Probabilistic model checking of complex biological pathways*

J. Heath[1], M. Kwiatkowska[2], G. Norman[2], D. Parker[2], and O. Tymchyshyn[2]

[1]School of Biosciences [2]School of Computer Science
University of Birmingham, Birmingham, B15 2TT, UK

Abstract. Probabilistic model checking is a formal verification technique that has been successfully applied to the analysis of systems from a broad range of domains, including security and communication protocols, distributed algorithms and power management. In this paper we illustrate its applicability to a complex biological system: the FGF (Fibroblast Growth Factor) signalling pathway. We give a detailed description of how this case study can be modelled in the probabilistic model checker PRISM, discussing some of the issues that arise in doing so, and show how we can thus examine a rich selection of quantitative properties of this model. We present experimental results for the case study under several different scenarios and provide a detailed analysis, illustrating how this approach can be used to yield a better understanding of the dynamics of the pathway.

1 Introduction

There has been considerable success recently in adapting approaches from computer science to the analysis of biological systems and, in particular, biochemical pathways. The majority of this work has relied on simulation-based techniques developed for discrete stochastic models [7]. These allow modelling of the evolution of individual molecules, whose rates of interaction are controlled by exponential distributions. The principal alternative modelling paradigm, using ordinary differential equations, differs in that it reasons about how the average concentrations of the molecules evolve over time. In this paper, as in [4, 3], we adopt the stochastic modelling approach, but employ methods which allow calculation of *exact* quantitative measures of the model under study.

We use probabilistic model checking [19] and the probabilistic model checker PRISM [9, 14] as a framework for the modelling and analysis of biological pathways. This approach is motivated by the success of previous work which has demonstrated the applicability of these techniques to the analysis of a wide variety of complex systems [11]. One benefit of this is the ability to employ the existing efficient implementations and tool support developed in this area. Additionally, we enjoy the advantages of model checking, for example, the use of

* Supported in part by EPSRC grants GR/S72023/01, GR/S11107 and GR/S46727 and Microsoft Research Cambridge contract MRL 2005-44.

both a formal model and specification of the system under study and the fact that the approach is exhaustive, that is, all possible behaviours of the system are analysed. Our intention is that the methods in this paper should be used in conjunction with the classical simulation and differential equation based approaches to provide greater insight into the complex interactions of biological pathways. This paper provides a detailed illustration of the applicability of probabilistic model checking to this domain through the analysis of a complex biological pathway called FGF (Fibroblast Growth Factor).

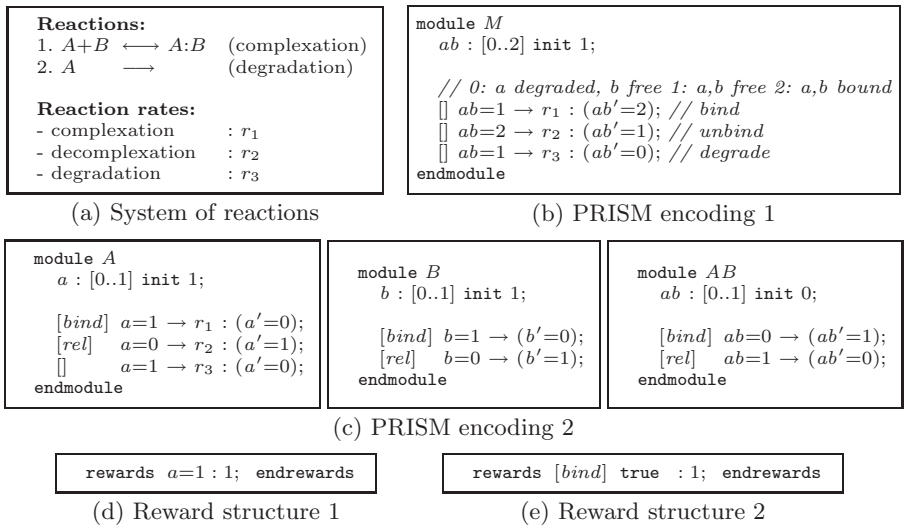
Related Work. The closest approach to that presented here is [4], where the probabilistic model checker PRISM is used to model the RKIP inhibited ERK pathway. The main difference is that in [4] the authors consider a "population" based approach to modelling using approximate techniques where concentrations are modelled by discrete abstract quantities. In addition, here we demonstrate how a larger class of temporal properties including reward-based measures are applicable to the study biological systems. Also related to the RKIP inhibited ERK pathway is [3], where it is demonstrated how the stochastic process algebra PEPA [8] can be used to model biological systems. The stochastic π-calculus [15] has been proposed as a model language for biological systems [18, 16]; this approach has so far been used in conjunction with stochastic simulation, for example through the tools BioSpi [16] and SPiM [12].

In parallel with the development of the PRISM model of the FGF pathway presented in this paper, we have constructed a separate π-calculus model [22, 13] and applied stochastic simulation through BioSpi. Although currently these works focus on different aspects of the pathway, in the future we aim to use this complex case study as a basis for investigating the advantages of stochastic simulation and probabilistic model checking.

2 Probabilistic Model Checking and PRISM

Probabilistic model checking is a formal verification technique for the modelling and analysis of systems which exhibit stochastic behaviour. This technique is a variant of *model checking*, a well-established and widely used formal method for ascertaining the correctness of real-life systems. Model checking requires two inputs: a description of the system in some high-level modelling formalism (such as a Petri net or process algebra), and specification of one or more desired properties of that system in temporal logic (e.g. CTL or LTL). From these, one can construct a model of the system, typically a labelled state-transition system in which each state represents a possible configuration and the transitions represent the evolution of the system from one configuration to another over time. It is then possible to automatically verify whether or not each property is satisfied, based on a systematic and exhaustive exploration of the model.

