A Study of the PDGF Signaling Pathway with PRISM

Qixia Yuan

Computer Science and Communications
University of Luxembourg, Luxembourg

School of Computer Science and Technology
Shandong University, China

Jun Pang

Computer Science and Communications
University of Luxembourg, Luxembourg

Sjouke Mauw

Computer Science and Communications
University of Luxembourg, Luxembourg

Panuwat Trairatphisan

Life Sciences Research Unit
University of Luxembourg, Luxembourg

Monique Wiesinger

Life Sciences Research Unit
University of Luxembourg, Luxembourg

Thomas Sauter

Life Sciences Research Unit
University of Luxembourg, Luxembourg

In this paper, we apply the probabilistic 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. We show that quantitative verification can yield a better understanding of the PDGF signaling pathway.

1 Introduction

Biological systems consist of separated 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 from a systematic view. Due to the similarity between biological systems and complex distributed/reactive systems studied in computer science [30], modeling and analyzing techniques developed in the field of formal methods can be applied to biological systems as well [5]. Due to efficient verification techniques, formal methods can analyze large systems exhibiting complex behaviors – this process is typically supported by automatic computer tools. Clearly, this 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 [10, 31, 11]).

In this paper, we explore the usage of model checkingfor 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 model checking approach, first introduced by Hart, Sharir and Pnueli [17], 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 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. Probabilistic verification has gained notable success in analyzing probabilistic systems including biological signaling pathways (e.g., see [25, 26]).

We use the probabilistic model checker PRISM [26] to yield a better understanding of the PDGF (Platelet-Derived Growth Factor) signaling pathway. PDGF, described approximately 30 years ago as a major mitogenic component of whole blood [37], 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 [37]. 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 [12, 1, 37, 8, 28]. 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 Section 3). In our study, there are three main goals: (1) analyze the dynamics of PDGF induced signaling, (2) analyze the influence of the crosstalk reactions and (3) analyze the importance of individual reaction 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 quantitative verification using PRISM.

Related work. PRISM has been used for analyzing a variety of biological systems. In [18], PRISM is used 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. In [24], PRISM is used to study the MAPK cascade. The authors explain how the biological pathway can be modeled in PRISM and how this enables the analysis of a rich set of quantitative properties.

Jha *et al.* [20] present the first algorithm for performing statistical model checking using Bayesian sequential hypothesis testing and test the performance of the algorithm on the FGF signaling pathway and several others. Schwarick and Heiner [32] give an interval decision diagram based approach to CSL model checking. They aim for efficient verification of biochemical network models and apply their approach to the MAPK cascade. Draábik *et al.* [9] apply the technique modular verification to analyzing the *lac* operon regulation.

Apart from probabilistic (or stochastic) model checking, other formal techniques have been used for studying biological systems in recent years as well. For instance, the process algebra bio-PEPA is used to model and analyze the NF-kappaB pathway [7]; Petri nets are used for quantitative predictive modeling of *C.elegans* vulval development [6]; and the rule-based modeling language Kappa is used to model epigenetic information maintenance [23]. More related work, especially case studies of applying formal methods in systems biology, can be found in [31, 11].

Outline of the paper. In Section 2, we give an overview of probabilistic model checking and the tool PRISM. Section 3 describes the PDGF signaling pathway. In Section 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 Section 5. Finally, we draw conclusion of this paper and discuss some future work in Section 6.

2 Probabilistic Model Checking and PRISM

We briefly introduce probabilistic verification and the probabilistic model checker – PRISM [26].

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 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.

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 2.1 *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.

Corresponding to CTMC models, we use Continuous Stochastic Logic (CSL) to specify properties of built models. CSL, originally introduced by Aziz *et al.* [2], provides a powerful means to specify both path-based and traditional state-based performance measures on CTMCs.

Definition 2.2 *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, $\sim \in \{<, \leq, \geq, >\}$, $p \in [0, 1]$, I is an interval of $\mathbb{R}_{\geq 0}$ and $r, t \in \mathbb{R}_{\geq 0}$.

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 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 $\sim p$. In this paper, we use a single type of 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. More precisely, $S_{\sim p}[\psi]$ asserts that the steady-state probability of being in a state satisfying ψ meets the bound $\sim p$.

