Molecular Biology **81**(1) (2003) 45–65

32. Heldin, C.H., Westermarkt, B., Wasteson, A.: Platelet-derived growth factor. Molecular and Celluar Endocrinology **39**(3) (1985) 169–187

33. Tallquist, M., Kazlauskas, A.: PDGF signaling in cells and mice. Cytokine & Growth Factor Review **15**(4) (2004) 205–213

34. Yokote, K., Hellman, U., Ekman, S., Saito, Y., Roennstrand, L., Saito, Y., Heldin, C.H., Mori, S.: Identification of Tyr-762 in the platelet-derived growth factor alpha-receptor as the binding site for Crk proteins. Oncogene **16**(10) (1998) 1229–1239

35. Gutiérrez-Uzquiza, A., Arechederra, M., Molina, I., Banos, R., Maia, V., Benito, M., Guerrero, C., Porras, A.: C3G down-regulates p38 MAPK activity in response to stress by Rap-1 independent mechanisms: involvement in cell death. Cellualar Signalling **22**(3) (2010) 533–542

36. Valgeirsdóttir, S., Paukku, K., Silvennoinen, O., Heldin, C.H., Claesson-Welsh, L.: Activation of Stat5 by platelet-derived growth factor (PDGF) is dependent on phosphorylation sites in PDGF beta-receptor juxtamembrane and kinase insert domains. Oncogene **16**(4) (1998) 505–515

37. Joazeiro, C.A., Wing, S.S., Huang, H., Leverson, J.D., Hunter, T., Liu, Y.C.: The tyrosine kinase negative regulator c-Cbl as a RING-type, E2-dependent ubiquitin-protein ligase. Science **286**(5438) (1999) 309–312

38. Hanahan, D., Weinberg, R.A.: Hallmarks of cancer: The next generation. Cell **144**(5) (2011) 646–674

39. Yu, J., Gutkind, J.S., Mahadevan, D., Li, W., Meyers, K.A., Pierce, J.H., Heidaran, M.A.: Biological function of PDGF-induced PI-3 kinase activity: its role in alpha PDGF receptor-mediated mitogenic signaling. Journal of Cell Biology **127**(2) (1994) 479–487

40. Bazenet, C.E., Gelderloos, J.A., Kazlauskas, A.: Phosphorylation of tyrosine 720 in the platelet-derived growth factor alpha receptor is required for binding of Grb2 and SHP-2 but not for activation of Ras or cell proliferation. Molecular and Cellular Biology **16**(12) (1996) 6926–6936

41. Reddi, A.L., Ying, G., Duan, L., Chen, G., Dimri, M., Douillard, P., Druker, B.J., Naramura, M., Band, V., Band, H.: Binding of Cbl to a phospholipase Cgamma1-docking site on platelet-derived growth factor receptor beta provides a dual mechanism of negative regulation. Journal of Biological Chemistry **282**(40) (2007) 29336–2947

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

43. Witt, J., Barisic, S., Schumann, E., Allgöwer, F., Sawodny, O., Sauter, T., Kulms, D.: Mechanism of PP2A-mediated IKKβ dephosphorylation: a systems biological approach. BMC Systems Biology **3** (2009) 71

44. Schmidt, H., Jirstrand, M.: Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology. Bioinformatics **22**(4) (2006) 514–515

45. Saltelli, A., Tarantola, S., Chan, K.P.S.: A quantitative model-independent method for global sensitivity analysis of model output. Technometrics **41**(1) (1999) 39–56

46. Davies, S., Reddy, H., Caivano, M., Cohen, P.: Specificity and mechanism of action of some commonly used protein kinase inhibitors. Biochemical Journal **351**(Pt 1) (2000) 95–105

47. Stewart, W.J.: Introduction to the Numerical Solution of Markov Chains. Princeton University Press (1994)

48. Tenenbaum, M., Pollard, H.: Ordinary Differential Equations: An Elementary Textbook for Students of Mathematics, Engineering, and the Science. Harper & Row (1985)

49. Kwiatkowska, M.Z., Norman, G., Parker, D., Tymchyshyn, O., Heath, J., Gaffney, E.: Simulation and verification for computational modelling of signalling pathways. In: Proc. 38th Winter Simulation Conference. (2006) 1666–1674

50. Han, T., Katoen, J.P., Mereacre, A.: Approximate parameter synthesis for probabilistic time-bounded reachability. In: Proc. 29 IEEE Real-Time Systems Symposium, IEEE Computer Society (2008) 173–182

51. Hahn, E.M., Hermanns, H., Zhang, L.: Probabilistic reachability for parametric Markov models. In: Proc. 16th Spin Workshop on Model Checking Software. Volume 5578 of LNCS., Springer (2009) 88–106

52. Andreychenko, A., Mikeev, L., Spieler, D., Wolf, V.: Parameter identification for Markov models of biochemical reactions. In: Proc. 23rd Conference on Computer Aided Verification. Volume 6806 of LNCS., Springer (2011) 83–98