STOCHASTIC SIMULATIONS OF BIOLOGICAL NETWORKS UNDER IMPULSES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MÜGE YAZICI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
STATISTICS

APRIL 2016

Approval of the thesis:

STOCHASTIC SIMULATIONS OF BIOLOGICAL NETWORKS UNDER IMPULSES

submitted by **MÜGE YAZICI** in partial fulfillment of the requirements for the degree of **Master of Science in Statistics Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver _____
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Ayşen Akkaya _____
Head of Department, **Statistics**

Assoc. Prof. Dr. Vilda Purutçuoğlu _____
Supervisor, **Statistics Dept., METU**

Examining Committee Members:

Assist. Prof. Dr. Ceylan Talu Yozgatlıgil _____
Statistics Dept., METU

Assoc. Prof. Dr. Vilda Purutçuoğlu _____
Statistics Dept., METU

Assoc. Prof. Dr. Yeşim Aydın Son _____
Medical Informatics Dept., METU

Assist. Prof. Dr. Ceren Vardar Acar _____
Statistics Dept., METU

Assist. Prof. Dr. Özlem Defterli _____
Mathematics - Computer Dept., Çankaya University

Date: 01.04.2016

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Müge Yazıcı

Signature:

ABSTRACT

STOCHASTIC SIMULATIONS OF BIOLOGICAL NETWORKS UNDER IMPULSES

Yazıcı, Müge

M.S., Department of Statistics

Supervisor: Assoc. Prof. Dr. Vilda Purutçuoğlu

April 2016, 88 pages

The impulses are one of the sources of fluctuations which lead to the sharp changes in the biochemical systems such as the changes in epidemic and population models. In this study, as the novelty, we implement these effects in exact stochastic simulation algorithms, namely, the Gillespie method and the first reaction method, by changing the chemical master equations (CME) with impulses. Hereby in the extension of these approaches with impulses, we consider different scenarios such as the sudden decrease of the number of molecules. Finally, we evaluate each of these alternatives under small and realistically large biochemical systems.

Keywords: Gillespie algorithm, chemical master equations, impulses.

ÖZ

İMPULSLAR ALTINDA BİYOLOJİK AĞLARIN STOKHASTİK SİMÜLASYON ALGORİTMALARI

Yazıcı, Müge
Yüksek Lisans, İstatistik Bölümü
Tez Yöneticisi: Doç. Dr. Vilda Purutçuoğlu

Nisan 2016, 88 sayfa

İmpulslar biyokimyasal sistemlerde salgın ve popülasyon modellerindeki değişimler gibi ani değişikliklere sebep olan temel kaynaklardan biridir. Bu çalışmada yenilik olarak Gillespie metod ve ilk reaksiyon metodu olarak bilinen tam stokastik simülasyon algoritmalarını, temel kimyasal denklemini (CME) impulsları içerecek şekilde uygulacağız. Gillespie algoritması, değişikliklerin etkisi olmaksızın temel kimyasal denklemler yardımıyla gerçek biyokimyasal sistemin rassal yapısını yaratabilen, bu alandaki en iyi bilinen metottur. Dolayısıyla bu metodun impulslar ile geliştirilmesinde, molekül sayısındaki ani artış impulsların bulunması gibi farklı senaryoları düşünmekteyiz. Sonuç olarak ise bu alternatiflerin her birini, küçük ve büyük sistemler altında değerlendirmekteyiz.

Anahtar Kelimeler: Gillespie algoritması, temel kimyasal denklemleri, impulslar.

To my husband, Recep, my mother, father and grandmother…

ACKNOWLEDGEMENTS

I would like to express my special thanks to my advisor, Assoc. Prof. Dr. Vilda Purutçuoğlu for her guidance. She is always consciousness and patience to me during my graduate and undergraduate period. Because of her support, I could be able to complete this study.

I would also like to thank my husband, Recep Ergül. He is always support, trust and encourage me. Without his love, I could not finish the thesis.

My family, Nuray Yazıcı, Hüseyin Yazıcı and Fatma Eral are really helpful to me. I am grateful to them for their special support.

I would thank to the Scientific Research Project (Project no: BAP-08-11-2014-007) of METU entitled as "Simulation of biochemical systems via stochastic simulation algorithms and extensions of these methods by including impulses" for its support.

I am very much thankful to Dr. Derya Altıntan (Department of Mathematic at Selçuk University) for her support and help in the computational part of this study.

I would thank to Assist. Prof. Dr. Ceylan Talu Yozgatlıgil for her support. She gives me experience to think detail and consider different angle.

I would appreciate to Assoc. Prof. Dr. Yeşim Aydın Son, Assist. Prof. Dr. Ceren Vardar Acar and Assist. Prof. Dr. Özlem Defterli to support and encourage of me. They are quite positive to me and gives me experience to evaluate results differently.

No forgetting my former manager, Tuna Doğan for his encourage of me. He is always understanding to me during my lecture period.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

CME: Chemical Master Equation

DNA: Deoxyribonucleic Acid

RNA: Ribonucleic Acid

DR: Direct Method

FR: First Reaction Method

TF: Transcription Factor

RNAP: The RNA-polymerase

RNase: Ribonuclease

SSA: Stochastic Simulation Algorithm

CPU: Central Processing Unit

DE: Differential Equation

SD: Average Search Depth

WD: Average Weighted Degree

CPU: Central Processing Unit

ODM: Optimal Direct Method

SDM: Sorting Direct Method

IFN: Interferon

TFBS: Transcription Factor Binding Site

CP: Cytoplasmic Phosphatase

NP: Nuclear Phosphatase

CHAPTER 1

INTRODUCTION

The biochemical modeling is the mathematical way of explaining the dynamic of biological systems such as the interaction of proteins and molecules. The theory of the biochemical systems was improved by Michael Savageau in 1960s. By understanding the effects, complexity and the importance of the interactions in the systems, the necessity of the computer has become crucial because the description of the cellular activations and the analyses of the data have turn into very complicated and costly calculations. Accordingly, the idea of analyzing these chemical reactions by deterministic methods has been concerned for many years. The advantage to study the deterministic methods is its explicit form of model parameters and its simplicity in calculation. Hereby, as the model parameters, it uses the mass action kinetics. The mass action assumes that the reactions are continuous and deterministic and the researcher needs a low level of information for the description of the biochemical systems. The stochastic modeling has been more researched in recent years and found more realistic to describe in vivo reactions (Turner, 2004). The stochastic processes are random, resulting in slightly different outputs in every application, particularly, in simulation because in this process, the changes in the system occur probabilistically. Furthermore, different from other biochemical models, the main interest in stochastic choices is the changes in the number of molecules of the system's elements under a homogenous environment (Bower and Bolouri, 2001).

Thereby, in stochastic model, the random fluctuations are considered when the volume of the system is small and the molecular population is low. The main fluctuation is discovered by using fluorescent probes. On the other hand, the impulses are one of the main items of fluctuations which cause sharp changes in biological systems and in this study; we add the impulses in stochastic simulations of

different dimensional systems to detect its effect in the calculation and validity of the results.

1.1 Aim of study

The impulses are one of the sources of fluctuations which lead to the sharp changes in the biochemical systems such as the changes in epidemic and population models. In this study, we use the Gillespie and the first reaction methods which are the most common exact stochastic simulation algorithms to implement the effects of impulses. The Gillespie algorithm and its extension, the first reaction algorithms are the two well-known methods which can generate the random nature of the actual biological system by means of chemical master equations (CME) without the effect of sharp fluctuations.

In the analyses, we consider different scenarios in the application of impulses to the systems such as impulses at fixed states of the species and at fixed periodic time. Then, we evaluate each of these alternatives under small and realistically large biochemical systems in terms of the computational demand. Later, we also perform the same analyses via the first reaction algorithm. Finally, we compare both outputs in order to assess the capacity of these methods with/without impulses.

1.2 Motivation

The model deals with the states of elements and interactions to obtain reliable information about the system. Regarding the given details about the system such as for the given reaction rates, the number of molecules or the amount of concentration of species, the model can explain the entire organism at different levels.