2.2   The model checker PRISM

PRISM [26] 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 is composed 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:

$$[a]\ g \to r : u;$$

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 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 *primed* value with respect to their *unprimed* value.

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; *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 to probabilistic model checking, 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

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) [22]. For further details on cell signaling see, for example, [4, 22].

Platelet-Derived Growth Factor (PDGF), described approximately 30 years ago as a major mitogenic component of whole blood [37] 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 [19]. PDGFR is a receptor tyrosine kinase, which in general transfer upstream signals to many downstream signaling pathways by phosphorylation. Up to now, five PDGF ligands 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 [35], which will activate Rap1 to positively influence c-Raf in the MAPK pathway [13], for Signal Transducer and Activator of Transcription (STAT), which might regulate the signal in parallel to the JAK-STAT pathway [34] 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 [21]. Based on intensive literature review, we have built a PDGF signal tranduction model in ODE format. This full model comprises 35 molecules. The essential part of the PDGF signaling pathway which are 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, are extracted for analysis in PRISM (as shown in Figure 1). This simplified model contains only 17 molecules, and the analysis in this paper focuses on this simplified model.

Figure 1 describes how signals are transduced in the PDGF pathway by activating or deactivating specific downstream pathways signaling molecules respectively. In this model, there are three inputs which are 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 signal and keep the cell survive at a basal level. There are three different types of arrows in the network: blue arrows, green arrows and red arrows. The
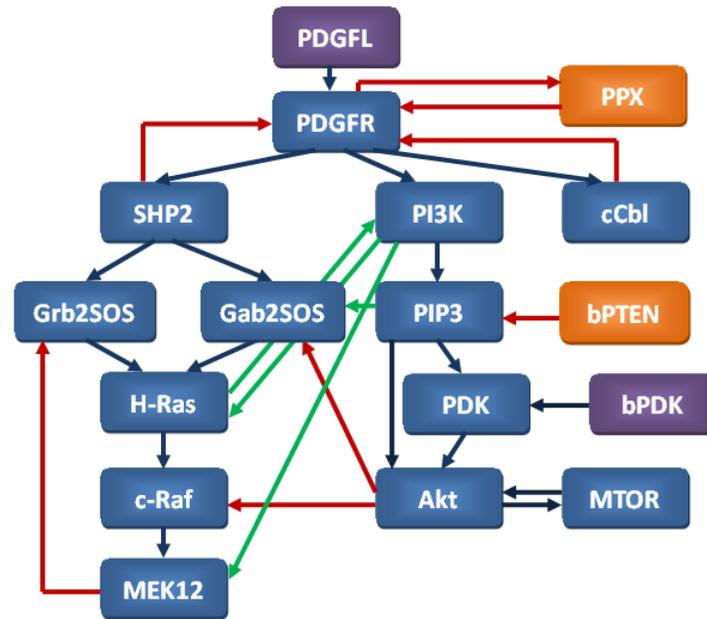
Figure 1: The extracted PDGF signaling pathway (blue arrows: main pathway, green arrows: positive crosstalk, red arrows: negative regulatory)

blue arrows represent the main activating interactions, indicating the two main downstream signaling pathways in the network. The MAPK pathway covers `SHP2`, `Grb2SOS`, `GabSOS`, `Ras`, `c-Raf` and `MEK12`. These molecules play a major role in the cellular proliferative circuit [16]. The PI3K/Akt pathway covers the molecules `PI3K`, `PIP3`, `PDK`, `bPTEN`, `bPDK` and `Akt`, also being important in viability circuit [16]. 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`. Then, 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 (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 [36, 3, 29]. 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.

Figure 2 contains the list of model reactions. Each node (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`.

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

Figure 2: List of biochemical reactions in the PDGF signaling pathway ODE model

For instance, reaction 2 describes that an active PPX gives a negative feedback to PDGFR, making it inactive. The parameters used in the model (all starting with letter k) in each reaction are the relative reaction rate constants normalized to the maximum value in each set of reactions. In this case, the values of parameters are ranging from 0 to 1. The initial set of parameters' values used in this model are derived from literature (e.g., [12, 1, 37, 8, 28]) and our own experimental observations of PDGFRα mutants. Further parameter optimization based on experimental data is ongoing, especially focusing on the unknown crosstalk reaction strength. However, in this paper, the initial set of parameters has been used. This might not fully correlate to the actual biological system. Therefore, we focus in our analysis on the more qualitative aspects already conserved by the network structure.