In probabilistic model checking, the models are augmented with quantitative information regarding the likelihood that transitions occur and the times at which they do so. In practice, these models are typically Markov chains or



(a) System of reactions (b) PRISM encoding 1

(c) PRISM encoding 2

(d) Reward structure 1 (e) Reward structure 2

Fig. 1. Simple example and possible PRISM representations

Markov decision processes. In this paper, it suffices to consider *continuous-time Markov chains* (CTMCs), in which transitions between states are assigned (positive, real-valued) rates, which are interpreted as the rates of negative exponential distributions. The model is augmented with rewards associated with states and transitions. Rewards associated with states (*cumulated rewards*) are incremented in proportion to the time spent in the state, while rewards associated with transitions (*impulse rewards*) are incremented each time the transition is taken.

Properties of these models, while still expressed in temporal logic, are now quantitative in nature. For example, rather than verifying that "the protein always eventually degrades", we may ask "what is the probability that the protein eventually degrades?" or "what is the probability that the protein degrades within T hours?". Reward-based properties include "what is the expected energy dissipation within the first T time units?" and "what is the expected number of complexation reactions before relocation occurs?".

PRISM [9, 14] is a probabilistic checking tool developed at the University of Birmingham. Models are specified in a simple state-based language based on Reactive Modules. An extension of the temporal logic CSL [1, 2] is used to specify properties of CTMC models augmented with rewards. The tool employs state-of-the-art symbolic approaches using data structures based on binary decision diagrams [10]. Also of interest, the tool includes support for PEPA [8] and has recently been extended to allow for simulation-based analysis using Monte-Carlo methods and discrete event simulation. For further details, see [14].

3 Modelling a simple biological system in PRISM

We now illustrate PRISM's modelling and specification languages through an example: the simple set of biological reactions given in Figure 1(a). We con-

sider two proteins A and B which can undergo complexation with rate r_1 and decomplexation with rate r_2. In addition, A can degrade with rate r_3.

We give two alternative approaches for modelling these reactions in PRISM, shown in Figures 1(b) and 1(c), respectively. A model described in the PRISM language comprises a set of *modules*, the state of each being represented by a set of finite-ranging *variables*. In approach 1 (Figure 1(b)) we use a single module with one variable, representing the (three) possible states of the whole system (which are listed in the italicised comments in the figure). The behaviour of this module, i.e. the changes in states which it can undergo, is specified by a number of *guarded commands* of the form $[]\ g \rightarrow r : u$, with the interpretation that if the predicate (guard) g is true, then the system is updated according to u (where $x' = ...$ denotes how the value of variable x is changed). The rate at which this occurs is r, i.e. this is the value that will be attached to the corresponding transition in the underlying CTMC.

In approach 2 (Figure 1(c)) we represent the different possible forms that the proteins can take (A, B and A:B) as separate modules, each with a single variable taking value 0 or 1, representing its absence or presence, respectively. To model interactions where the state of several modules changes simultaneously, we use *synchronisation*, denoted by attaching action labels to guarded commands (placed inside the square brackets). For example, when the *bind* action occurs, variables a and b in modules A and B change from 1 to 0 and variable ab in module AB changes from 0 to 1. In this example, the rate of each combined transition is fully specified in module A and we have omitted the rates from the other modules. More precisely, PRISM assigns a rate of 1 to any command for which none is specified and computes the rate of a combined transition as the product of the rates for each command. Note that independent transitions, involving only a single module, can also be included, as shown by the modelling of degradation (which only involves A), by omitting the action label.

In general, a combination of the above two modelling approaches is used. In simple cases it is possible to use a single variable, but as the system becomes more complex the use of separate variables and synchronisation becomes more desirable. We will see this later in the paper.

Properties of CTMCs are specified in PRISM using an extension of the temporal logic CSL. We now give a number of examples for the model in Figure 1(c).

- *What is the probability that the protein A is bound to the protein B at time instant T?* ($\mathcal{P}_{=?}[\texttt{true}\ \mathcal{U}^{[T,T]}\ ab{=}1]$);
- *What is the probability that the protein A degrades before binding to the protein B?* ($\mathcal{P}_{=?}[ab{=}0\ \mathcal{U}\ (a{=}0 \wedge ab{=}0)]$);
- *During the first T time units, what is the expected time that the protein A spends free?* ($\mathcal{R}_{=?}[\mathcal{C}^{\leq T}]$, assuming a reward structure which associates reward 1 with states where the variable a equals 1 - see Figure 1(d));
- *What is the expected number of times that the proteins A and B bind before A degrades?* ($\mathcal{R}_{=?}[\mathcal{F}\ (a{=}0 \wedge ab{=}0)]$, assuming a reward of 1 is associated with any transition labelled by *bind* - see Figure 1(e)).

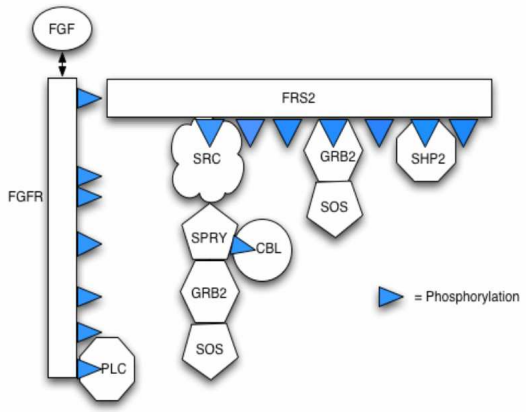


Fig. 2. Diagram showing the different possible bindings in the pathway

4 Case study: FGF

Fibroblast Growth Factors (FGF) are a family of proteins which play a key role in the process of cell signalling in a variety of contexts, for example wound healing. The mechanisms of the FGF signalling pathway are complex and not yet fully understood. In this section, we present a model of the pathway which is based on literature-derived information regarding the early stages of FGF signal propagation and which incorporates several features that have been reported to negatively regulate this propagation [6, 21, 5, 20].

Our model incorporates protein-protein interactions (including competition for partners), phosphorylation and dephosphorylation, protein complex relocation and protein complex degradation (via ubiquitin-mediated proteolysis). Figure 2 illustrates the different components in the pathway and their possible bindings. Below is a list of the reactions included in the model. Further details are provided in Figure 3.

