Probabilistic Model Checking for Biochemical Reaction Systems

Ratana Ty Ken-etsu Fujita Ken'ichi Kawanishi
Graduated School of Science and Technology
Gunma University

Abstract Probabilistic model checking is an automated verification technique for verifying the correctness and performance of probabilistic models. This paper shows how to use this verification technique to apply to biochemical reaction systems where the numbers of molecules are small and the reactions are governed by stochastic rates with various scales. We work on a case study analysing Fibroblast Growth Factor (FGF) molecular signaling modeled as continuous-time Markov chains (CTMCs). We verify this model by using model checker PRISM and quasi-stationary distributions for CTMCs.

1. Introduction

Probabilistic model checking [1] is an automated verification technique for verifying the correctness and performance of probabilistic models. It involves the systematic analysis of a probabilistic model, typically a variant of a Markov chain. This technique is an important component in the design and analysis of software and hardware systems. It has been widely used to study about systems including biological systems such as biochemical reaction networks and signaling pathways.

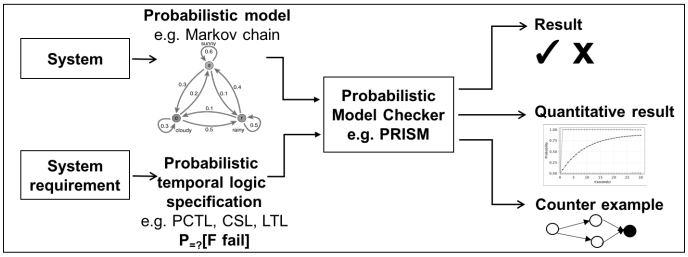


Figure 1: Probabilistic model checking

1.1. Modelling molecular networks

There are two established frameworks for modelling molecular reactions, the continuous deterministic approach and the discrete stochastic approach [2].

- In the continuous deterministic approach, one approximates the number of molecules using a continuous function that represents the change in molecule concentration using differential equations (ODEs) based on mass action kinetics. This approach is suitable for modelling large numbers of molecules and the solutions give the (average) concentration of each molecular species over time.
- The discrete stochastic approach models the stochastic evolution of populations of molecules, where reactions are discrete events, governed by stochastic rates typically assumed to be constant and dependent on the number of molecules, which admits their modelling in terms of continuous-time Markov chains. This approach is more accurate in cases where the numbers of molecules are small.

1.2. Aim

This paper shows how probabilistic model checking can be used to apply to biochemical reaction systems where the numbers of molecules are small and the reactions are governed by stochastic rates with various scales. We observe the mechanisms of the biochemical reaction before reaching the absorbing state by analising the numbers and time of visiting as well as staying at any states by both model checking tool PRISM and quasi-stationary distributions.

2. Probabilistic Model Checking

We refer to [8] as the standard text of model checking including probabilistic systems.

2.1. Probabilistic model checking for continuous-time Markov Chains

Markov chains (MCs) are viewed as operational models for probabilistic systems and at the same time structures for Continuous Stochastic Logic (CSL) [9]. MCs are handled as a transition system with probabilities rather than a sequence of random variables, and logic serves for expressing properties of probabilistic systems in a systematic method.

A continuous-time Markov chain (CTMC) [1] is defined by the following tuple

$$\mathcal{C} = (\mathsf{S}, \bar{s}, \mathsf{P}, \mathsf{E}, \mathsf{AP}, \mathsf{L}),$$

where S is a finite set of states, $\bar{s} \in \mathsf{S}$ is a distinguished initial state, P is a transition probability matrix such that $\sum_{s' \in \mathsf{S}} \mathsf{P}(s, s') = 1$, $\mathsf{E}{:}\mathsf{S} \to \mathbb{R}_{\geq 0}$ is the exit rate, AP is a set of atomic propositions, and $\mathsf{L}{:}\mathsf{S} \to 2^{\mathsf{AP}}$ is a labelling function.

In a CTMC $\mathcal{C}$ [1], the residence time of a state $s \in \mathsf{S}$ is a random variable governed by an exponential distribution with exit rate parameter $\mathsf{E}(s)$. The probability to exit state s in t time units is given by $\int_0^t \mathsf{E}(s) \cdot e^{-\mathsf{E}(s)\tau} d\tau$. The probability of transition from state s to another state s' in t time units is equal to $\mathsf{P}(s, s') \cdot \int_0^t \mathsf{E}(s) \cdot e^{-\mathsf{E}(s)\tau} d\tau$.