4 Modelling and Property Specifications in PRISM

4.1 PRISM model

We now describe how to build a PRISM model for the PDGF signaling pathway as presented in the previous section. Our model represents a single instance of the signaling pathway, meaning there can be at most one element of each molecule. In the single instance model, the molecules' steady state, which is expressed as a probability in PRISM, corresponds to the molecule density in the real-life experiments.

Each of the nodes (or molecules) of the pathway except for the PDGFL in Figure 1 is represented by a separate PRISM module. Since PDGFL, bPTEN and bPDK remain the same after the reactions they are involved, we set them as const boolean values true. Figures 3(a) and 3(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,

```
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

Figure 3: PRISM modules and rewards

since 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 Figure 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 Figures 3(a) and 3(b), there are commands with label `bkoff1` in both of the modules `PDGFR` and `PPX`. The two commands are used to model the reaction (2) in Figure 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. Totally we have modeled all the 17 molecules in 14 PRISM modules (`PDGFL`, `bPTEN` and `bPDK` are modeled as a constant).

As mentioned in Section 2.2, PRISM models can be augmented with information about rewards. We construct rewards to calculate the time for a molecule being active. Figure 3(c) shows the rewards for calculating the active state of `PDGFR`. Each time `PDGFR` is in 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

There are three main goals for this study: (1) analyze the dynamics of PDGF induced signaling, (2) analyze the influence of the crosstalk reactions as defined in Section 3, and (3) analyze the importance of

Model	States	Transitions
`WildType`	589,824	7,145,472
`SHP2Mutant`	36,864	373,248
`PI3KMutant`	589,824	7,096,320
`cCb1Mutant`	294,912	3,357,696
`RasPI3KMutant`	589,824	7,047,168
`Aktc-RafMutant`	786,423	10,264,576

Table 1: Model statistics in PRISM

individual reaction 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 examine 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 probabilties of molecule `MEK12` and `Akt` change when a certain reaction is removed.

Below we list 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.

- $P_{=?}[F^{[t,t]}\text{PDGFR} = 1]$
  The probability that the molecule `PDGFR` is active at time instant t.

- $R_{=?}^{\{\text{``PDGFRactive''}\}}[C <= t]$
  The expected time of `PDGFR` being active by time t. It refers to the reward structure "PDGFRactive" as defined in Figure 3(c).