1. An FGF ligand binds to an FGF receptor (FGFR) creating a complex of FGF and FGFR.
2. The existence of this FGF:FGFR dimer leads to phosphorylation of FGFR on two residues Y653 and Y654 in the activation loop of the receptor.
3. The dual Y653/654 form of the receptor leads to phosphorylation of other FGFR receptor residues: Y463, Y583, Y585, Y766 (in this model we only consider Y766 further).
4. and 5. The dual Y653/654 form of the receptor also leads to phosphorylation of the FGFR substrate FRS2, which binds to both the phosphorylated and dephosphorylated forms of the FGFR.
6. FRS2 can also be dephosphorylated by a phosphotase, denoted Shp2.
7. A number of effector proteins interact with the phosphorylated form of FRS2. In this model we include Src, Grb2:Sos and Shp2.
8. and 9. These are two methods of attenuating signal propagation by removal (i.e. relocation) of components. In step **8.** if Src is associated with the phos-

1. FGF binds to FGFR
FGF+FGFR ↔ FGFR:FGF (k_{on} = 5e+8M^{-1}s^{-1}, k_{off}=1e−1s^{-1})
2. Whilst FGFR:FGF exists
FGFR Y653 → FGFR Y653P (k_{cat}=0.1s^{-1})
FGFR Y654 → FGFR Y654P (k_{cat}=0.1s^{-1})
3. When FGFR Y653P and FGFR Y654P
FGFR Y463 → FGFR Y463P (k_{cat}=70s^{-1})
FGFR Y583 → FGFR Y583P (k_{cat}=70s^{-1})
FGFR Y585 → FGFR Y585P (k_{cat}=70s^{-1})
FGFR Y766 → FGFR Y766P (k_{cat}=70s^{-1})
4. FGFR binds FRS2
FGFR+ FRS2 ↔ FGFR:FRS2 (k_{on} = 1e+6M^{-1}s^{-1}, k_{off}=2e−2s^{-1})
5. When FGFR Y653P, FGFR Y654P and FGFR:FRS2
FRS2 Y196 → FRS2 Y196P (k_{cat}=0.2s^{-1})
FRS2 Y290 → FRS2 Y290P (k_{cat}=0.2s^{-1})
FRS2 Y306 → FRS2 Y306P (k_{cat}=0.2s^{-1})
FRS2 Y382 → FRS2 Y382P (k_{cat}=0.2s^{-1})
FRS2 Y392 → FRS2 Y392P (k_{cat}=0.2s^{-1})
FRS2 Y436 → FRS2 Y436P (k_{cat}=0.2s^{-1})
FRS2 Y471 → FRS2 Y471P (k_{cat}=0.2s^{-1})
6. Reverse when Shp2 bound to FRS2:
FRS2 Y196P → FRS2 Y196 (k_{cat}=12s^{-1})
FRS2 Y290P → FRS2 Y290 (k_{cat}=12s^{-1})
FRS2 Y306P → FRS2 Y306 (k_{cat}=12s^{-1})
FRS2 Y382P → FRS2 Y382 (k_{cat}=12s^{-1})
FRS2 Y436P → FRS2 Y436 (k_{cat}=12s^{-1})
FRS2 Y471P → FRS2 Y471 (k_{cat}=12s^{-1})
FRS2 Y392P → FRS2 Y392 (k_{cat}=12s^{-1})
7. FRS2 effectors bind phosphoFRS2:
Src+FRS2 Y196P ↔ Src:FRS2 Y2196P (k_{on} = 1e+6M^{-1}s^{-1}, k_{off}=2e−2s^{-1})
Grb2+FRS2 Y306P ↔ Grb2:FRS2 Y306P(k_{on} = 1e+6M^{-1}s^{-1}, k_{off}=2e−2s^{-1})
Shp2+FRS2 Y471P ↔ Shp2:FRS2 Y471P(k_{on} = 1e+6M^{-1}s^{-1}, k_{off}=2e−2s^{-1})
8. When Src:FRS2 we relocate/remove
Src:FRS2 → relocate out ($t_{1/2}$=15min)
9. When Plc:FGFR it degrades FGFR
PLC+FGFRY 766 ↔ PLC:FGFR 766(k_{on} = 1e+6M^{-1}s^{-1}, k_{off}=2e−2s^{-1})
PLC:FGFR 766 → degFGFR ($t_{1/2}$=60min)
10. Spry appears in time-dependent manner:
→ Spry ($t_{1/2}$=15min)
11. Spry binds Src and is phosphorylated:
Spry+Src ↔ Spry Y55:Src (k_{on} = 1e+5M^{-1}s^{-1}, k_{off}=1e−4s^{-1})
Spry Y55:Src → Spry Y55P:Src (k_{cat}=10s^{-1})
Spry Y55P+Src ↔ Spry Y55P:Src (k_{on} = 1e+5M^{-1}s^{-1}, k_{off}=1e−4s^{-1})
Spry Y55P+Cbl ↔ Spry Y55P:Cbl (k_{on} = 1e+5M^{-1}s^{-1}, k_{off}=1e−4s^{-1})
Spry Y55P+Grb2 ↔ Spry Y55P:Grb2(k_{on} = 1e+5M^{-1}s^{-1}, k_{off}=1e−4s^{-1})
12. phosphoSpry binds Cbl which degrades/removes FRS2
Spry Y55P:Cbl+FRS2 ↔ FRS-Ubi (k_{cat}=8.5e−4s^{-1})
FRS2-Ubi → degFrs2 ($t_{1/2}$=5min)
13. Spry is dephosphorylated by Shp2: (when Shp2 bound to FRS2)
Spry Y55P → Spry Y55 (k_{cat}=12s^{-1})
14. Grb2 binds Sos
Grb2+Sos ↔ Grb2:Sos (k_{on} = 1e+5M^{-1}s^{-1}, k_{off}=1e−4s^{-1})

Fig. 3. Reaction rules for the pathway

phorylated FRS2, this leads to relocation (i.e. endocytosis and/or degradation of FGFR:FRS2). In step **9.** if Plc is bound to Y766 of FGFR, this leads to relocation/degradation of FGFR.