On the other side, the description of the biological system is challenging because of its complexity. Because the vivo/vitro analyses of the biological systems can be expensive and limited. In this situation, we can perform the simulation method so that we can visualize the behavior of the real system artificially but under realistic assumption. Furthermore, as we have a chance to alter the inputs of the systems in every simulation, we enable to predict the activation of the system under new conditions. By this way we can better detect the crucial components of the system,

their presumed effects and finally find new biologically interesting questions about the system.

Hereby, in this study, we aim to compare the performance of the stochastic simulation algorithms of the biological systems and to evaluate their performances under abrupt changes. In real systems, such abrupt changes are called as impulses and in biological networks; these changes are performed in the description of cancer, diabetes or similar diseases whose attacks are observed suddenly and stochastically.

Therefore, in the application, we choose distinct dimensional and actual systems. In the calculation, we apply the JAK-STAT pathway under realistic complexity and the Lotka Voltera model as the toy example. We also investigate the PKC Pathway and Lysis–Lysogeny model as moderate networks. Finally, we use the combination of the JAK-STAT and PKC pathways as a single and large pathway. In the simulation, we perform the Gillespie and the First Reaction Algorithms which are the two well-known stochastic simulation algorithms. In both methods, the assumption is that the molecules of the species, which constitute a system, mix well in a constant volume and under a constant temperature. Indeed, we decrease/increase suddenly the number of molecules for each model and search the effect of these sudden changes in the computational demand to decide on the best exact algorithm under impulses.

CHAPTER 2

LITERATURE BACKGROUND

2.1 Biological Modeling

A biological model is a biological description constructed to explain the biological behavior of a system whose major elements are species, proteins, or genes through time and their interactions between each other (Wilkinson, 2006). In terms of biological description, it is always better to deal with detailed and precise model since it provides us to obtain reliable results which can validate the actual behavior of the biological process. Hereby, the modeling tools are needed to explain the complex structure of molecular and cellular interactions. Moreover, there is a strong relation between modeling and experiment as well. For instance, after constructing a model, it is tried to obtain experimental results which can be validated by the suggested model. In fact, the modeling and the computer simulation are corporate in order to improve predictions about the complex biological systems.

Thereby, a mathematical model in biology aims to describe a biological/biochemical process in a numerical and more explicit way. In such description, we typically need to explain the major activation in the system that can be listed as the translation, transcription, gene regulation, cellular signaling, DNA damage and repair process, homeostatic processes, cell cycle and the apoptosis. By this way, the function of a tissue, organ and even the entire organism at a higher level can be described under certain conditions (Wilkinson, 2006; Bower and Bolouri, 2001).

Accordingly, the simulation of a system can help us to generate the behavior of the data artificially. Hereby by simulating the model, we can suggest a prediction of the interested system. On the other hand, if we do experiments to obtain the real data, then we can have a chance to compare the actual data with simulated measurements.

If both results are conformed, we can rely on the model (Wilkinson, 2006; Purutçuoğlu, 2010).

On the other hand, although the modeling has many advantages, there can be some problems in the sense that it can be hard to explain the entire organisms due to the actual complexity in the system. For instance, the actual organisms have detailed variables for each individual cell and the suggested model cannot have ability to include all such gene specific details in a single mathematical expression (Bower and Bolouri, 2001).

Hereby, below, we initially describe several well-known definitions such as DNA, RNA, and protein, which are commonly used in biology and then present main approaches in mathematical modeling of biological processes. Later, we combine these explanations within the simulation of the systems.

2.1.1 Major Biological Terms

In this part, first of all, we represent certain major components/elements in all algorithms. These are DNA, RNA, receptor, ion channel, enzymes and transcription factor.

Then, we explain the responsibility of each element in the regulation of the gene. Here, we briefly describe the translation and the transcription process.

DNA, deoxyribonucleic acid, is a cell material including all genetic information, called genes. It is allocated in the nucleus of all eukaryotes, whereas, it is placed in the cytosol for the prokaryotic cells. Each cell of the organism has the same DNA and it is composed of four chemical bases, which are adenine (A), guanine (G), cytosine (C), and thymine (T). Each base should be paired certain base such that the adenine is matched with thymine and the guanine is paired by cytosine. DNA can replicate itself and after the cell division, there exists one more DNA chain having the same genetic information with the mother cell (National Institutes of Health, 2014).

On the other hand, *RNA, ribonucleic acid,* is the other main macromolecules of the cells. RNA is a transmittor between DNA and proteins and provides to make a

protein. In fact, RNA is produced in the nucleus for eukaryotes. Then, it goes out to cytosol and is transcribed there. After it attaches ribosomes that represent the sites of cohics translation of the messenger RNA and protein, it is translated into proteins. DNA and RNA look like each other. But there are some small structural differences between them in the sense that RNA has a single-strand, on the contrary, DNA is double-stranded. In addition, the uracil base in DNA is substituted for the thymine base in RNA (The RNA Society, 2014).

The gene is a specific type of proteins having two regions, namely, the coding region and regulatory region. The protein is encoded in coding region while the regulatory region is related part for the control of the gene. The receptor, ion channel and enzymes describe different types of proteins. The amino acids construct proteins and each protein has different functions due to its distinct amino acid sequence (Bower and Bolouri, 2001).

Finally, *the transcription factor, TF*, is a class of proteins, which is responsible for binding to the DNA sequences and controls the transcription of the genetic information from DNA to mRNA. Furthermore, it can affect the gene expression alone or with other proteins positively (as an activator) and negatively (as a repressor) (Wilkinson, 2006).

By using these listed biological cellular elements, we can describe the major functions in a system. These are the transcription and the translation that are also known as the *central dogma*. *The transcription* is a process that initiates the transport of the genetic information from DNA to mRNA. This step starts when TF binds to DNA. The prokaryotes have basic transcription states than eukaryotes in the sense that their biological process can be described by few reactions. On the other hand, as the eukaryotes have more complex states, their transcription's description becomes complex too. The RNA-polymerase (RNAP), which refers to an enzyme producing RNA, originates this process. RNAP fastens the activation by attaching the up-stream of the interested gene which is called the promoter region. After this step, *the translation* process begins and mRNA binds to ribosomes by three bases at a time. The underlying three-base combination is called *the codon*. In fact, a codon is translated into one of 20 amino acids. Therefore, the protein is produced in this step.

On the other hand, *the transportation* is a process between transcription and translation. In this process, mRNA is moved from nucleus to cytoplasm.

If the mRNA transcript of a gene of interest is transformed to nothing, this is named as *the degradation*. The ribonuclease, also denoted by RNase, is an RNA-degrading enzyme (Wilkinson, 2006).

The transcription process can be modeled as follows:

$$g+TF1 \leftrightarrow TF1.g \quad (1)$$

$$TF1.g+TF2 \leftrightarrow TF2.TF1.g \quad (2)$$

$$RNAP+TF2.TF1.g \leftrightarrow RNAP.TF2.TF1.g \quad (3)$$

$$RNAP.TF2.TF1.g \rightarrow TF2.TF1.g+RNAP+r \quad (4)$$

In Equation 1 and 2, g, TF1 and TF2 represent the gene, transcription factor 1 and 2, in order. They are single entities in the model. RNAP binds to the combination of gene and transcription factors and initiates the transcription process. TF2.TF1.g and r denote the complete proteins and the mRNA transcript of the gene (Wilkinson, 2006). Finally, the two-sided arrow between reactions denotes the reverse reaction between both sides of the given reaction.

2.1.2 Chemical Reactions

The change and speed of a system can be expressed with chemical equations. The chemical reaction, which is briefly shown as below, is a way to express a biological activation so that it can be represented by a mathematical description.

$$m_a A \xrightarrow{k} m_b B \quad (5)$$

where m_a and m_b are called *the stoichiometric coefficient* of the *reactant* and *product*, respectively. In this description, the reactant shows the species written in the left hand side of the arrow and the product indicates the species given on the right hand side of the arrow. For Equation 5, the reactant and the product are presented by

A and *B*, in order. Here, *k* is the *rate constant* of the reaction which is dependent on the temperature and the volume of the environment. This value is a representation of the time for the execution of a reaction. Hereby, Equation 5 implies that m_a molecules of *A* reacts to produce m_b molecules of *B* while the molecules move with respect to the Brownian motion (Turner, 2004; Bower and Bolouri, 2001).