- $S_{=?}[\text{PDGFR} = 1]$
  The long-run probability that `PDGFR` is active.

5 Results

We use PRISM to construct the PDGF model described in Section 4.1 and analyze the set of properties listed in Section 4.2. As described in Section 3,

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 Figure 2 are included. Subsequently, we obtain the mutant models by removing the mutant arrows (as mentioned in Section 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 pointed to `SHP2`. The third and the fourth ones are `PI3KMutant` and `cCblMutant`. Following the same process, we remove the mutant arrows pointed to `PI3K` and `cCbl` separately in the last two models. The three removed mutant arrows correspond to the reactions (7), (8) and (6) in Figure 2. The size of the models in PRISM are shown in Table 1 (Table 1 also shows the size of models `RasPI3KMutant` and `Aktc-RafMutant` which will be discussed later). For each model, we compute the probability of each molecule being active at time instance t, which is summarized in Figure 4.
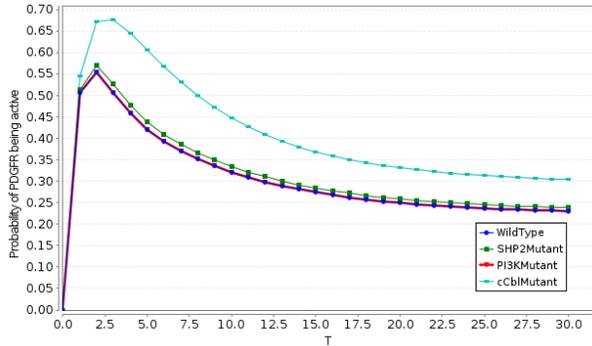
(a) Probability of being active (`PDGFR`)
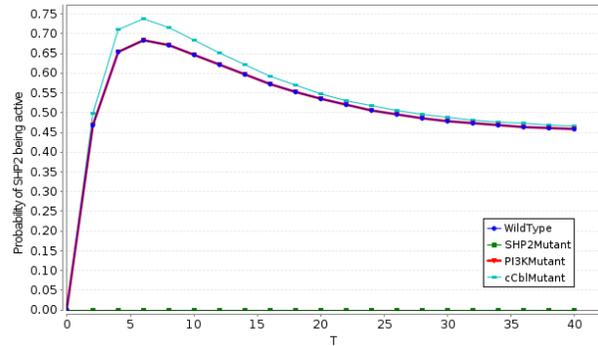
(b) Probability of being active (`SHP2`)
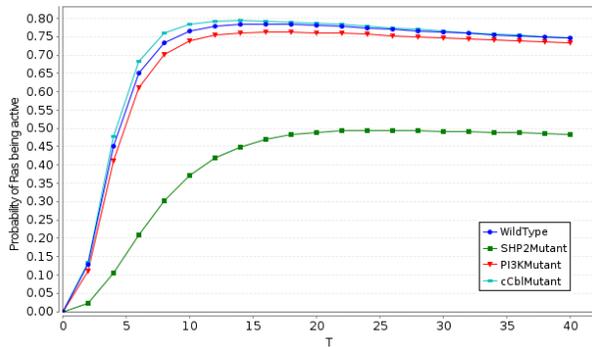
(c) Probability of being active (`Ras`)
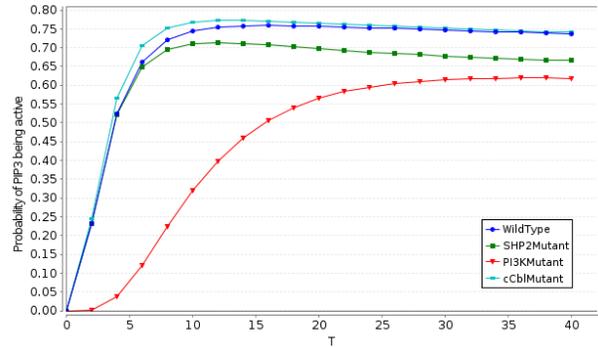
(d) Probability of being active (`PIP3`)
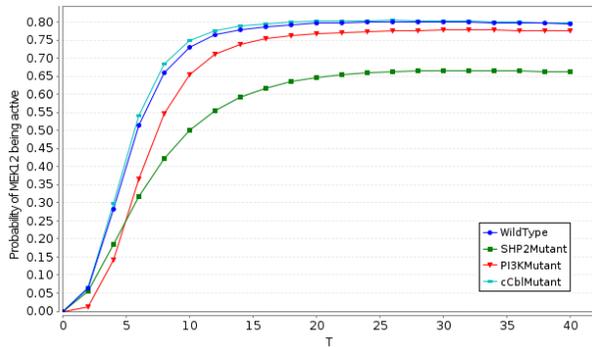
(e) Probability of being active (`MEK12`)
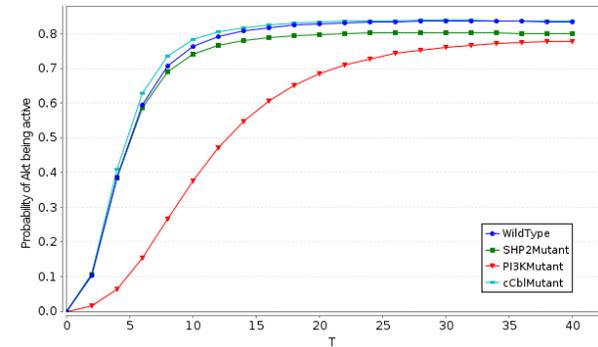
(f) Probability of being active (`Akt`)

Figure 4: Probabilities of molecules being active

Figure 4 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 Figure 4(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 Figure 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. The red curve in Figure 4(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 in the full PDGF signaling pathway model there is another molecule PLCg giving a fast positive crosstalk to PI3K, which is removed in our simplified PDGF model here.

If we focus on the light blue curve of Figures 4(a) - 4(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 Figures 4(b), 4(c) and 4(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 Figures 4(d) and 4(f) show that the influence of PI3K to the molecules in the PI3K/Akt pathway, which is similar to that of SHP2 to the MAPK pathway.

We analyze the long-run probability of the molecules being active after PDGFL stimulation (shown in Table 2). The result demonstrates the activities of each molecule in this model at steady state.

Molecule	Probability	Molecule	Probability
PDGFR	0.22	Grb2SOS	0.55
SHP2	0.45	Ras	0.72
GabSOS	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	0.00