10. The signal attenuator Spry is a known inhibitor of FGFR signalling and is synthesised in response to FGFR signalling. Here we include a variable to regulate the concentration of Spry protein in a time dependent manner.

11. We incorporate the association of Spry with Src and concomitant phosphorylation of Spry residue Y55.
12. The Y55 phosphorylated form of Spry binds with Cbl, which leads to ubiquitin modification of FRS2 and a degradation of FRS2 through ubiquitin-mediated proteolysis.
13. The Y55P form of Spry is dephosphorylated by Shp2 bound to FRS2 Y247P.
14. Grb2 binds to the Y55P form of Spry. In our model Spry competes with FRS2 for Grb2 as has been suggested from some studies in the literature.

Note that this model is not intended to, and cannot be, a fully accurate representation of a real-world FGF signalling pathway. Its primary purpose at this stage of development is as a tool to evaluate biological hypotheses that are not easily obtained by intuition or manual methods. To this end, the model is an abstraction as argued in [17], created to facilitate predictive "in silico" experiments for a range of scenarios. Results of such "in silico genetics" experiments based on simulations of a stochastic π-calculus model of the above set of reactions are described in [22] (see also [13]).

We explicitly draw attention to the following issues. The reactions selected are based upon their current biological interest rather than complete understanding of the components of FGF signalling. Indeed, at this stage we have ignored many reactions that could prove significant in regulation of FGFR signalling in real cells. However, the design permits the incorporation of further modifications to the core model as biological understanding advances. The model is idealised in that it does not take into account variations in composition, affinities or rate constants that might occur in different cell types or physiological conditions. However, a useful computational modelling approach should accommodate future quantitative or qualitative modifications to the core model.

5 Modelling in PRISM

We now describe the specification in PRISM of the FGF model from the previous section. We employ a combination of the two approaches discussed in Section 3. Each of the basic elements of the pathway, including all possible compounds and receptors residues (FGF, FGFR, FRS2, Plc, Src, Spry, Sos, Grb2, Cbl and Shp2) is represented by a separate PRISM module. Synchronisation between modules is used to model reactions involving interactions of multiple elements. However, the different forms which each can take (for example, which other compounds it is bound to) are represented by one or more variables within the module.

Our model represents a single instance of the pathway, i.e. there can be at most one of each compound. This has the advantage that the resulting state space is relatively small (80,616 states); however, the model is highly complex due to the large number of different interactions that can occur in the pathway (there are over 560,000 transitions between states). Furthermore, as will be demonstrated later in the paper, the model is sufficiently rich to explain the roles of the components in the pathway and how they interact. The study of a

```
formula Frs = relocFrs2=0 ∧ degFrs2=0; // FRS2 not relocated or degraded
module FRS2
  FrsUbi     : [0..1] init 0; // ubiquitin modification of FRS2
  relocFrs2  : [0..1] init 0; // FRS2 relocated
  degFrs2    : [0..1] init 0; // FRS2 degraded
  Y196P      : [0..1] init 0;... Y471P : [0..1] init 0; // phosphorylation of receptors
  // compounds bound to FRS2
  FrsFgfr    : [0..1] init 0; // 0: FGFR not bound, 1: FGFR bound
  FrsGrb     : [0..2] init 0; // 0: Grb2 not bound, 1: Grb2 bound, 2: Grb2:Sos bound
  FrsShp     : [0..1] init 0; // 0: Shp2 not bound, 1: Shp2 bound
  FrsSrc     : [0..8] init 0;
  // 0: Src not bound       1: Src bound,            2: Src:Spry
  // 3: Src:SpryP,          4: Src:SpryP:Cbl,        5: Src:SpryP:Grb
  // 6: Src:SpryP:Grb:Cbl,  7: Src:SpryP:Grb:Sos,    8: Src:SpryP:Grb:Sos:Cbl
                                  ...
  // phosphorylation of receptors (5)
  [] Frs∧Y653P=1∧Y654P=1∧FrsFgfr=1∧Y196P=0 → 0.2 : (Y196P′=1); // Y196
                                  ...
  [] Frs∧Y653P=1∧Y654P=1∧FrsFgfr=1∧Y471P=0 → 0.2 : (Y471P′=1); // Y471
  // dephosphorylation of Y196 (6) - remove Src if bound
  []          Frs∧FrsShp=1∧Y196P=1∧FrsSrc=0 → 12 : (Y196P′=0);
  [src_rel]  Frs∧FrsShp=1∧Y196P=1∧FrsSrc>0 → 12 : (Y196P′=0)∧(FrsSrc′=0);
                                  ...
  // dephosphorylation of Y471 (6) - remove Shp2 since bound
  [shp_rel] Frs∧FrsShp=1∧Y471P=1 → 12 : (Y471P′=0)∧(FrsShp′=0);
                                  ...
  // Src:FRS2→degFRS2 [8]
  [] Frs∧FrsSrc>0 → 1/(15*60) : (relocFrs2′=1);
                                  ...
  // Spry55p:Cbl+FRS2→Frs-Ubi [12]
  [] Frs∧FrsSrc=4,6,8 ∧ FrsUbi=0 → 0.00085 : (FrsUbi′=1);
  // FRS2-Ubi→degFRS2 [12]
  [] Frs∧FrsUbi=1 → 1/(5*60) : (degFrs2′=1);
                                  ...
  // Grb2+Sos↔Grb2:Sos [14]
  [sos_bind_frs]  Frs∧FrsGrb=1  → 1       : (FrsGrb′=2);          // Grb:FRS2
  [sos_bind_frs]  Frs∧FrsSrc=5,6→ 1       : (FrsSrc′=FrsSrc+2); // Grb:SpryP:Src:FRS2
  [sos_rel_frs]   Frs∧FrsGrb=2  → 0.0001 : (FrsGrb′=1);          // Grb:FRS2
  [sos_rel_frs]   Frs∧FrsSrc=7,8→ 0.0001 : (FrsSrc′=FrsSrc−2); // Grb:SpryP:Src:FRS2
                                  ...
endmodule
```

Fig. 4. Fragment of the PRISM module for FRS2 and related compounds

single instance of the pathway is also motivated by the fact that the same signal dynamics (Figure 7(a)) were obtained in [22, 13] for a model where the number of molecules of each type were initially set to 100. Fragments of the PRISM code for the modules representing FRS2, Src and Sos are given in Figures 4, 5 and 6, respectively. The full version is available from the PRISM web page [14].