The *petri nets* are mathematical explanation of models with a graphical representation. These systems are simply described as below:

Petri Net = (P, T, Pre, Post, M) where P is a finite set of places, T is a finite set of transitions and M is the current state of the system (Wilkinson, 2006). Accordingly, if we consider the following model for the genes *A*, B_2, AB_2 and *C*, we can define a system that can be also represented as in Table 1.

$$A + B_2 \leftrightarrow AB_2 \quad (6)$$

$$A \rightarrow A + C \quad (7)$$

$$C \rightarrow C + B \quad (8)$$

Table 1: Description of the system in Equation 6-9 in order, pre and post terms

	Reactants (Pre)					Products (Post)				
	AB_2	A	C	B	B_2	AB_2	A	C	B	B_2
Reaction 1 **(Equation 6)**		1			1	1				
Reverse Reaction of Reaction 1	1						1			1
Reaction 2 **(Equation 7)**		1					1	1		
Reaction 3 **(Equation 8)**			1					1	1	
Reaction 4 **(Equation 9)**				2						1

The reactant (pre) and products (post) matrices of this model are shown in Table 1 and the difference between these matrices, which is called as the *net effect matrix*, is represented as below.

$$\text{Net effect} = \begin{bmatrix} 1 & -1 & 0 & 0 & -1 \\ -1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -2 & 1 \end{bmatrix}. \tag{10}$$

As seen beforehand, there can be more than one reactant and product in a reaction. For instance,

$$m_a A + m_c C \xrightarrow{k} m_b \tag{11}$$

indicates a reaction with three substrates, namely, *A, B* and *C* with m_a, m_b, and m_c stoichiometric coefficients, respectively. Hence, it presents that the m_a amount of *A* molecules reacts with the m_c amount of *C* molecules to form the m_b amount of molecule *B* (Wilkinson, 2006).

Also the system is called *equilibrium* if there is no change in the number of molecules. In other words, all of the inputs are used and only the outputs are presented in the system.

Finally, if the biological activation is explained via a list of reactions, it generates a network whose genes are the elements of the reaction list (Bower and Bolouri, 2001). In the following part, we initially describe the major mathematical methods to model these sets of genes and then, combine this idea in the simulation of networks under stochasticity.

2.2 Modeling Approaches of Biological Systems

There are three major approaches in modeling the biological systems. These are:

i. *Boolean models,*
ii. *Deterministic models,*
iii. *Stochastic models.*

The “Boolean models” represent the state as a list of positions while “Deterministic models” present it as a list of concentrations. Finally, the “Stochastic models” deal with the list of the actual number of molecules. More details about these major approaches are given in the following parts. On the other hand, apart from these main techniques, there exists some alternative methods which lie between the Boolean model and the deterministic approaches as well as between the deterministic and the stochastic choices.

Table 2: The main differences of the deterministic and stochastic modelings.

	Deterministic Models	**Stochastic Models**
Equations	Based on the ordinary differential equations	Based on the chemical master equations
Representation of the amount of species	Concentration amount of the species	Number of molecules of the species
Parameters of the models	Reaction rates	Stochastic reaction rate constants

2.2.1 Boolean Models

The Boolean models represent a state with 0 and 1 values in such a way that 0 means not-expressed and 1 displays the expressed gene. Actually, it can be thought like a dummy variable and deals with whether the gene is active or inactive. Accordingly, for a system with n elements, the system contains second possible states and we can express a truth table indicating the changes in the system for every chance in states for small biological networks because all possible next states can be predicted in small systems. Moreover, it can be expressed using the operators AND (1 if all inputs are 1), OR (1 if any of the inputs are 1) and NOT (1 if all inputs are 0) (Bower and Bolouri, 2001; Jong, 2002).

2.2.2 Models of Differential Equations

The differential equation (DE) models assume that the states are continuous and they change over time by the change in the concentration of the genes as follow.

$$C \xrightarrow{k} C' \tag{12}$$

In this equation, *C* describes the reactant and *C′* is the product. In this expression, the concentration of *C′* increases while the concentration of C decreases. Hereby, the derivative of the concentrations for *C* is equal to the multiplication of concentration of *C′* and gives the changes in the concentration of *C* during the reaction as shown below. Finally, the meaning of minus sign presents the decrease in the *C* concentration.

$$\frac{d[C]}{dt} = -k[C] \tag{13}$$

On the other hand, if we write this expression to describe the change in product *C′*, it becomes

$$\frac{d[C\prime]}{dt} = k[C] \tag{14}$$

since the concentration of *C* raises continuously. From Equation 13 and 14, it is seen that the rate constant depends only on the reactants. If there are n reactants, the derivative becomes the multiplication of *k* and *n* reactants and the sign is minus, whereas, the sign of products is taken as positive. For instance

$$A + B + \cdots + N \longrightarrow AB \ldots N \tag{15}$$

can be written as

$$\frac{d[A]}{dt} = -k[A][B] \ldots [N] \tag{16}$$

…

$$\frac{d[A]}{dt} = -k[A][B] \ldots [N] \tag{17}$$

and finally,

$$\frac{d[AB\ldots N]}{dt} = -k[A][B]\ldots[N] \quad (18)$$

If the reactant or product has a form of multiplication with a constant, it can be written as a multiplication of concentrations and constants in the derivative form too. On the other side, if the reactant or the product has an index, the index can be written as a power of concentrations in the derivative. For instance,

$$2A + 3B \longrightarrow A_2B_3 \quad (19)$$

can be converted as the following set of differential equations.

$$\frac{d[A]}{dt} = -6k[A]^2[B]^3, \quad (20)$$

$$\frac{d[B]}{dt} = -6k[A]^2[B]^3, \quad (21)$$

$$\frac{d[A_2B_3]}{dt} = k[A]^2[B]^3. \quad (22)$$

So the underlying relation in the rate of reaction which is the reactant concentration with respect to the power of stoichiometry, is called as the mass action kinetics. In general, the DE modeling explains successfully the gene regulation of the system (Wilkinson, 2006; Bolouri, 2008; Jong, 2002).

2.2.3 Stochastic Models

The stochastic simulation provides to understand the distribution related to time and gives information about the shape of data. In fact, stochastic methods explain the changes in the number of molecules in a system by using a probabilistic approximation. By applying the Monte Carlo technique, the system simulation can be also done to find the statistics. Accordingly, a reaction can be shown as in Equation 23,

$$A \xrightarrow{c} B. \quad (23)$$

In Equation 23, it can be defined that the probability of *B* molecules generated from molecule *A* is proportionally found by $c \times dt$. Here, *c* is the stochastic reaction rate constant and *dt* shows the change in time. In addition, the number of molecule *B* at

the end of the reaction is computed as *c[A]dt* where [.] indicates the number of molecule for a gene.

Finally, this modeling type is successful in explaining the stochastic behavior of a system like the transcription and the translation (Turner, 2004; Wilkinson, 2006).

2.2.4 Other Models

In addition to the Boolean models, differential equations and stochastic models which are three main models in the gene regulation, there are also some models to express biological systems, called hybrid methods. They are the kinetic logic model, continuous logical model, Langevin approach and the Fokker-Planck equation.

The kinetic logic model places between the Boolean and the differential equation (DE) methods by expressing the states as discrete values. Actually, it is one step along Boolean models. The state of each gene of this model is expressed as a discrete value. Under this model, it is assumed that all the states do not change at the same time, which is also known as the asynchronicity property, and the change in the system is shown by partial linear/nonlinear functions for a given threshold. So unlike the Boolean approach, there are more than two states such as low, nearly low, strong, and nearly strong levels. Thereby, it needs more complicated information about the data and the changes of the states are independent on each other. In addition, this model needs more data than Boolean models. Therefore, it provides reliable predictions.

The continuous logical model generates models between the kinetic logic and DE models and on comparison with the kinetic logic method, the states are still discrete, whereas, it requires more detailed data. In fact, DE is used to describe the transition from one state to another.