Table 2: Steady state probabilities of molecules

PRISM supports reward properties (see Section 4). Figure 5 shows the expected time of six molecules being active by time instant t. These six molecules are the same as in Figure 4. All the six curves tend to be linear after time instant 12, which shows that the state of the 6 molecules start to be in a steady state after time instant 12.

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. Totally, 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 re-
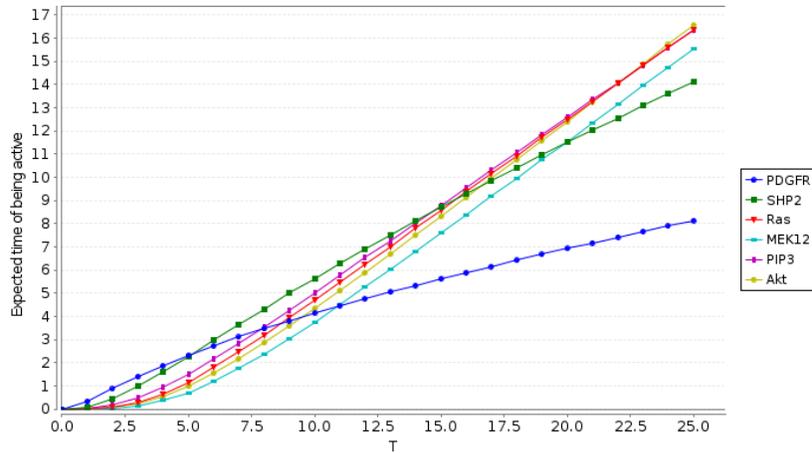
Figure 5: Expected time of being active by time *t*
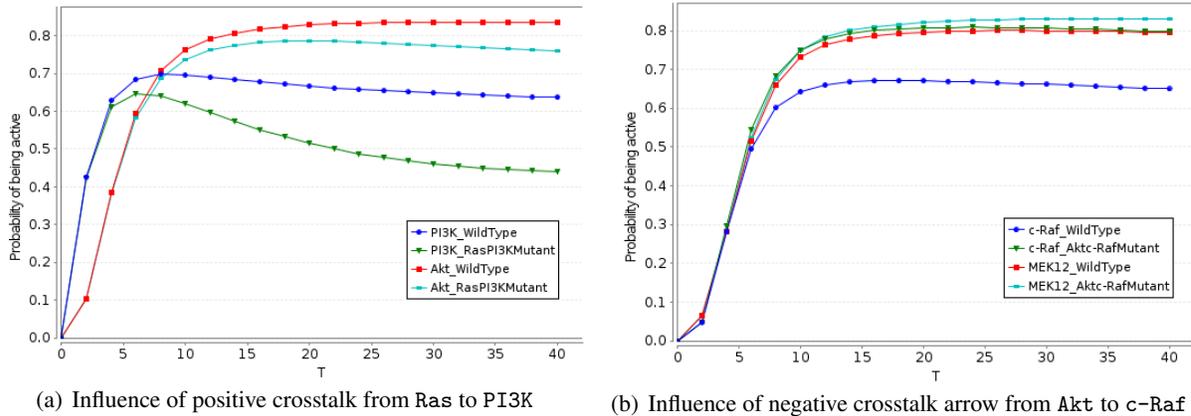
(a) Influence of positive crosstalk from `Ras` to `PI3K`

(b) Influence of negative crosstalk arrow from `Akt` to `c-Raf`

Figure 6: Influence of crosstalks

moved. 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 analyze the influences of the two crosstalks by comparing the `WildType` model to the two mutant models, respectively. The size of each PRISM model is summarized in Table 1.

Figure 6 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 Figure 6(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 Figure 6(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 Figure 6(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 to `Akt` is smaller than that to `PI3K`. Just like the situation in Figures 4(b), 4(c) and 4(e), the more downstream the molecules are,
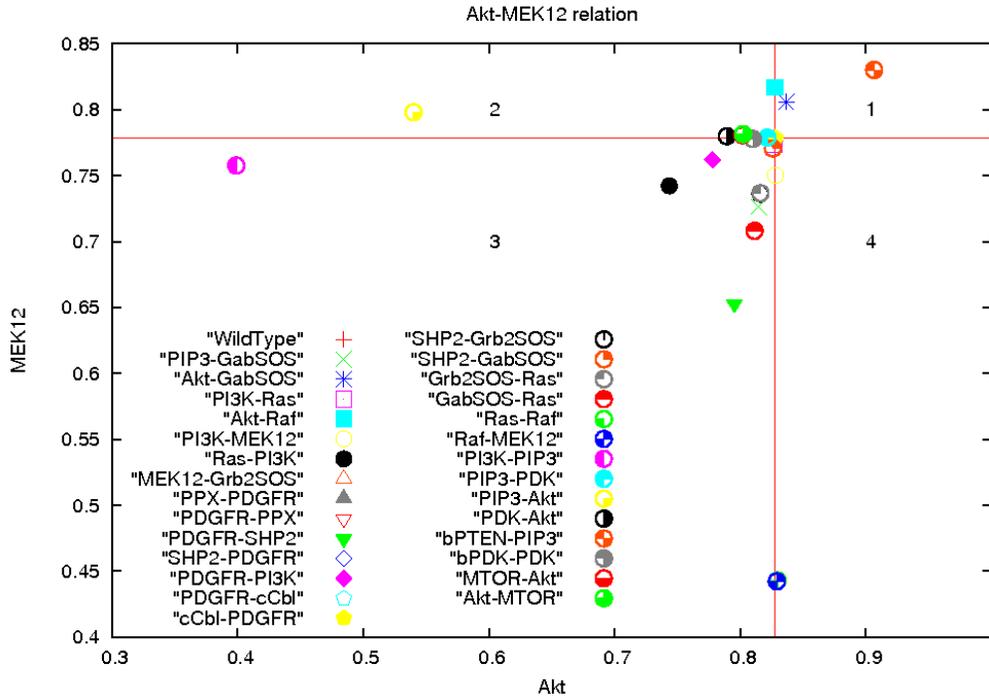
Figure 7: Steady state probabilities of `Akt` and `MEK12` in different models

the smaller the influence becomes. The results in Figure 6(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`.

For the last goal to analyze the importance of individual reaction on downstream signaling molecules, we compute the steady state probabilities of `Akt` and `MEK12` in 31 different models, each of which is obtained by removing one reaction from the `wildtype` model. For example, 'PIP3-GabSOS' model is obtained by removing reaction No.20 in Figure 2). The results are shown in Figure 7. We draw two lines in the figure to divide the axis into four areas, letting the dot of the wildtype model lie at the cross of the two lines. Apparently, the dots in different areas show that the removed reactions can bring different influence to the steady state probabilities of `Akt` and `MEK12`. The reactions (corresponding to the dots) in area 1 can decrease the steady state probabilities of both `Akt` and `MEK12`, while those in area 3 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 impact on the steady state probability of `Akt`. Similarly, the reactions on the vertical line have little impact on `MEK12`. These observations have biological implications. `Akt` and `MEK12` are downstream molecules in the signal transduction process that regulate different cellular functions. Signal from `Akt` keeps the cells to survive from apoptosis and signal form `MEK12` regulates the cells growth and proliferation. In cancer, both of these two main pathways (see Figure 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.