Figure 4 shows the module for FRS2. It contains variables representing whether FRS2 is currently: undergoing ubiquitin modification (*FrsUbi*); relocated (*relocFrs2*); degraded (*degFrs2*); and bound to other compounds (*FrsFgfr*, *FrsGrb*, *FrsShp* and *FrsSrc*). It also has variables representing the phosphorylation status of each of FRS's receptors ($Y196P, \ldots, Y471P$).

The first set of commands given in Figure 4 correspond to the phosphorylation of receptors in FRS (reaction **5** in Figure 3). Since the only variables that are updated are local to this module, the commands have no action label, i.e. we do not require any other module to synchronise on these commands. The guards of these commands incorporate dependencies on the current state both of FRS2

```
module SRC
  Src : [0..8] init 1;
  // 0: Src bound to FRS2,  1: Src not bound,        2: Src:Spry
  // 3: Src:SpryP,          4: Src:SpryP:Cbl,        5: Src:SpryP:Grb
  // 6: Src:SpryP:Grb:Cbl,  7: Src:SpryP:Grb:Sos,    8: Src:SpryP:Grb:Sos:Cbl

  // Src+FRS2196P↔Src:FRS2 (7)
  [src_bind] Src>0 → (Src′=0);
  [src_rel]  Src=0 → (Src′=FrsSrc);
  // Spry+Src→Spry55:Src or Spry55P+Src→Spry55P:Src (11)
  [spry_bind] Src=1 → 1 : (Src′=Spry+1);
  // Spry+Src←Spry55:Src (11)
  [spry_rel] Src=2 → 0.01 : (Src′=1);
  // Spry55P+Src←Spry55P:Sr (11)c
  [spry_rel] Src>2 → 0.0001 : (Src′=1);
  // Spry55:Src→Spry55P:Src (11)
  [] Src=2 → 10 : (Src′=3);
  // SpryP+Cbl↔SpryP:Cbl (11)
  [cbl_bind_src] Src=3,5,7→ 1       : (Src′=Src+1);
  [cbl_rel_src]  Src=4,6,8→ 0.0001 : (Src′=Src−1);
  // SpryP+Grb↔SpryP:Grb (11)
  [grb_bind_src] Src=3,4  → 1       : (Src′=Src+2*Grb);
  [grb_rel_src]  Src=5,6  → 0.0001 : (Src′=Src−2); // SOS not bound
  [grb_rel_src]  Src=7,8  → 0.0001 : (Src′=Src−4); // SOS bound
                          ...
endmodule
```

Fig. 5. PRISM module for Src and related compounds

```
module SOS
  Sos : [0..1] init 1;

  // Grb2+Sos↔Grb:Sos
  [sos_bind]     Sos=1 → (Sos′=0); // Grb2 free
  [sos_bind_frs] Sos=1 → (Sos′=0); // Grb2:FRS2 or to Grb2:SpryP:SRC:FRS2
  [sos_rel]      Sos=0 → (Sos′=1); // Grb2 free
  [sos_rel_frs]  Sos=0 → (Sos′=1); // Grb2:FRS2 or to Grb2:SpryP:SRC:FRS2
                 ...
endmodule
```

Fig. 6. PRISM module for Sos

itself and of other compounds. More precisely, FGFR must be bound to FRS2 and certain receptors of FGFR must have already been phosphorylated.

Elsewhere, in Figure 4, we see commands that use synchronisation to model interactions with other compounds, e.g. the release of Src (the commands labelled *src_rel*) and the binding and release of Sos (the commands labelled *sos_bind_frs* and *sos_rel_frs*). Note the corresponding commands in modules *SRC* (Figure 5) and *SOS* (Figure 6). In each of these cases, as discussed in Section 3, the rate of the combined interaction is specified in the *FRS2* module and is hence omitted from the corresponding commands in *SRC* and *SOS*. Also, in the module for Sos (Figure 6), there are different action labels for the binding and release of Sos with Grb2; this is because Grb2 can be either free or bound to a number of different compounds when it interacts with Sos. For example, Grb2 can be bound to Frs2 (through reaction **7**) or Spry (through reaction **11**), and Spry can in turn be bound to Src, which can also be bound to FRS2.

Notice how, in the commands for binding and unbinding of Src with FRS2 in Figure 4 (labelled *sos_bind_frs* and *sos_rel_frs*), we can use the value of *FrsSrc*

to update the value of *Src*, rather than separating each case into individual commands. Also worthy of note are the updates to *Src* in Figure 5 when either Grb2 or Grb2:Sos bind to Src. To simplify the code, we have used a single command for each of these possible reactions, and therefore updates which either increment or decrement the variable *Src* by 2 or 4 (the variable *Grb* takes value 1 if Grb2 is not bound to Sos and value 2 if Sos is bound).

6 Property specification

Our primary goal in this case study is to analyse the various mechanisms previously reported to negatively regulate signalling. Since the binding of Grb2 to FRS2 serves as the primary link between FGFR activation and ERK signalling, we examine the amount of Grb2 bound to FRS2 as the system evolves. In addition, we investigate the different causes of degradation which, based on the system description, can be caused by one of the following reactions occurring:

- when Src:FRS2 is present, FRS2 is relocated (reaction **8**);
- when Plc:FGFR is present, it degrades FGFR (reaction **9**);
- when phosphoSpry binds to Cbl, it degrades FRS2 (reaction **12**).