The Langevin approach is a method between DE and stochastic models. Different from DE, a noise is added to the model where the noise, i.e. error term, has zero mean and no correlation as shown in Equations 24 and 25. *e(t)*, noise function, is random and does not depend on the current state.

$$dc/dt = f(c), \tag{24}$$

$dc/dt = f(c) + e(t).$ (25)

In Equation 24, the given expression indicates DE modeling, whereas, Equation 25 describes the Langevin approach of the same expression. Here, *e(t)* denotes the *random noise term* and it is a function of time. The error term is not dependent on the previous state because of its no-correlation.

The Fokker-Planck equation deals with both discrete and continuous variables. Thereby, at the beginning of modeling, it starts by a stochastic model and then, it transforms the model to the differential equation. The change in the probability distribution is the main interest of this equation and this change is continuous and has a discrete distribution based on concentration of molecules (Bower and Bolouri, 2001).

2.3 Mass-Action Stochastic Kinetics

To better understand the biological processes, the associated reactions need to be understood. So the biochemical modeling is a helpful method for making clear the reaction of proteins and molecules.

Hereby, as described in Section 1.2, if there is a system of reactions included a species and r reactions (r>a), the qualitative structure of the reaction network can be denoted by N= (P, T, Pre, Post, M) which is called as the Petri Net. In this representation, P shows the set of species and T is the set of reactions. Pre presents the path from places to transitions and Post is the path from transitions to places. Finally, M denotes the current state of system as used beforehand.

On the other hand, each reaction has a hazard function whose components are the current state x and the rate constant c, i.e , hi(x,ci) for the ith reaction. Regarding the type of the reaction, this function is computed in different ways. Below, we list most widely used reaction types and their hazard functions.

1) Hazard function is equal to the constant rate in the *zeroth-order reactions* as shown in Equation 26 for the product gene *X* and the reaction rate *c*. Hence, the hazard of this reaction is found as *h(x,c)=c*.

   $\Phi \xrightarrow{c} X.$ (26)

2) The hazard function is a combination of a constant rate and a molecule *X* in the *first-order reactions* as presented in Equation 27. Then, the associated hazard is calculated via *h(x,c)=cx*.

$$X \xrightarrow{c} ? \tag{27}$$

3) If a particular pair of molecules reacts, the hazard is the combination of the pair of molecules and the constant rate as described below. The hazard for Equation 28 is computed as *h(a,b,c)=abc* and for Equation 29 if is found as $h(a) = \frac{a(a-1)}{2}$.

$$A + B \xrightarrow{c} ? \tag{28}$$

$$2A \xrightarrow{c} ? \tag{29}$$

4) In higher order reactions as represented below, the calculation of the hazard is obtained as follows:

$$kX \xrightarrow{c} ? \tag{30}$$

where *c* is the rate constant and *k* is the stoichiometry coefficient of *x*. Hence, the hazard is calculated by$h(x, c) = c\binom{x}{k}$ for Equation 30.

2.4 Computational Cost Analysis

In other to assess the performance of different simulation methods in terms of the computational demand, certain measures are suggested in the literature. Among alternatives, the most widely used ones can be listed as the average search depth, average weighted degree and the central processing unit. In our analyses, we also choose these measures to compare the findings of distinct simulation algorithms. Below we define each with more details.

Average search depth: A biochemical system has many reactions such as *r* = 1,2,…R. Here, *r* shows which reaction happens in the system. Accordingly, the *average search depth*, denoted by *SD*, indicates the average number of operations by following expression.

$$SD=\frac{\sum_j^R j\beta_j}{\sum_j^R \beta_j} . \tag{31}$$

In the equation above, *j* represents which reaction occurs during the simulation and β_j stands for the time of the occurrence of the *j*th reaction. Apart from *SD*, some systems apply the dependency graph before the simulation.

Average weighted degree: The *average weighted degree*, shown by *WD*, calculates the average degree for the dependency graph for each occurring reaction. This term is related to the nodes and the pathway between edges of nodes. Here, the *degree* implies how many reactions are affected by the occurrence of the reaction. Thus *WD* is computed by the equation below.

$$WD=\frac{\sum d_j\beta_j}{\sum \beta_j} \tag{32}$$

in which d_j represents the degree of the *j*th node and β_j describes the number of firing times of the *j*th reaction as used in expression of *SD*.

Central processing unit: The *central processing unit* (CPU) is the necessary unit to understand and translate the program codes. RAM (shared memory) has many CPUs connections such that they can access each private memory and are independent on each other. Due to this feature, CPUs can be used to compare the performance of systems.

2.5 Stochastic Simulation Algorithms

The gene regulation is generally modeled by deterministic mathematical formulation (DE) in detail. This model defines the chemical species as continuous and chemical interaction as DE (McCollum and Peterson, 2006). However, the DE method does not give effective results on the molecular level. Therefore, the stochastic algorithm is used to model the biochemical systems. This algorithm defines species as discrete values and chemical interactions as random processes. The stochastic simulation algorithm (SSA) of the biochemical reaction, which is exact and homogenous, is based on the *master equations,* also named as the *Chapman-Kolmogorov equation*.

This algorithm analyzes each reaction stochastically under a time evaluation of species' populations. Here, the main assumption is the system in thermal equilibrium. Actually, these equations are used in small systems while Monte Carlo is preferred in large systems. Because the master equations cannot be solved numerically when the system is complex. In the representation of this type of equations, let's *Y* be the number of molecules in the reactant parts in a reaction. Moreover, assume that the system has *r* reactions and *v* known as the state change vector denote the net effect matrix. Hence, the probability of occurrence of the *j*th reaction while changing state from (Y-v_j) to (Y) at time interval [t,t+dt] is described by

$$\frac{\partial P(Y,t)}{\partial t} = \sum_{j=1}^{r} \{ h_j(Y - v_j) P(Y - v_j, t) dt - h_j(Y) P(Y, t) dt \}. \quad (33)$$

In Equation 33, $P(Y,t)$ shows the probability for Y amount of molecules in the time t and dt denotes the changes in time. Accordingly, $h_j(Y\text{-}v_j)$ presents the hazard function for the state (Y-v_j) when the net effect is described by v_j for the jth reaction ($j=1,\ldots$ r).

There are three methods to stochastically generate a system. The main aim of these algorithms is to answer two questions. The first interest is when the next reaction occurs and the second one is which reaction will occur next. These methods are similar in accuracy and deviation since they are all exact simulation algorithms. But they differ within each other in terms of their computational time. In the following part, we explain each of these algorithms in details (Wilkinson, 2006).

2.5.1 Gillespie Algorithm

The Gillespie algorithm, also called as *the direct method*, is the fastest simulator in the calculation of time, especially, for small systems (Wilkinson, 2006, Gillespie, 1977; Turner, 2004). However, its application is hard in heterogeneous situation and not computationally efficient in simulation of large networks as the calculation takes long time.

The Gillespie algorithm is known as discrete event simulation and its procedure is as follows for a system with *r* reactions and *n* species. The system at time $t=0$ with rate

constants c_1, c_2,....,c_r and the molecule numbers for each species x_1, x_2,....,xn are initialized.

i) The hazard functions, $h_i(x,c_i)$ are calculated for each reaction $(i=1,...,r)$.

ii) The combined reaction hazard is calculated via $h_0(x,c)=\sum_{i=1}^{r} h_i\,(x, c_i)$.

iii) The time interval for the next event, i.e, reaction, is generated from the exponential density with rate h_0, i.e. $Exp(h_0(x,c))$.

iv) The reaction is simulated with probabilities $h_i(x,c_i)/h_0(x,c)$ under the assumption that the reactions are independent of each other.

v) The outputs x and t are updated via $x:=x+S^{(j)}$ and $t:=t+t'$ respectively. Here $S^{(j)}(j=1,...,n)$ indicates the jth column of the stoichiometry matrix and t denotes the time interval. In the stoichiometry matrix, the rows show the reaction and the columns denote the number of species. Hence, it is a $(r\times n)$ matrix for totally r reactions and n species $(r\times n)$.

vi) If $t<T$, the algorithm is repeated from step (ii) until $t\geq T$, where T is the predetermined total time.

vii) The hazard functions, hi(x,ci) are calculated for each reaction (i=1,…,r).

2.5.2 First Reaction Method

The first reaction method is a simulation method which uses shorter calculation and is based on the Gillespie method (Gillespie, 1992; Turner, 2004). Hereby, it has advantage over Gillespie for the complicated systems.