Alternatively, a CTMC can be defined by specifying the rates matrix $\mathsf{R} : \mathsf{S} \times \mathsf{S} \to \mathbb{R}_{\geq 0}$ where $\mathsf{R}(s, s')$ is the rate of transitioning from state s to s'. A transition can only occur between states s and s' if $\mathsf{R}(s, s') > 0$. The probability of this transition from s within t time-units is $1 - e^{-\mathsf{R}(s,s') \cdot t}$. If there is more than one state s' for which $\mathsf{R}(s, s') > 0$, the first transition to be taken place determines the next state of the CTMC. The exit rate $\mathsf{E}(s)$ is then equal to $\sum_{s' \neq s} \mathsf{R}(s, s')$ and the embedded discrete-time Markov chain (DTMC) is given by

$$\mathsf{P}(s, s') = \begin{cases} \frac{\mathsf{R}(s,s')}{\mathsf{E}(s)} & \text{if } \mathsf{E}(s) \neq 0 \text{ and } s \neq s', \\ 1 & \text{if } \mathsf{E}(s) = 0 \text{ and } s = s', \\ 0 & \text{otherwise.} \end{cases}$$

2.2. Continuous Stochastic Logic (CSL)

To specify quantitative properties of CTMCs, the logic CSL [2] has been proposed. CSL distinguishes between state formulas (Φ) and path formulas (Ψ) and includes $\mathbf{P}_{\sim p}[\Psi]$ (probabilistic operator), $\mathbf{S}_{\sim p}$ (steady state), path operators $\mathbf{X}\Phi$ (next state), as well as $\Phi \mathbf{U}^{[t,t']} \Phi$ (time bounded until). The syntax of CSL is given by

$$\Phi ::= \mathsf{true} \mid \mathsf{a} \mid \neg\Phi \mid \Phi \vee \Phi \mid \Phi \wedge \Phi \mid \mathbf{P}_{\sim p}[\Psi] \mid \mathbf{S}_{\sim p}[\Phi]$$
$$\Psi ::= \mathbf{X}\Phi \mid \Phi \mathbf{U}^{[t,t']} \Phi \mid \Phi \mathbf{U} \Phi$$

where a is an atomic proposition, $\sim \in \{<, \leq, >, \geq\}$, $p \in [0, 1]$, and $t, t' \in \mathbb{R}_{\geq 0}$.

2.3. Model checking for CSL over CTMCs

The infinitesimal generator matrix [2] for the CTMC $\mathcal{C} = (\mathsf{S}, \bar{s}, \mathsf{P}, \mathsf{E}, \mathsf{AP}, \mathsf{L})$ is the matrix $\mathsf{Q} : \mathsf{S} \times \mathsf{S} \to \mathbb{R}$ defined as follows:

$$\mathsf{Q}(s, s') = \begin{cases} \mathsf{R}(s, s') & \text{if } s \neq s', \\ -\sum_{s' \neq s} \mathsf{R}(s, s') & \text{otherwise.} \end{cases}$$

2.4. Probabilistic model checker PRISM

PRISM [4] is a tool for formal modelling and analysis of systems that exhibit random or probabilistic behaviour. It is an open source application developed at the University of Birmingham and University of Oxford. PRISM can build and analyse some types of probabilistic models: discrete-time Markov chains (DTMCs), continuous-time Markov chains (CTMCs), Markov decision processes (MDPs), probabilistic automata (PAs), probabilistic timed automata (PTAs). The property specification language of PRISM is based on temporal logics: probabilistic computation tree logic

(PCTL), continuous stochastic logic (CSL), linear temporal logic (LTL), probabilistic computation tree logic* (PTCL*)[1] and extensions for quantitative specifications, costs and rewards.

2.4.1. Rewards

PRISM can be used to compute properties such as how long the system has spent in a given state, or how often a certain state transition has taken place based on "rewards".