(a) Probability of being active (`PDGFR`) (b) Probability of being active (`Ras`)
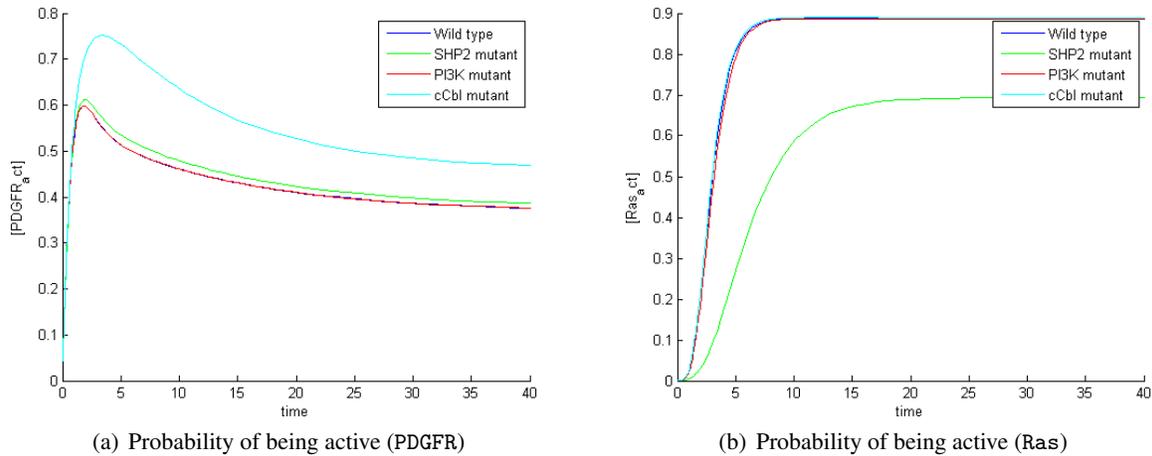
Figure 8: Results from ODE simulations

6   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 and extracted it for the analysis in a probabilistic model checker. We focused on a simplified model of the PDGF signaling pathway which contains 17 molecules while the full model has 35 molecules. We have analyzed the dynamics of PDGF induced signaling, the influence of crosstalk reactions in the signaling pathway and the importance of individual reaction on downstream signaling molecules. Our experiment results show that quantitative verification can provide us a better understanding of the PDGF signaling pathway, especially the result discussed in the end of Sect. 5 potentially can give rise to better behavior prediction of the pathway. (The PRISM model and property specifications can be found at `satoss.uni.lu/jun/models/PDGF.zip`.)

Except for model checking, we have also analyzed the model using ODE simulations. The results for the two molecules, `PDGFR` and `Ras`, are shown in Figure 8(a) and Figure 8(b) respectively. We can see that probabilistic verification and ODE simulations produce comparable results: the shape of the curves are similar while the maximum probabilities and the steady state probabilities differ a little. The reason of the difference might be that the approach based on verification leads to more accurate results when the number of molecules is small as indicated in [27]. In general, ODE analysis perform well for larger number of molecules, but not for small numbers, as time trajectories for average concentrations can be misleading if numbers of molecules small. We plan to have a detailed comparison between our analysis using probabilisitic verification techniques and the analysis using ODE simulations. As indicated in the end of Section 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., [15, 14]) can help to synthesize values which are consistent with the lab experimental results.

References

[1] K. Anan, T. Morisaki, M. Katano, A. Ikubo, H. Kitsuki, A. Uchiyama, S. Kuroki, M. Tanaka & M. Torisu (1996): *Vascular endothelial growth factor and platelet-derived growth factor are potential angiogenic and metastatic factors in human breast cancer*. Surgery 119(3), pp. 333–339.