The step of this algorithm for r reactions and n species in a system can be explained as follows:

i) The system at time $t=0$ with rate constants and the numbers of molecule for species are initialized.

ii) The hazard function for each reaction, h_i $(i=1,...,r)$ is calculated.

iii) The time interval for each reaction, t_i ($i=1,\ldots,r$) is computed by $t_i=Exp(h_i(x,c_i))$, where c_i is the reaction rate for the ith reaction.

iv) The smallest t_i is chosen as a time interval for the next reaction.

v) The number molecules x and the time t are updated via $t:=t+t_j$ and $x:=x+S^{(j)}$ in which $S^{(j)}$ denotes the jth row of the stoichiometry matrix.

vi) If $t<T$, step (ii) is returned and the algorithm is repeated until $t\geq T$ for the total time T.

The methods of the direct and the first reaction are similar to each other. But the first reaction method does not use *T-1* reaction times. Hence, it is less efficient than the direct method.

2.5.3 Gibson-Bruck Algorithm

The Gibson-Bruck algorithm, also known as *the next reaction method*, is more efficient than Gillespie and faster than the first reaction method in large systems since the time interval and the hazard function are calculated together in this method (Gibson and Bruck, 2000; Turner, 2004). This method modifies *T-1* reaction times to reuse for the system.

The calculation of this algorithm for a system with r reactions and n species has the following steps.

i) The rate constants and the number of molecules are initialized for $t=0$. Then the initial reaction hazards and the time interval depending on the hazard function are calculated.

ii) The smallest time interval t_j ($j=1,\ldots,r$) is chosen.

iii) The state vector x, the hazard function and the time based on the jth reaction are updated.

iv) For each reaction $j \neq i$, the hazard and the time are computed by $h_i' = h_i(x,c_i)$ and $t_i=t+(h_i/h_i')(t_i-t)$, respectively. Then, the old h_i is deleted. Finally, the

system is updated via $t:=t+t_i$ and $x:=x+S^{(j)}$ for the time and the number of reactions, in order.

v) If $t<T$, step (ii) is returned and the algorithm is repeated until $t>T$ for the total time interval T.

2.5.4 Optimal Direct Method

The optimal direct method (ODM) developed by Cao et al. (2004) is the most efficient method of SSA for the formulation. The main idea of this method is to reduce the complexity in the selection stage of reactions which need to be simulated stochastically. To optimize the underlying selection, we aim to get the total number of stochastically simulated reactions very small. Accordingly, the algorithm modifies the reactions in such a way that we can decrease the frequency of the occurrence of reactions by using pre-simulations of the system so that the presumed speed of each reaction can be categorized like slow, moderately fast and very fast reactions. Then, the reactions having slow or moderate speeds are generated later and the speediest ones simulated at first. Such a change in the simulation of certain reactions reduces the computational demand in the overall simulation of the system since the highly stochastically behaved reactions, in place of all, get the priority in the simulation.

Hence with more mathematical details, if we simulate the *j*th reaction with a reaction time β_j, we can reallocate this reaction by optimizing the average search depth *SD* of the system. The optimized *SD*, denoted by *SD**, is obtained by sorting reactions in a decreasing order based on how often they react. This order provides a smaller *SD** by optimizing the cost of simulation. Hereby to find β_j, we need one of more presimulations of the system as, typically, no prior information is provided about β_j's in the simulation of the complex biochemical networks. If

$$\beta_i' > \beta_j' \text{ for i<j} \tag{34}$$

when i and j denote two reactions in a system, this method moves the reaction happening most frequently, i.e. β_j', to the beginning of the reaction order so that the computational time can decrease.

On the other hand, if the average weighted degree *WD* is very smaller than the total number of reaction chains *R*, i.e. $WD << R$, then we compute the hazards of the reactions chains affected by the last reaction. Under such situations, we need to calculate only the *WD* amount of hazards in order to find the cost for the calculation of all *R* amounts of hazards.

2.5.5 Sorting Direct Method

The optimal direct method needs to execute several pre-simulations to determine the reaction order. However, this method cannot be effective for sharp changes because pre-simulations cannot effectively predict the system behavior. Therefore, *the sorting direct method* (SDM) is detected to eliminate the pre-simulation. This method modifies the changes. Indeed, the gene regulation network model explains this modification.

$$G1 \rightarrow G1+mRNA1 \quad (35)$$

$$mRNA1 \rightarrow mRNA1+Protein1 \quad (36)$$

$$mRNA1 \rightarrow \emptyset \quad (37)$$

$$Protein1 \rightarrow \emptyset \quad (38)$$

$$G2 \rightarrow G2+mRNA2 \quad (39)$$

$$mRNA2 \rightarrow mRNA2+Protein2 \quad (40)$$

$$mRNA2 \rightarrow \emptyset \quad (41)$$

$$Protein2 \rightarrow \emptyset \quad (42)$$

$$G2+Protein1 \rightarrow G2\ Induced \quad (43)$$

$$G2\ Induced \rightarrow G2\ Induced+mRNA2 \quad (44)$$

This network explains the transcription and translation Gene1/Gene2 into Protein1/Protein2. Gene2 is induced as a result of the combination of Gene2 and Protein1. In fact, Gene2 Induced transcribes at a high accelerated rate.

The difference between the ODM and SDM is shown in the Figure 1. The sharp changes occur for ODM while SDM remains constant because of its adaption to the changes. When Gene2 reacts more frequently than Gene1, Gene 1 can be overemphasized during the pre-simulation. The pre-simulation cannot effectively predict the behavior of the reaction for this situation. Therefore, the average search depth increases and the system performance decreases. If the firing of the reaction frequencies has constant frequency, ODM predicts the optimal order for the reactions. Hereby, in general, SDM performs as well as or better than ODM and have better strategy for large shifts.

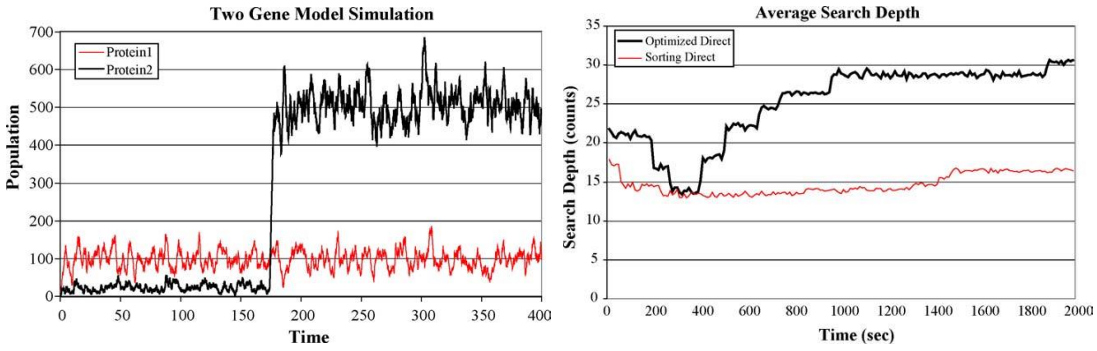

Figure 1: Average search depth for ODM and SDM for a system with 20 genes (McCollum and Peterson, 2006).

2.6 Impulses

The natural activity is generally dynamic and can have sudden changes which are called as the *impulses.* The shocks, natural disasters and vaccinations are some examples of the impulses. Hereby, these dynamic changes can happen into two different ways, namely, deterministically and stochastically. If the concentration of species in the system is low and the stochastic fluctuations have a significant effect

on the system, the stochastic impulses are used to describe the sudden changes in the model. For instance, the impulsive effect can be seen in the diabetes model. When the diabetic patient eats overdose sugary food, then this causes an instant effect, i.e. impulse, in the blood measure of the patient. The diabetic model and the explanation of its impulsive effect have been also worked on the study of Belle et al. (2009) and Jonge et al. (2014). Apart from diabetes, most of the cancer diseases can be explained via an external and big stimulus of the growth factor, resulting in impulses in the regular activation of the pathway. For instance, the MAPK-ERK system whose mis-function causes cancer, triggers by an external stimulus to its growth factor (Kolch, 2000; Kolch et al., 2005; Hornberg, 2005). Therefore, it is mainly used in oncogene researches. Besides the MAPK-ERK pathway, other EGF receptor signaling networks (Haley and Gullick, 2008) can be modelled by impulses and the full description of their reaction lists can be found from KEGG, GO or Ensembl databases. More details about the biological information in these databases can be also seen in Purutçuoğlu and Ayyıldız (2014).