Below, we present a list of the various properties of the model that we have analysed, and the form in which they are supplied to the PRISM tool. For the latter, we define a number of *atomic propositions*, essentially predicates over the variables in the PRISM model, which can be used to identify states of the model that have certain properties of interest. These include a_{grb2}, which indicates that Grb2 is bound to FRS2 (i.e. those states where the variable *FrsGrb* of Figure 4 is greater than zero), and a_{src}, a_{plc} and a_{spry}, corresponding to the different causes of degradation/relocation given above. For properties using expected rewards (with the $\mathcal{R}_{=?}[\cdot]$ operator), we also explain the reward structure used.

A. *What is the probability that Grb2 is bound to FRS2 at the time instant T?* ($\mathcal{P}_{=?}[\texttt{true}\ \mathcal{U}^{[T,T]}\ a_{grb2}]$);
B. *What is the expected number of times that Grb2 binds to FRS2 by time T?* ($\mathcal{R}_{=?}[\mathcal{C}^{\leq T}]$, where a reward of 1 is assigned to all transitions involving Grb2 binding to FRS2);
C. *What is the expected time that Grb2 spends bound to FRS2 within the first T time units?* ($\mathcal{R}_{=?}[\mathcal{C}^{\leq T}]$, where a reward of 1 is assigned to states where Grb2 is bound to FRS2, i.e. those satisfying atomic proposition a_{grb2});
D. *What is the long-run probability that Grb2 is bound to FRS2?* ($\mathcal{S}_{=?}[a_{grb2}]$);
E. *What is the expected number of times Grb2 binds to FRS2 before degradation or relocation occurs?* ($\mathcal{R}_{=?}[F\ (a_{src} \vee a_{plc} \vee a_{spry})]$, with rewards as for **B**);
F. *What is the expected time Grb2 spends bound to FRS2 before degradation or relocation occurs?* ($\mathcal{R}_{=?}[F\ (a_{src} \vee a_{plc} \vee a_{spry})]$, with rewards as for **C**);
G. *What is the probability that each possible cause of degradation/relocation has occurred by time T?* (e.g. $\mathcal{P}_{=?}[\neg(a_{src} \vee a_{plc} \vee a_{spry})\ \mathcal{U}^{[0,T]}\ a_{src}]$ in the case Src causes relocation);

H. *What is the probability that each possible cause of degradation/relocation occurs first?* (e.g. $\mathcal{P}_{=?}[\neg(a_{src} \vee a_{plc} \vee a_{spry})\, \mathcal{U}\, a_{plc}]$ in the case when Plc causes degradation);

I. *What is the expected time until degradation or relocation occurs in the pathway?* ($\mathcal{R}_{=?}[\mathcal{F}\ (a_{src} \vee a_{plc} \vee a_{spry})]$ where all states have reward 1).

7 Results and analysis

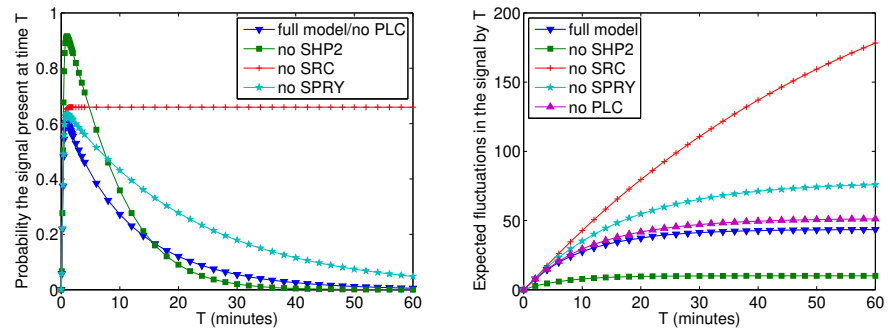
We used PRISM to construct the FGF model described in Section 5 and analyse the set of properties listed in Section 6. This was done for a range of different scenarios. First, we developed a base model, representing the full system, in which we suppose that initially FGF, unbound and unphosphorylated FGFR, unphosphorylated FRS2, unbound Src, Grb2, Cbl, Plc and Sos are all present in the system (Spry arrives into the system with the half-time of 10 minutes).

Subsequently, we performed a series of "in silico genetics" experiments on the model designed to investigate the roles of the various components of the activated receptor complex in controlling signalling dynamics. This involves deriving a series of modified models of the pathway where certain components are omitted (Shp2, Src, Spry or Plc), and is easily achieved in a PRISM model by just changing the initial value of the component under study. For example, to remove Src from the system we just need to change the initial value of the variable *Src* from 1 to 0 (see Figure 5).

For each property we include the statistics for 5 cases: for the full pathway and for the pathway when either Shp2, Src, Spry or Plc is removed. Figures 7(a)–(c) show the transient behaviour (i.e. at each time instant T) of the signal (binding of Grb2 to FRS2) for the first 60 minutes, namely properties **A**, **B** and **C** from the previous section. Table 1 gives the the long-run behaviour of the signal, i.e. properties **D**, **E** and **F**. The latter three results can be regarded as the values of the first three in "the limit", i.e. as either T tends to infinity or degradation occurs. Figures 7(d)–(f) show the transient probability of each of the possible causes of relocation or degradation occurring (property **G**). Table 2 shows the results relating to degradation in the long-run (properties **H** and **I**).

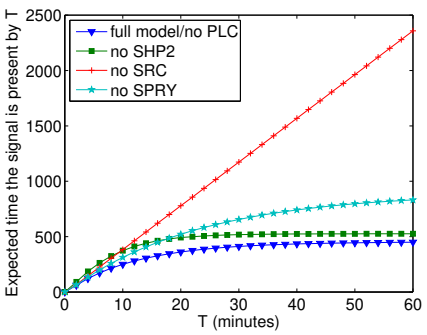
We begin with an analysis of the signal (binding of Grb2 to FRS2) in the full model, i.e. see the first plot ("full model") in Figure 7 and the first lines of Tables 1 and 2. The results presented demonstrate that the probability of the signal being present (Figure 7(a)) shows a rapid increase, reaching its maximum level at about 1 to 2 minutes. The peak is followed by a gradual decrease in the signal, which then levels off at a small non-zero value. In this time interval Grb2 repeatedly binds to FRS2 (Figure 7(b)) and, as time passes, Grb2 spends a smaller proportion of time bound to FRS2 (Figure 7(c)).