3. Biological Case Study

3.1. Fibroblast Growth Factor (FGF)

FGF [1] are family of proteins which play an important role in cell signaling. FGF have been involved with e.g. skeletal development and wound healing, etc. FGF function in many organisms from worm to humans. The mechanisms of FGF signaling are not well understood, and several hypotheses exist.

3.2. Fibroblast Growth Factor Receptor (FGFR)

FGFR are extracellular protein. They receive the signal from the growth signal to promote fibroblast to grow in normal cell. FGFR can only interact with FGF at the cell surface, but relocation causes FGFR to move inside the cell.

3.3. A simple set of reactions

The simplified set of reactions based on the role of FGF in receptor is given below. An FGF (molecule) can bind to an FGF receptor (FGFR) to form the compound FGF:FGFR. When the compound FGF:FGFR is formed, FGFR can become phosphorylated. Binding and phosphorylation, both reactions are also reversible. Finally, when FGFR is phosphorylated, it can be relocated. We assume that FGF disappears if it is bound to FGFR when relocation occurs.

FGF binds FGFR:

$$FGF + FGFR \rightarrow FGFR : FGF \qquad \text{rate } k_1 = 5000\, M^{-1}s^{-1} \qquad (1)$$

FGF releases FGFR:

$$FGFR : FGF \rightarrow FGF + FGFR \qquad \text{rate } k_2 = 0.002\, s^{-1} \qquad (2)$$

Phosphorylation[2] of FGFR (whilst FGFR:FGF):

$$FGFR \rightarrow FGFRP \qquad \text{rate } k_3 = 0.1\, s^{-1} \qquad (3)$$

Dephosphorylation[3] of FGFR:

$$FGFRP \rightarrow FGFR \qquad \text{rate } k_4 = 0.01\, s^{-1} \qquad (4)$$

Relocation of FGFR (whilst FGFRP):

$$FGFR \rightarrow relocFGFR \qquad \text{rate } k_5 = 1/60\, min^{-1} \qquad (5)$$

Each reaction has an associated kinetic rate $(k_1, \ldots, k_5)$. The kinetic rate k_1 has units $M^{-1}s^{-1}$, where M refers to the molar concentration, i.e., the number of moles per litre.

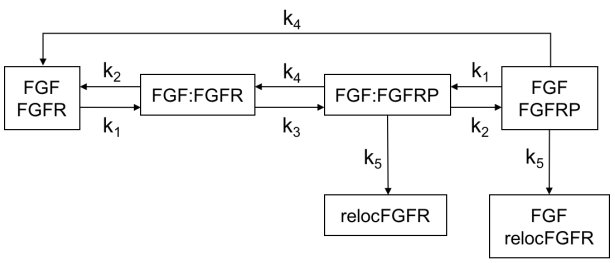


Figure 2: System of reactions of FGF

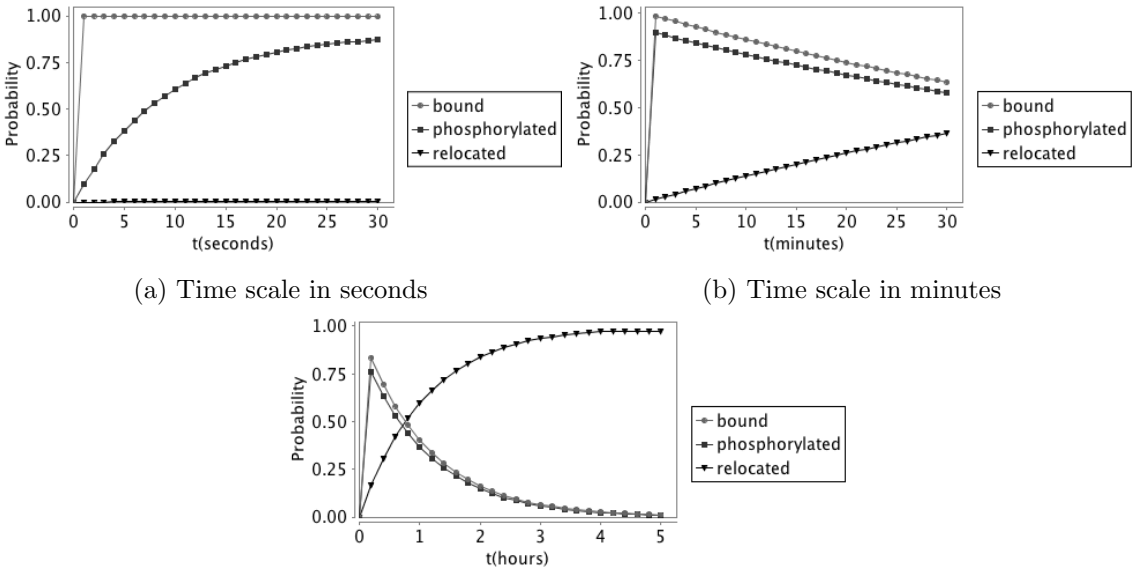
[1] PCTL* can be used to specify properties of DTMCs and MDPs (or untimed properties of CTMCs).
[2] The addition of a phosphoryl (PO_3^{2-}) group to a molecule
[3] The process by which phosphate ($PO_4^{3}-$) group are removed from a molecule by a phosphatase.