The deterministic impulses are simply based on the ordinary differential equations (ODE) with impulses and the random behavior is ignored in this approach. In order to define the time of impulses, the hybrid strategies are used. The hybrid methods basically split the species and the reactions into two distinct groups, namely, continuous and discrete. The former describes fast reactions and associated species and the latter represents relatively slow reactions and related species. On the other hand, the time of impulses in the deterministic approach is called as the jump times that are determined by the discrete groups of reactions in the implementation of the hybrid methods.

Hence, as previously presented, the simple ODE system can be summarized as shown in Equation (45).

$$f(y) = \frac{d_y}{d_t} \tag{45}$$

where $y=(y_1,\ldots, y_a)$ is the state vector when the system is defined by a number of species and r number of reactions. Accordingly, the deterministic impulse model is based on this general ODE modeling with a jump function, called impulse function,

as indicated in Equations 46 and 47 for a given initialization of the system via the (t_0, y_0) point. Here, t_0 and y_0 refer to the initial time and the associated initial state, respectively.

$$y' = f(t, y) \qquad t \neq \tau_l(y) \tag{46}$$

$$\Delta y = y^{+}\big(\tau_l(y)\big) - y\big(\tau_l(y)\big) \qquad l = 1,2,\dots \tag{47}$$

In Equation 47, Δy denotes the impulse function at time $t = \tau_l$. When the impulse occurs at time τ_1, the state vector $y(t) = y(t;t_0,y_0)$ is transferred to the new position $y(t) = y(t;\tau_1, y_1^{+}(\tau_1))$ where$y_1^{+}(\tau_1))$ is new $y(t)$ point for $\tau_l - 1(y) < t \leq \tau_l(y)$.

On the other side, the stochastic impulses are based on CME, given in Equation 33 with impulses at fixed or varying times. Hereby, the time of impulses can be a continuous function, threshold function or a fixed value from a poisson distribution. For the continuous function, the impulse occurs at time t=$\tau_l(x)$, where x is the d-dimensional vector of y. For the threshold function, the time of jumps depends on a threshold defined by x_T, (t: $x(t) \geq x_T$). Finally, in the last case, the time of jumps is determined from a random value from a poisson distribution. In this study, we merely deal with the stochastic impulses.

As the application of the stochastic simulation methods with and without impulses, we consider different dimensional systems, namely, small, moderately large and large systems. As the small system, we choose the Lotka-Volterra model having 3 reactions and 2 species (i.e. prey and predator). We run the Gillespie algorithm and the first reaction algorithm. Then, we include impulses under two scenarios: Abrupt changes are observed at fixed and at random time points by adding impulses to the state vector for the time interval [t_s, t_{s+1}].

CHAPTER 3

APPLICATION

3.1 Description of Systems

In order to evaluate the performance of stochastic simulation algorithms, we select different dimensional systems, namely, small, moderately large and realistically large networks. For the toy system, we choose the Lotka-Voltera model (Wilkinson, 2006) as it is one of the well-known networks, in particular, for comparative studies. The description of this system is shown as below.

$$Y_1 \rightarrow 2Y_1 \tag{48}$$

$$Y_1 + Y_2 \rightarrow 2Y_2 \tag{49}$$

$$Y_2 \rightarrow \emptyset \tag{50}$$

The set of reactions represented in Equation 48-50, describes the relations between the predator and the prey with 2 species and 3 reactions. Here, Y_1 and Y_2 denote the *prey* and the *predator* species, respectively. With more details, Equation 48 indicates the reproduction of the prey and Equation 49 identifies that the prey behaves as a food supply for the predator. Thereby, the reproduction of the predator depends on the population of the prey. Finally, Equation 50 represents the death of predators.

On the other hand, for the moderately large system, we use the reaction list of the PKC pathway. This system stands for a model of the protein kinase C signal transduction pathway, related to the significant neural functions, especially, for the memory and the learning process. Moreover, it has importance at neuronal functions such as the synaptic long-term potentiation (LTP) and the depression (LTD) (Manninen, 2006).

PKC is activated by the rachidonic acid (AA), Ca^{2+} and the diacylglycerol (DAG). AA, cis-unsaturated fatty acid, affects the neural activities and the synapses during LTP. Hereby, the reaction of AA and Ca^{2+} triggers the sensitivity of the GAP-43 phosphorylation that is related to the synaptic connections. In fact, GAP43 is the main element of this pathway and increases during the phosphorylation of the LTP persistence. Furthermore, the PKC pathway is Ca^{2+} dependent and needs to DAG that is the product of phosphatidylinositol, for the activation. PKC translocates from the cytosol to the membrane by stimulating the variability in basal. In the reaction, AA affects the phosphorylation of GAP-43 by interacting with Ca^{2+} and DAG (Schaechter and Benowitz, 1993). Hence, the description of the system can be described by the following reaction list in Appendix B.

The other system for the assessment of large systems, we apply the JAK-STAT pathway (Maiwald et al., 2010), with realistic complexity. This network is one of the major signaling pathways controlling the immune system based on the regulation of interferons. The interferons (IFN's) are the proteins synthesized by host cells in reply to pathogens like viruses or bacteria and are used for the bridge between cells to activate defenses of the immune system against pathogens. The proteins are included by the family of glycoproteins known as cytokines and are probed into the three classes based on the type of receptors, namely, Type I IFN, Type II IFN and Type III IFN. In this study, we discuss Type I IFN which is a large subgroup of the IFN proteins.

On the other hand, IFNα is the family of Type I IFN and it stimulates immune responses against the viral infection by binding a receptor (IFNAR) of IFNα containing of IFNAR1 and IFNAR2 chains.

Hereby, Type I IFN activates the JAK/STAT pathway. In fact, the JAK activation triggers by the cellular events such as the cell migration or the cell proliferation (Rawlings, 2004) and it is suggested that the effect of this pathway is reduced by the mutation. The interaction of cytokines and receptors determines the cellular behavior. As a result of the binding ligand to the receptor, the activator of STAT1 and STAT2 which is the member of JAK1, i.e. the Janus kinases and TYK2, the tyrosine kinases 2 are activated. Specifically, the IFNAR dimer stimulates IRF9,

which is the interferon regulatory factor 9, and this causes to form ISGF3, that is known as the interferon stimulated gene factor 3. SOCS, which is the suppressor of cytokine signaling, affects negatively the cytokine signaling. Accordingly, IRF9 has a significant influence on the IFNα signal transduction, resulting in the achievement of the antiviral response. In Figure 2, we briefly summarize the biological process of the JAK-STAT pathway for a simple visual representation and the complete list of reactions for this system is presented in Appendix A.

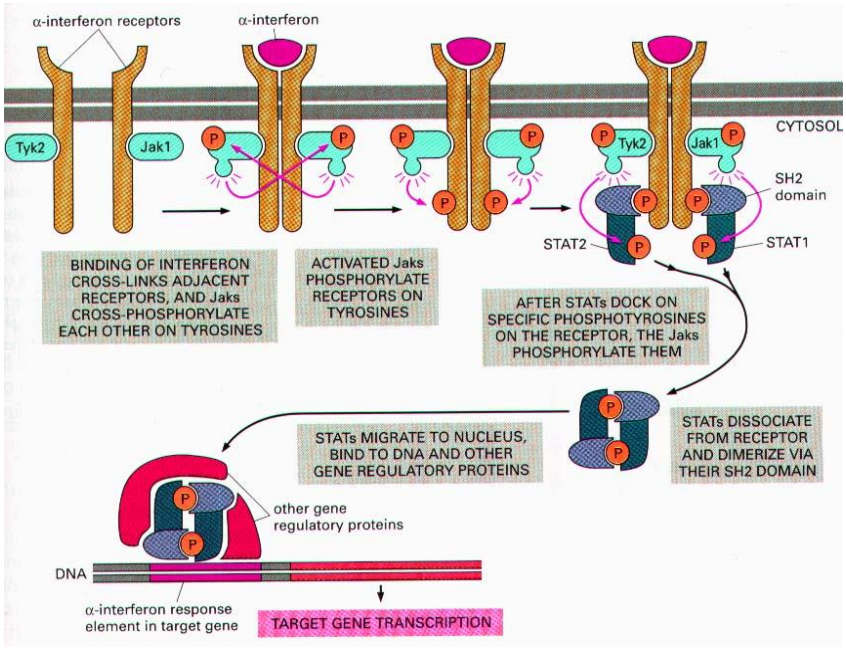

Figure 2: The simple presentation of the JAK-STAT pathway (Selice Wordexpress, 2011).