[2] A. Aziz, K. Sanwal, V. Singhal & R. Brayton (2000): *Model Checking Continuous Time Markov Chains*. ACM Transactions on Computational Logic 1(1), pp. 162–170.

[3] C. E. Bazenet, J. A. Gelderloos & A. Kazlauskas (1996): *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), pp. 6926–6936.

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

[5] N. Bonzanni, K. Anton Feenstra, W. J. Fokkink & E. Krepska (2009): *What Can Formal Methods Bring to Systems Biology?* In: Proc. 2nd World Congress on Formal Methods, LNCS 5850, Springer, pp. 16–22.

[6] N. Bonzanni, E. Krepska, K. Anton Feenstra, W. J. Fokkink, T. Kielmann, H. E. Bal & J. Heringa (2009): *Executing multicellular differentiation: quantitative predictive modelling of C.elegans vulval development*. Bioinformatics 25(16), pp. 2049–2056.

[7] F. Ciocchetta, A. Degasperi, J. K. Heath & J. Hillston (2010): *Modelling and Analysis of the NF-appaB Pathway in Bio-PEPA*. Transactions on Computational Systems Biology 12, pp. 229–262.

[8] J. Cools, D. J. DeAngelo, J. Gotlib, E. H. Stover, R. D. Legare, J. Cortes, J. Kutok, J. Clark, I. Galinsky, J. D. Griffin, N. C. Cross, A. Tefferi, J. Malone, R. Alam, S. L. Schrier, J. Schmid, M. Rose, P. Vandenberghe, G. Verhoef, M. Boogaerts, I. Wlodarska, H. Kantarjian, P. Marynen, S. E. Coutre, R. Stone & D. G. Gilliland (2003): *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), pp. 1201–1214.

[9] P. Drábik, A. Maggiolo-Schettini & P. Milazzo (2010): *Modular verification of interactive systems with an application to biology*. In: Proc. 1st Workshop on Interactions between Computer Science and Biology, ENTCS 268, Elsevier, pp. 61–75.

[10] J. Fisher & T. A. Henzinger (2007): *Executable cell biology*. Nature Biotechnology 25(11), pp. 1239–1249.

[11] J. Fisher & N. Piterman (2010): *The Executable Pathway to Biological Networks*. Briefings in Functional Genomics and Proteomics 9(1), pp. 79–92.

[12] T. R. Golub, G. F. Barker, M. Lovett & D. G. Gilliland (1994): *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), pp. 307–316.

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

[14] E. M. Hahn, H. Hermanns & L. Zhang (2009): *Probabilistic reachability for parametric Markov models*. In: Proc. 16th Spin Workshop on Model Checking Software, LNCS 5578, Springer, pp. 88–106.

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

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

[17] S. Hart, M. Sharir & A. Pnueli (1983): *Termination of probabilistic concurrent programs*. ACM Transactions on Programming Languages and Systems 5(3), pp. 356–380.