4. Model Analysis by PRISM

4.1. Modelling a single number of molecule

We give two alternative approaches [2] to model the reaction system of one of each FGF and FGFR molecule. Both correspond to the model CTMCs and a set of modules. In our first approach, there are two main modules, FGF and FGFR, each with variables representing their current state. Reactions involving more than one species are modelled using synchronisation. In our second approach, we use a single module with one variable representing the six possible states of the whole system. This approach is simple and easy to understand. However, when the system becomes more complex the use of separate variables and synchronisation as our first approach becomes more desirable since we can easily modify and express properties.

4.2. Model analysis



(a) Time scale in seconds

(b) Time scale in minutes

(c) Time scale in hours

Figure 3: Probabilities over three time scales for the FGF system to be in one of three states

Figure 3 shows results analysed with PRISM for the probability that at time instant t, FGFR is: bound to FGF; phosphorylated; relocated. Results are plotted for ranges of t over 3 different time scales (a) seconds, (b) minutes and (c) hours. Figure 3 (a) shows that in the initial evolution of the system FGF and FGFR bind very quickly after that FGFR becomes phosphorylated while there is almost no chance of FGFR relocating. Figures 3 (b) and 3 (c) show that as time passed the chance that FGF and FGFR are bound diminishes, and the chance that FGFR is phosphorylated diminishes faster, eventually FGFR become relocated.

4.3. Modelling larger number of molecules

In this section we extend our model in previous section by increasing more than one of each FGF and FGFR molecule to occur in the system. We give two possible approaches.

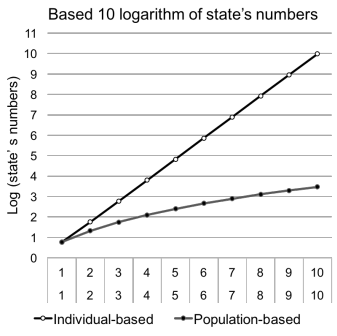
- **Individual-based model**: each individual molecule is modeled separately.
- **Population-based model**: each type of molecule or species are modeled separately rather than the state of each individual component.

4.3.1. The state-space explosion problem

Figure 4 lists the numbers of states and transitions along with their graphs, where X-axis measures the number of FGF and FGFR molecules and the Y-axis measures the base 10 logarithm of the numbers of states and transitions. As can be seen, there is a rapid increase in the number of states and transitions for the individual-based model and quickly make model checking infeasible.

This model will suffer from the well-known state-space explosion problem as the complexity of the system increases where there is an exponential growth in the state space. However, there is also an increase in the number of states for the population-based model but in the way far more gradual increase in the state space, which allow us to analyse much larger models.