In addition, we investigate the Lysis–Lysogeny model. In fact, the viral and the temperant phages can select the pathway lysis and lysogenic circuit. The selection depends on the competition of the cro and CI protein. Here, CI and cro are required for the lysogenic and the lytic cycle, in order. The T4 phage is the protype of the viral phage. On the other hand, the λ phage is the prototype of the temperant phage. Indeed, the lysogenic cycle is the reproduction of viral issues and the lytic means that the cell integrity is distorted. The mechanism of the lytic system is initiated with the transcription and the translation. Actually, the cycle of the lytic phage starts with a

stick to the cell-wall. The virus DNA enters to the cell in order to synthesize the virus protein. Therefore, the virus DNA can multiply themselves. Then, the lysogenic cycle starts with an entrance dependent on three proteins, CI, CII, CIII. The CI protein is produced after the viral DNA interacts with an individual cell. Then, the production of CII and CIII starts. CII provides the production of CI and CIII prevents the degradation of CII by enzymes. After the completion of the integration, only CI produces. Finally, the phage is integrated with the bacterial genome and this stage is called as the pro-phage and the associated bacteria are named as the lysogen bacteria. Generally, these systems have low concentration of species and slow reaction rates. The complete list of reactions for this system is presented in Appendix B.

Finally, we investigate the convergence of the JAK-STAT and PKC pathway to observe the response of the methods for large systems. Here, the reaction of the STAT1 and GATA4 proteins trigger the convergence of the JAK-STAT and PKC pathways. In this extended pathway, the vasoactive hormone and AII change the gene transcription. AII reacts with AT1R and AT1R modifies the biological effects of AII. AT1R is generally used for the treatment of the cardiovascular disease. In fact, some STAT molecules are activated by AII. Hereby, the link of two pathways, which are JAK-STAT and PKC, are based on the AT1R and AII activation. The description of the additional reactions which bind these two systems are presented in Appendix part.

3.2 Comparison of Methods

3.2.1 Comparison for long and short time

The LV, PKC, JAK-STAT and Lysis – Lysogeny systems are investigated for the time t= 5, 20 and 50 for both the Gillespie and the First Reaction Methods. In simulations, the initial number of molecules is taken as 100 arbitrarily for each system. On the other hand, the reaction rate constants are equated as used in most of the comparative studies about these systems whose details are given for the LV system in Wilkinson, 2006. As seen in Figures 3, 4 and 5, the increase in the number of preys affects the number of predators due to the fact that the predators need foods which are preys in our system as described in the reaction list as well. Accordingly,

when the number of preys decreases, their main nutritional sources are shortened, resulting in the extinction of these species. Indeed, a decrease in the time interval enables us to better visualize the fluctuation in the system. On the other hand, while the number of species decreases, the total hazard falls down as represented in Figure 3, 4, 5. In the end, when we compare the computational demands of the direct and the first reaction algorithms for this system with respect to the selected criteria S, D and CPU, the direct method indicates a less computation cost with respect to the first reaction method under no impulsive scenarios.

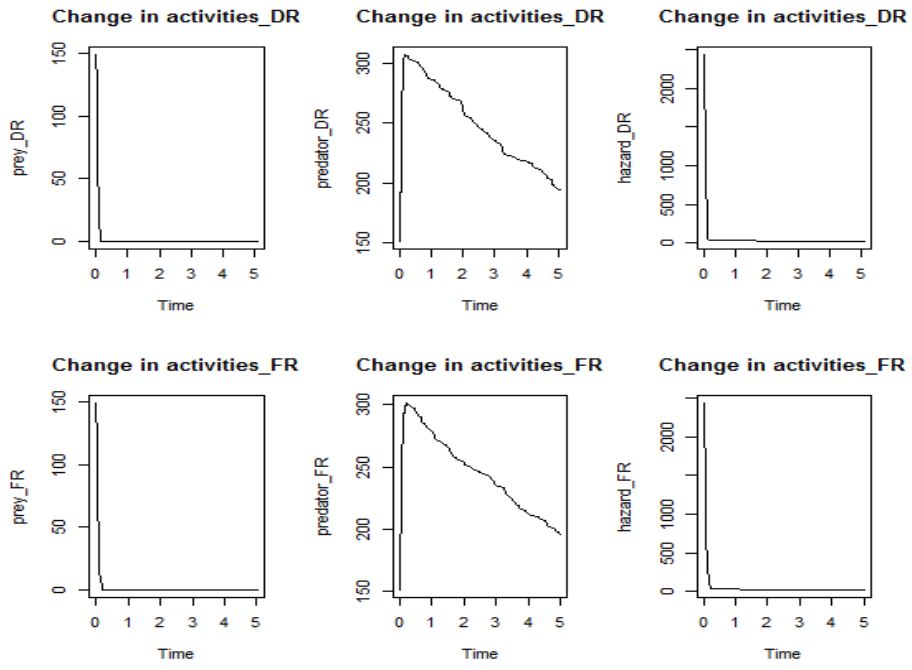

Figure 3: The plots indicate the changes in the number of preys, predators and the changes in hazards of the LV system via *Direct Method (DR)* and *First Reaction Method (FR)* for **the time 5** when the system has no impulse.

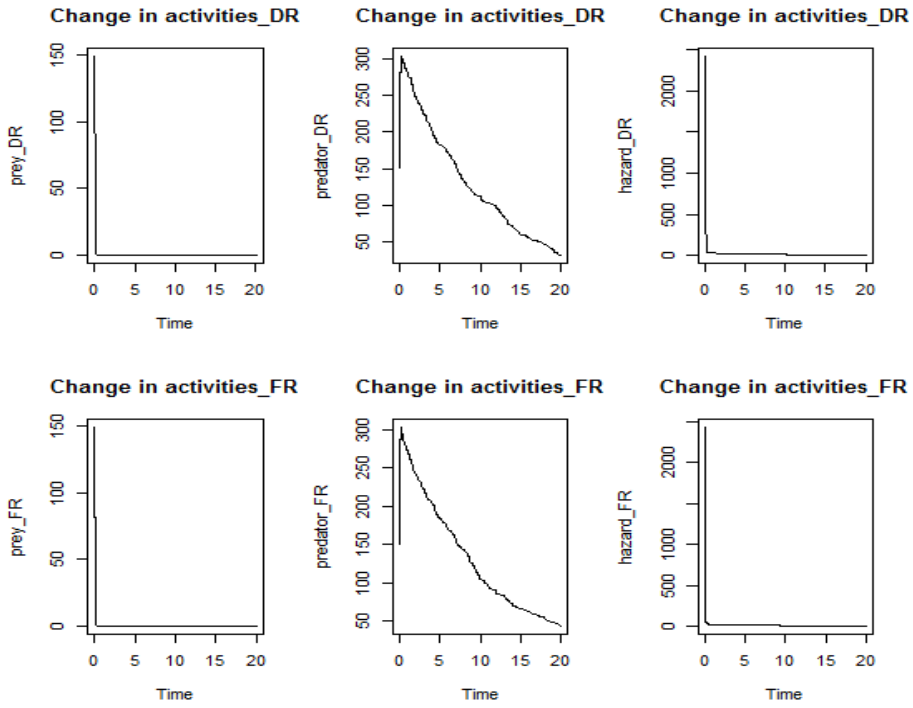

Figure 4: The plots indicate the changes in the number of preys, predators and the changes in *hazards of the LV system via Direct Method (DR) and First* Reaction Method (FR) for **the time 20** when the system has no impulse.