[18] J. Heath, M. Kwiatkowska, G. Norman, D. Parker & O. Tymchyshyn (2008): *Probabilistic model checking of complex biological pathways*. Theoretical Computer Science 319(3), pp. 239–257.

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

[20] S. K. Jha, E. M. Clarke, C. J. Langmead, A. Legay, A. Platzer & P. Zuliani (2009): *A Bayesian Approach to Model Checking Biological Systems*. In: Proc. 7th Conference on Computational Methods in Systems Biology, *LNCS* 5688, Springer, pp. 218–234.

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

[22] G. Krauss (2008): *Biochemistry of Signal Transduction and Regulation*. Wiley-VCH, Weinheim.

[23] J. Krivine, V. Danos & JA. Benecke (2009): *Modelling epigenetic information maintenance: a Kappa tutorial*. In: Proc. 21st Conference on Computer Aided Verification, *LNCS* 5643, Springer, pp. 17–32.

[24] M. Z. Kwiatkowska, G. Norman & D. Parker (2008): *Using probabilistic model checking in systems biology*. SIGMETRICS Performance Evaluation Review 35, pp. 14–21.

[25] M. Z. Kwiatkowska, G. Norman & D. Parker (2009): *Algorithmic Bioprocesses*, chapter Quantitative Verification Techniques for Biological Processes, pp. 391–409. Springer.

[26] M. Z. Kwiatkowska, G. Norman & D. Parker (2010): *Symbolic Systems Biology*, chapter Probabilistic Model Checking for Systems Biology, pp. 31–59. Jones and Bartlett.

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

[28] M. Miettinen & J. Lasota (2006): *Gastrointestinal stromal tumors: Pathology and prognosis at different sites*. Seminars in Diagnostic Pathology 23(2), pp. 70–83.

[29] A. L. Reddi, G. Ying, L. Duan, G. Chen, M. Dimri, P. Douillard, B. J. Druker, M. Naramura, V. Band & H. Band (2007): *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), pp. 29336–2947.

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

[31] A. Sadot, J. Fisher, D. Barak, Y. Admanit, M. J. Stern, E. Jane Albert Hubbard & D. Harel (2008): *Toward Verified Biological Models*. IEEE/ACM Transactions on Computational Biology Bioinformatics 5(2), pp. 223–234.

[32] M. Schwarick & M. Heiner (2009): *CSL Model Checking of Biochemical Networks with Interval Decision Diagram*. In: Proc. 7th Conference on Computational Methods in Systems Biology, *LNCS* 5688, Springer, pp. 296–312.

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

[34] S. Valgeirsdóttir, K. Paukku, O. Silvennoinen, C. H. Heldin & L. Claesson-Welsh (1998): *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), pp. 505–515.

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

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

[37] J. Yu, C. Ustach & H.-R. Kim (2003): *Platelet-derived growth factor signaling and human cancer*. Journal of Biochemistry and Molecular Biology 36(1), pp. 49–59.