The rapid increase in the signal is due the relevant reactions (the binding of Grb2 to FRS2 triggered by phosphorylation of FRS2, which requires activated FGFR to first bind to FRS2) all occurring at very fast rates. On the other hand, the decline in the signal is caused either by dephosphorylation of FRS2 (due to Shp2 being bound to FRS2) or by relocation/degradation of FRS2.

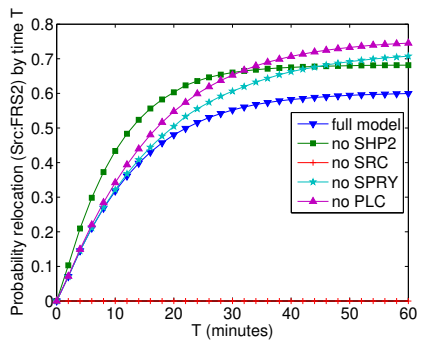


(a) Probability bound (Grb2:FRS2)

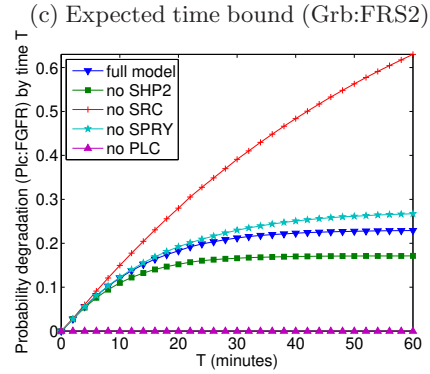
(b) Expected bindings (Grb2:FRS2)



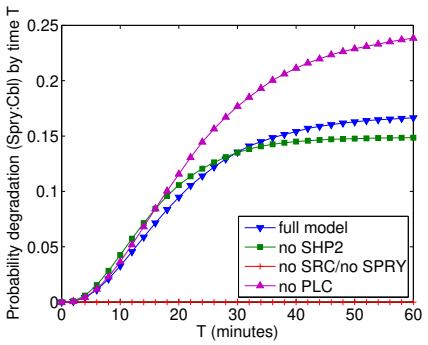
(c) Expected time bound (Grb:FRS2)



(d) Probability relocated (Src:FRS2)



(e) Probability degraded (Plc:Fgfr)



(f) Probability degraded (Spry:Cbl)

Fig. 7. Transient numerical results

Dephosphorylation of FRS2 is both fast and allows Grb2 to rebind (as FRS2 can become phosphorylated again). The overall decline in signal is due to relocation of FRS2 caused by bound Src which takes a relatively long time to occur (Table 2 and Figure 7(d)). Degradation caused by Spry has little impact since it is not present from the start and, by the time it appears, it is more likely that Grb2 is no longer bound or Src has caused relocation (Table 2, Figure 7(d) and Figure 7(f)).

	probability bound	expected no. of bindings	expected time bound (min)
full model	7.54e-7	43.1027	6.27042
no Shp2	3.29e-9	10.0510	7.78927
no Src	0.659460	283.233	39.6102
no Spry	4.6e-6	78.3314	10.8791
no Plc	0.0	51.5475	7.56241

Table 1. Long run and expected reachability properties for the signal

	probability of degradation/relocation			expected
	Src:FRS2	Plc:FGFR	Spry:Cbl	time (min)
full model	0.602356	0.229107	0.168536	14.0258
no Shp2	0.679102	0.176693	0.149742	10.5418
no Src	-	1.0	0.0	60.3719
no Spry	0.724590	0.275410	-	16.8096
no Plc	0.756113	-	0.243887	17.5277

Table 2. Probability and expected time until degradation/relocation in the long run

The fact that the signal levels out at a non-zero value (Table 1) is caused by Plc degrading the FGF receptor bound to FRS2 and Grb2. More precisely, after FGFR is degraded by Plc, no phosphorylation of partner FRS2 residues is possible. The signal stays non-zero since neither Src-mediated relocation and degradation, nor Shp-mediated dephosphorylation, are possible when respective FRS2 residues are not active. The non-zero value is very small because it is more likely that Src has caused relocation (Table 2). The repeated binding of Grb2 to FRS2 (Figure 7(b)) is caused by the dephosphorylation of FRS2, which is soon phosphorylated again and allows Grb2 to rebind. The decrease in the proportion of time that Grb2 is bound to FRS2 is due to the probability of FRS2 becoming relocated/degraded increasing as time passes (Figure 7(d)–(f)).

Next, we further illustrate the role of the components by analysing models in which different elements of the pathway are not present.

Shp2. Figure 7(a) shows that the peak in the signal is significantly larger than that seen under normal conditions. By removing Shp2 we have removed, as explained above, the fast reaction for the release of Grb2 from FRS2, and this justifies the larger peak. The faster decline in the signal is due to there being a greater chance of Src being bound (as Shp2 causes the dephosphorylation of FRS2, it also causes the release of Src from FRS2), and hence the increased chance relocation (Figure 7(d) and Table 2). These observations are also the cause for the decrease in the time until degradation/relocation when Shp2 is removed (Table 2) and the fact that the other causes of degradation/relocation are less likely (Figures 7(e)–(f) and Table 2). Dephosphorylation due to bound Shp2 was responsible for the large number of times that Grb2 and FRS2 bind (and unbind) in the original model; we do not see such a large number of bindings once Shp2 is removed (Figure 7(b) and Table 1).

Src. As Figure 7(a) demonstrates, the suppression of Src is predicted to have a major impact on signalling dynamics: after a fast increase, the signal

fails to decrease substantially. This is supported by the results presented in both Figures 7(d)–(f) and Table 2 which show that Src is the main cause of signal degradation, and by removing Src the time until degradation or relocation greatly increases. The failure of Spry to degrade the signal (Figure 7(f) and Table 2) is attributed to its activation being downstream of Src. Note that, this also means that Plc is the only remaining cause of degradation.