(N,M)	States	
	Individual-based	Population-based
(1,1)	6	6
(2,2)	57	21
(3,3)	591	56
(4,4)	6293	126
(5,5)	67635	252
(6,6)	728881	462
(7,7)	7853235	792
(8,8)	84480445	1287
(9,9)	906804555	2002
(10,10)	9709573985	3003



(N,M)	Transitions	
	Individual-based	Population-based
(1,1)	11	11
(2,2)	199	57
(3,3)	3541	193
(4,4)	58437	509
(5,5)	902071	1140
(6,6)	13223143	2275
(7,7)	186158393	4166
(8,8)	2537193169	7137
(9,9)	33671076115	11593
(10,10)	436986997683	18029

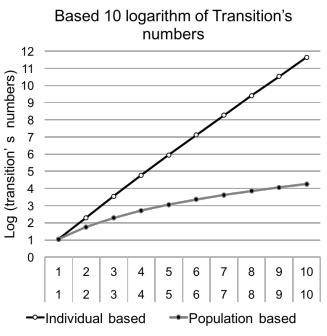


Figure 4: States and transitions of individual-based and population-based model

4.3.2. Population-based model analysis by PRISM

In order to compute the numbers of transitions to any states before relocation occurs, we give labels denoted by $[X] \in \{[\text{unbind1}], [\text{bind1}], [\text{unbind2}], [\text{bind2}], [\text{phos1}], [\text{phos2}], [\text{unbindp}]\}$ to the transitions in the system as in Figure 5. Table 1 shows the numbers of each transition by simulation and verification.

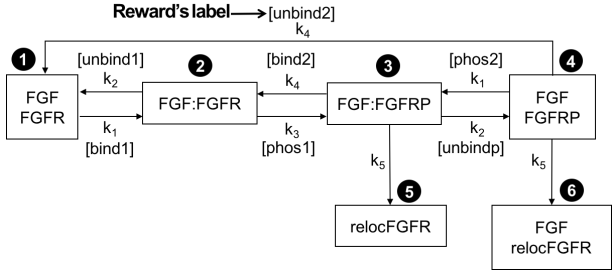


Figure 5: Labels of transitions

Table 1: Expected numbers of transition before relocation occurs by PRISM

State	Label	Number of transitions					
		Simulation (Number of samples)				Verification	Total
		(10)	(100)	(1000)	(10000)		
1	[unbind1]	0.3000	0.7300	0.7490	0.7504	0.7400	0.7400
	[unbind2]	0.0000	0.0000	0.0000	0.0000	< 0.0001	
2	[bind1]	1.4000	1.8400	1.8160	1.7304	1.7400	37.7400
	[bind2]	27.7000	33.4400	34.7100	35.8794	36.0000	
3	[phos1]	45.2000	42.0900	34.5710	36.3347	37.0000	44.2000
	[phos2]	7.2000	7.3900	7.1140	7.2736	7.2000	
4	[unbindp]	9.9000	7.1600	7.0290	7.1517	7.2000	7.2000
Total numbers of transition							89.8000

Table 2 below lists the time spent at states 1–4 before relocation occurs. Table 3 lists the

probability of reaching relocating states (absorbing states) 5 and 6. As can be seen, the probability of reaching state 5 is really high whereas there is almost no chance that it reaches state 6.

Table 2: Expected time staying at any states before relocation occurs by PRISM

State	1	2	3	4	Total
Time of staying	3.4800×10^{-4}	3.7000×10^{2}	3.6000×10^{3}	1.4000×10^{-3}	$\approx 3.9700\times10^{3}$

Table 3: Probability of reaching relocating (absorbing) state by PRISM

State	5	6
Probability	9.9999960×10^{-1}	3.9999902×10^{-7}

5. Analysis by Absorbing Markov chain

The transition rate matrix R and the corresponding infinitesimal generator Q of the continuous-time Markov chain are given by

$$\mathsf{R} = \begin{pmatrix} 0 & k_1 & 0 & 0 & 0 & 0 \\ k_2 & 0 & k_3 & 0 & 0 & 0 \\ 0 & k_4 & 0 & k_2 & k_5 & 0 \\ k_4 & 0 & k_1 & 0 & 0 & k_5 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathsf{Q} = \begin{pmatrix} -q_1 & k_1 & 0 & 0 & 0 & 0 \\ k_2 & -q_2 & k_3 & 0 & 0 & 0 \\ 0 & k_4 & -q_3 & k_2 & k_5 & 0 \\ k_4 & 0 & k_1 & -q_4 & 0 & k_5 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

where $q_1 = k_1, q_2 = k_2 + k_3, q_3 = k_2 + k_4 + k_5$ and $q_4 = k_1 + k_4 + k_5$. The transition probability matrix P for the embedded discrete-time Markov chain is given by