Spry. The model fails to reproduce the role of Spry in inhibiting the activation of the ERK pathway by competition for Grb2:Sos. More precisely, our results show that the suppression of Spry does not result in signal reduction. This can be explained by the differences in system designs: under laboratory conditions the action of Spry is measured after Spry is over-expressed, whereas, under normal physiological conditions, Spry is known to arrive slowly into the system. Removing Spry removes one of the causes of degradation, and therefore increases the other causes of degradation/relocation (Figures 7(d)–(e) and Table 2). Moreover, the increase in the probability of Plc causing degradation/relocation leads to an increase in the chance of Grb2 and FRS2 remaining bound (Table 2).

Plc. While having a modest effect on transient signal expression, the main action of Plc removal is to cause the signal to stabilise at zero (Table 1). This is due to Plc being the only causes of degradation/relocation not relating to FRS2. The increase in time until degradation (Table 2) is also attributed to the fact that, by removing Plc, we have eliminated one of the possible causes of degradation. This also has the effect that the other causes of relocation/degradation are more likely (Figure 7(d), Figure 7(f) and Table 2).

8 Conclusions

In this paper we have shown that probabilistic model checking can be a useful tool in the analysis of biological pathways. The technique's key strength is that it allows the calculation of exact quantitative properties for system events occurring over time, and can therefore support a detailed, quantitative analysis of the interactions between the pathway components. By developing a model of a complex, realistic signalling pathway that is not yet well understood, we were able to demonstrate, firstly, that the model is robust and that its predictions agree with biological data [22, 13] and, secondly, that probabilistic model checking can be used to obtain a wide range of quantitative measures of system dynamics, thus resulting in deeper understanding of the pathway.

We intend to perform further analysis of the FGF pathway, including an investigation into the effect that changes to reaction rates and initial concentrations will have on the pathway's dynamics. Future work will involve both comparing this probabilistic model checking approach with simulation and ODEs, and also investigation of how to scale the methodology yet further.

References

1. A. Aziz, K. Sanwal, V. Singhal, and R. Brayton. Verifying continuous time Markov chains. In *Proc. CAV'96*, volume 1102 of *LNCS*, pages 269–276. Springer, 1996.

2. C. Baier, B. Haverkort, H. Hermanns, and J.-P. Katoen. Model checking continuous-time Markov chains by transient analysis. In *Proc. CAV'00*, volume 1855 of *LNCS*, pages 358–372. Springer, 2000.
3. M. Calder, S. Gilmore, and J. Hillston. Modelling the influence of RKIP on the ERK signalling pathway using the stochastic process algebra PEPA. *Transactions on Computational Systems Biology*, 2006. To appear.
4. M. Calder, V. Vyshemirsky, D. Gilbert, and R. Orton. Analysis of signalling pathways using continuous time Markov chains. *Transactions on Computational Systems Biology*, 2006. To appear.
5. I. Dikic and S. Giordano. Negative receptor signalling. *Curr Opin Cell Biol.*, 15:128–135, 2003.
6. V. Eswarakumar, I. Lax, and J. Schlessinger. Cellular signaling by fibroblast growth factor receptors. *Cytokine Growth Factor Rev.*, 16(2):139–149, 2005.
7. D. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, 1977.
8. J. Hillston. *A Compositional Approach to Performance Modelling*. Cambridge University Press, 1996.
9. A. Hinton, M. Kwiatkowska, G. Norman, and D. Parker. PRISM: A tool for automatic verification of probabilistic systems. In *Proc. TACAS'06*, volume 3920 of *LNCS*, pages 441–444. Springer, 2006.
10. M. Kwiatkowska, G. Norman, and D. Parker. Probabilistic symbolic model checking with PRISM: A hybrid approach. *International Journal on Software Tools for Technology Transfer (STTT)*, 6(2):128–142, 2004.
11. M. Kwiatkowska, G. Norman, and D. Parker. Probabilistic model checking in practice: Case studies with PRISM. *ACM SIGMETRICS Performance Evaluation Review*, 32(4):16–21, 2005.
12. A. Phillips and L. Cardelli. A correct abstract machine for the stochastic pi-calculus. In *Proc.BioCONCUR'04*, ENTCS. Elsevier, 2004.
13. `www.cs.bham.ac.uk/~oxt/fgfmap.html`.
14. PRISM web site. `www.cs.bham.ac.uk/~dxp/prism`.
15. C. Priami. Stochastic π-calculus. *The Computer Journal*, 38(7):578–589, 1995.
16. C. Priami, A. Regev, W. Silverman, and E. Shapiro. Application of a stochastic name passing calculus to representation and simulation of molecular processes. *Information Processing Letters*, 80:25–31, 2001.
17. A. Regev and E. Shapiro. Cellular abstractions: Cells as computation. *Nature*, 419(6905):343, 2002.
18. A. Regev, W. Silverman, and E. Shapiro. Representation and simulation of biochemical processes using the pi- calculus process algebra. In *Pacific Symposium on Biocomputing*, volume 6, pages 459–470. World Scientific Press, 2001.
19. J. Rutten, M. Kwiatkowska, G. Norman, and D. Parker. *Mathematical Techniques for Analyzing Concurrent and Probabilistic Systems*, volume 23 of *CRM Monograph Series*. AMS, 2004.
20. J. Schlessinger. Epidermal growth factor receptor pathway. Sci. STKE (Connections Map), `http://stke.sciencemag.org/cgi/cm/stkecm;CMP\_14987`.
21. M. Tsang and I. Dawid. Promotion and attenuation of FGF signaling through the Ras-MAPK pathway. *Science STKE*, pe17, 2004.
22. O. Tymchyshyn, G. Norman, J. Heath, and M. Kwiatkowska. Computer assisted biological reasoning: The simulation and analysis of FGF signalling pathway dynamics. Submitted for publication.