$$\mathsf{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{k_2}{k_2+k_3} & 0 & \frac{k_3}{k_2+k_3} & 0 & 0 & 0 \\ 0 & \frac{k_4}{k_2+k_4+k_5} & 0 & \frac{k_2}{k_2+k_4+k_5} & \frac{k_5}{k_2+k_4+k_5} & 0 \\ \frac{k_4}{k_1+k_4+k_5} & 0 & \frac{k_1}{k_1+k_4+k_5} & 0 & 0 & \frac{k_5}{k_1+k_4+k_5} \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Our P matrix is so-called absorbing Markov chain [5]. The transition probability matrix of the absorbing Markov chain is of the general form written by

$$\mathsf{P} = \begin{pmatrix} \mathsf{P}_T & \mathsf{P}_{T,A} \\ \mathsf{0} & \mathsf{I} \end{pmatrix},$$

where I is the identity matrix of appropriate dimension. In our case, we have two absorbing states 5 and 6 and hence the dimension is two. The sub-stochastic matrix P_T corresponding to the transient states and $\mathsf{P}_{T,A}$ are given by

$$\mathsf{P}_T = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{k_2}{k_2+k_3} & 0 & \frac{k_3}{k_2+k_3} & 0 \\ 0 & \frac{k_4}{k_2+k_4+k_5} & 0 & \frac{k_2}{k_2+k_4+k_5} \\ \frac{k_4}{k_1+k_4+k_5} & 0 & \frac{k_1}{k_1+k_4+k_5} & 0 \end{pmatrix}, \quad \mathsf{P}_{T,A} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \frac{k_5}{k_2+k_4+k_5} & 0 \\ 0 & \frac{k_5}{k_1+k_4+k_5} \end{pmatrix}.$$

To compute the average number of times a specific transient state is visited before the absorbing states are entered, we define the fundamental matrix N by

$$\mathsf{N} = (\mathsf{I} - \mathsf{P}_T)^{-1}.$$

The (i, j) entry $N_{i,j}$ of N is the average number of times the absorbing Markov chain will be in state j before the absorbing states are entered, given that it started in state i. In our case, we can numerically compute N as follow:

$$\mathsf{N} = \begin{pmatrix} 1.7400 & 37.7400 & 44.2000 & 7.2000 \\ 0.7400 & 37.7400 & 44.2000 & 7.2000 \\ 0.7200 & 36.7200 & 44.2000 & 7.2000 \\ 0.7200 & 36.7200 & 44.2000 & 8.2000 \end{pmatrix}.$$

This means that the biochemical reaction starting from state 1 will, on average, be expected to be in state 2, where the FGF molecule bound with FGFR molecule, 37.74 times before reaching absorbing states.

Let n_i be the mean time before reaching the absorbing states starting from a specific transient state i, and let $\mathsf{n} = (n_i)$ be the column vector composed of n_i for i =1,2,3 and 4. In discrete-time Markov chains, the mean time is the mean number of steps. We can obtain n by multiplying the fundamental matrix N with a column vector $\mathsf{1}$ that contains only 1's as

$$\mathsf{n} = \mathsf{N1} = \begin{pmatrix} 90.8800 \\ 89.8800 \\ 88.8400 \\ 89.8400 \end{pmatrix}.$$

This tells us that it takes on average 90.88 time steps before reaching relocation states with the absorbing states starting from state 1. It should be noted that the total number of transitions in Table 1 is equal to $n_1 - 1$.

We can obtain the parallel continuous-time analogues. Let Q_T be the infinitesimal sub-generator of the continuous-time absorbing Markov chain. In our case, it is given by

$$\mathsf{Q}_T = \begin{pmatrix} -q_1 & k_1 & 0 & 0 \\ k_2 & -q_2 & k_3 & 0 \\ 0 & k_4 & -q_3 & k_2 \\ k_4 & 0 & k_1 & -q_4 \end{pmatrix}.$$

We define a matrix T whose (i, j) entry $T_{i,j}$ records the mean time that the continuous-time absorbing Markov chain will be in a transient state j before the absorbing states are entered, given that it started in a transient state i. Then it holds that $\mathsf{T} = (-\mathsf{Q}_T)^{-1}$. In our case, we can numerically compute T as follow:

$$\mathsf{T} = \begin{pmatrix} 3.4800 \times 10^{-4} & 3.7000 \times 10^{2} & 3.6000 \times 10^{3} & 1.4400 \times 10^{-3} \\ 1.4800 \times 10^{-4} & 3.7000 \times 10^{2} & 3.6000 \times 10^{3} & 1.4400 \times 10^{-3} \\ 1.4400 \times 10^{-4} & 3.6000 \times 10^{2} & 3.6000 \times 10^{3} & 1.4400 \times 10^{-3} \\ 1.4400 \times 10^{-4} & 3.6000 \times 10^{2} & 3.6000 \times 10^{3} & 1.6400 \times 10^{-3} \end{pmatrix}.$$

Let t_i be the mean time of the continuous-time Markov chain before reaching the absorbing states, given that it started from a transient state i. If we denote by $\mathsf{t} = (t_i)$ the column vector composed of t_i for i =1,2,3 and 4, then we have $\mathsf{t} = \mathsf{T1}$ and our example gives numerical values of t as

$$\mathsf{t} = \begin{pmatrix} 3.9700 \times 10^{3} \\ 3.9700 \times 10^{3} \\ 3.9600 \times 10^{3} \\ 3.9600 \times 10^{3} \end{pmatrix}.$$

It should be noted that t_1 is equal to the total expected time by PRISM in Table 2.

6. Probability Distributions Before Absorption

We analyze the behavior of absorbing Markov chains before absorption occurs. For this, we consider two types of probability distributions. First, we define a ratio between the mean time the Markov chain spends at each transient state and the mean time to absorption. The ratio gives us probability distributions depending on each initial transient state. Second, we consider so-called quasi-stationary distribution of absorbing Markov chains. In our case, the existence of the quasi-stationary distribution is guaranteed because the set of transient states is finite and irreducible.

6.1. Probability distributions by ratio of mean times

For transient states i and j of the discrete-time absorbing Markov chain, we define $p(j \mid i)$ by

$$p(j \mid i) = \frac{N_{i,j}}{n_i}.$$

It should be noted that n_i is finite and $n_i > 0$ for each transient state $i \in \{1, 2, 3, 4\}$. Because $\mathsf{n} = \mathsf{N1}$, we can interpret $\mathsf{p}(\cdot \mid i) = (p(j \mid i))$ as a probability mass function for each transient state $i \in \{1, 2, 3, 4\}$.

For transient states i and j of the continuous-time absorbing Markov chain, we define $q(j \mid i)$ by

$$q(j \mid i) = \frac{T_{i,j}}{t_i}.$$

As is the case of the discrete-time Markov chain, we can interpret $\mathsf{q}(\cdot \mid i) = (q(j \mid i))$ as a probability mass function for each transient state $i \in \{1, 2, 3, 4\}$.

6.2. Quasi-stationary distributions

We briefly introduce the quasi-stationary distributions of absorbing Markov chains. For the quasi-stationary distributions, we refer the reader to the seminal work [6, 7].

Let x be the quasi-stationary distribution of P_T on the transient states for the discrete-time Markov chain, under the condition that absorption has not yet taken place. It is known that x is given by the left eigenvector of P_T for the eigenvalue ρ with maximal real part, that is,

$$\mathsf{x}\mathsf{P}_T = \rho\mathsf{x}.$$

Because P_T is a nonnegative and irreducible matrix in our case, the eigenvalue ρ is the Perron–Frobenius eigenvalue of P_T. Since P_T is sub-stochastic, we must have $\rho < 1$. We can obtain the left eigenvector x as a unique strictly positive vector such that $\mathsf{x1} = 1$.

Similarly, let y be the quasi-stationary distribution of Q_T on the transient states for the continuous-time Markov chain, under the condition that absorption has not yet taken place. It is also known that y is given by the left eigenvector of Q_T for the eigenvalue $-\alpha$ with maximal real part, that is,

$$\mathsf{y}\mathsf{Q}_T = -\alpha\mathsf{y}.$$

Because Q_T is irreducible, $-\alpha$ is unique and simple, and we must have $-\alpha < 0$. It also holds that y can be chosen strictly positive vector such that $\mathsf{y1} = 1$.

7. Comparison of Probabilities Computed by PRISM and Those of Quasi-stationary Distributions

Table 4: Comparing probabilities of visiting states 1–4 computed by PRISM with quasi-stationary distributions

State	PRISM			Qusasi-stationary distributions
	Transitions	Visits	Probability	Probability
1	0.7400	1.7400	0.0191	0.0082
2	37.7400	37.7400	0.4153	0.4157
3	44.2000	44.2000	0.4864	0.4946
4	7.2000	7.2000	0.0792	0.0815
Total	89.8800	90.8800	1.0000	1.0000

Table 5: Comparing probabilities of staying at states 1–4 computed by PRISM with quasi-stationary distributions

State	PRISM		Qusasi-stationary distributions
	Time	Probability	Probability
1	3.4800×10^{-4}	0.0000	0.0000
2	3.7000×10^{2}	0.0932	0.0911
3	3.6000×10^{3}	0.9068	0.9089
4	1.4000×10^{-3}	0.0000	0.0000
Total	3.9700×10^{3}	1.0000	1.0000

As can be seen in Tables 4 and 5, the probability of visiting and staying at any states before relocation occurs obtained by model checker PRISM and quasi-stationary distributions are approximately the same.

8. Concluding Remarks and Further Work

In this research, we studied probabilistic model checking [3] for continuous-time Markov chains, and how to analyse this model with model checker PRISM [4]. We worked with a model which contains one molecule each of FGF and FGFR of which signalling mechanisms are not well understood and several hypotheses exist. We also showed how to model a molecular interactions with different stochastic rate using model checker PRISM. We considered two alternative approaches [2]: individual-based model and population-based model.

We analysed the population-based model when there are one of each FGF and FGFR molecule. By comparing the results of PRISM with those of quasi-stationary distribution over absorbing Markov chains, we obtained the probability of visiting and staying at any states before relocation occurs in both techniques are approximately the same. In PRISM, we can specify the initial state and compute properties such as the time spent in a given state or how often a certain state transition has taken by using a feature named "reward", however we need to put labels to every transitions and states, which makes the model become more complex and easily meet the state-space explosion problem as large numbers of reactions in the system occur. In contrast, in quasi-stationary distributions, we cannot specify the initial state of the system, however we can define numbers of visits or time of staying and the approximated probability in all states in the system at a time. For justifying this approximation method, it would be worthwhile to investigate that the quasi-stationary distributions can provide an upper bound for the probability mass function in terms of stochastic ordering.

As our future work, we want to work on more further and detail about Probabilistic model

checking of larger numbers of molecules of biological system and work on the combination of quasi-stationary distribution and PRISM to compute the system properties without giving labels in each transition.

References

[1] M. Kwiatkowska and C. Thachuk: Probabilistic Model Checking for Biology. In Software Safety and Security, IOS Press, pp. 165–190, 2014.

[2] M. Kwiatkowska, G. Norman and D. Parker: Probabilistic Model Checking for Systems Biology. In M. Sriram Iyengar (editor), Symbolic System Biology, pp. 31–59, Jones and Bartlett, 2010.

[3] M. Kwiatkowska, G. Norman and D. Parker: Stochastic Model Checking. In M. Bernardo and J. Hillston (editors), Formal Methods for the Design of Computer, Communication and Software Systems: Performance Evaluation (SFM'07), LNCS Vol. 4486, pp. 220–270, 2007.

[4] PRISM website: http://www.prismmodelchecker.org

[5] D. J. Barnes and D. Chu: Introduction to Modeling for Biosciences. Springer, 2010.

[6] J. N. Darroch and E. Seneta: On quasi-stationary distributions in absorbing discrete-time finite Markov chains. Journal of Applied Probability, Vol. 2, pp. 88–100, 1965.

[7] J. N. Darroch and E. Seneta: On quasi-stationary distributions in absorbing continuous-time finite Markov chains. Journal of Applied Probability, Vol. 4, pp. 192–196, 1967.

[8] C. Baier and J-P. Katoen: Principle of Model Checking. The MIT Press, 2008.

[9] K. Fujita: Theorem Proving and Model Checking. Proc. of the 2016 IEICE General Conference, AI-3-1, SS-30–31, 2016.