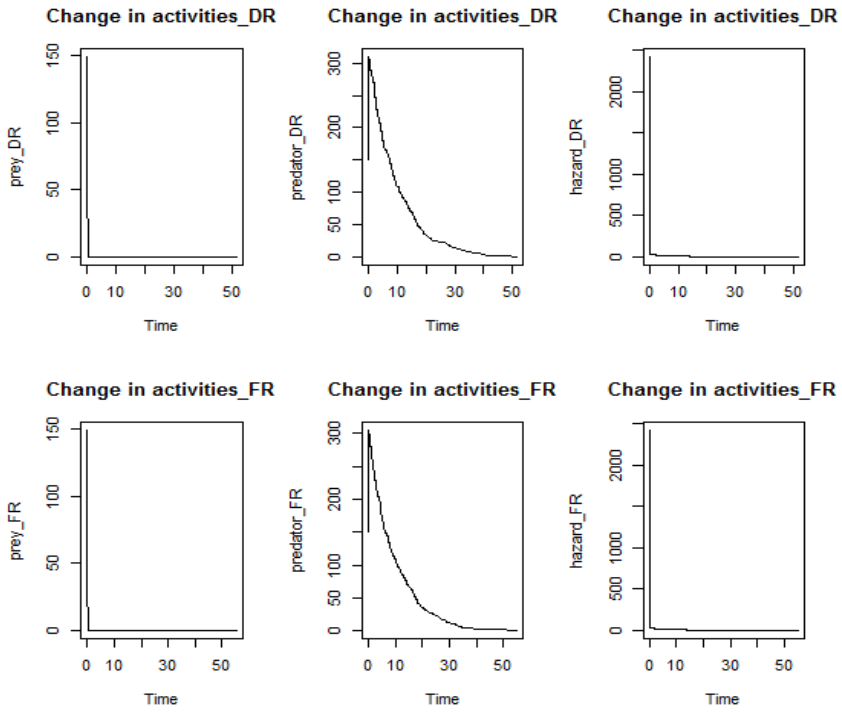

Figure 5: The plots indicate the changes in the number of preys, predators and the changes in hazards of the LV system via *Direct Method (DR)* and *First Reaction Method (FR)* for **the time 50** when the system has no impulse.

It can be seen in the Figure 3, 4, 5 there is no significant difference between Gillespie and First Reaction Method for short or long runs. Both systems have almost the same results.

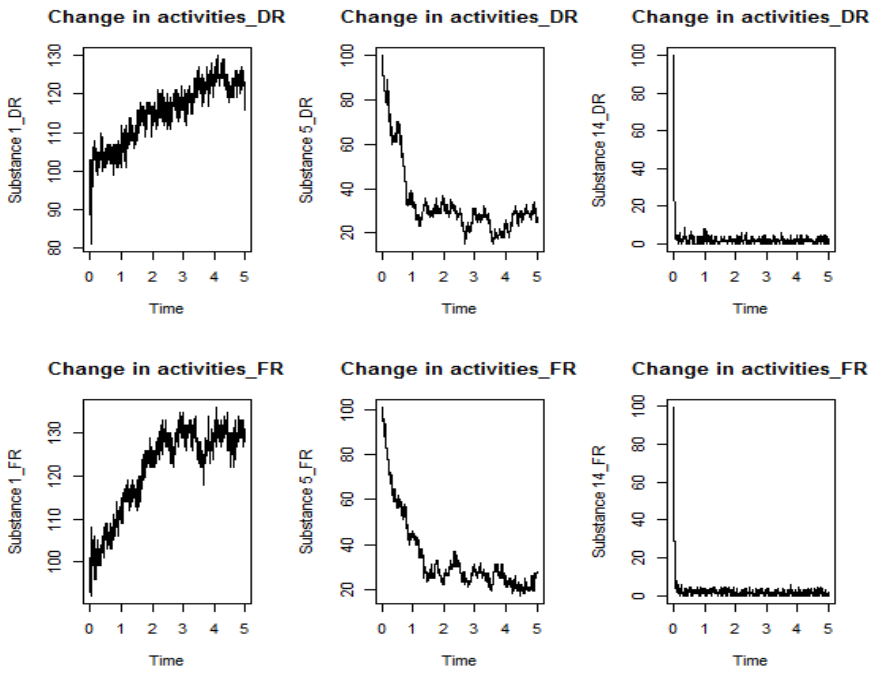

Figure 6: The plots indicate the changes in the number of the Substance 1, Substance 5 and Substance 14 of the PKC system via *Direct Method (DR)* and *First Reaction Method (FR)* for **the time 5** when the system has no impulse.

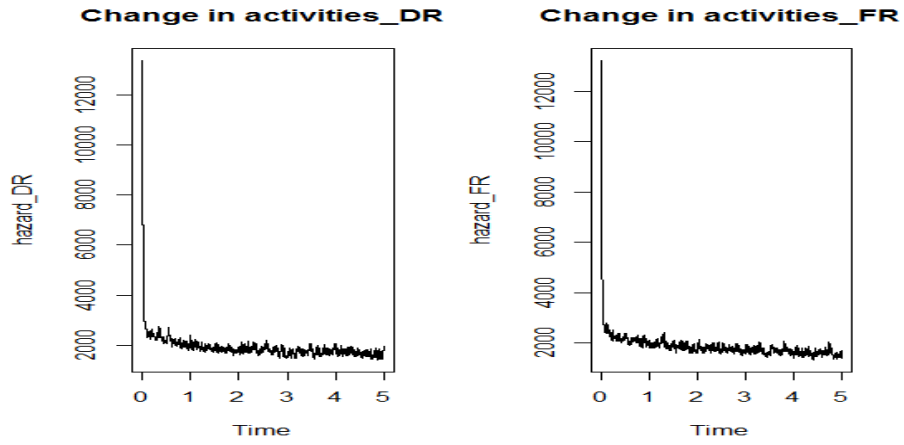

Figure 7: The plots indicate the changes in hazards of the PKC system via *Direct Method (DR)* and *First Reaction Method (FR)* for **the time 5** when the system has no impulse.

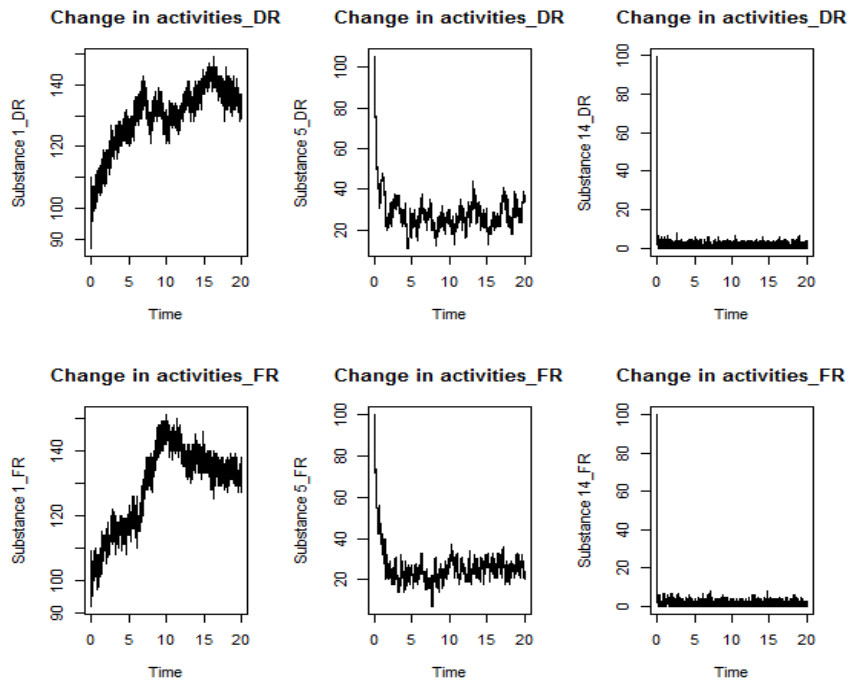

Figure 8: The plots indicate the changes in the number of the Substance 1, Substance 5 and Substance 14 of the PKC system via *Direct Method (DR)* and *First Reaction Method (FR)* for **the time 20** when the system has no impulse.

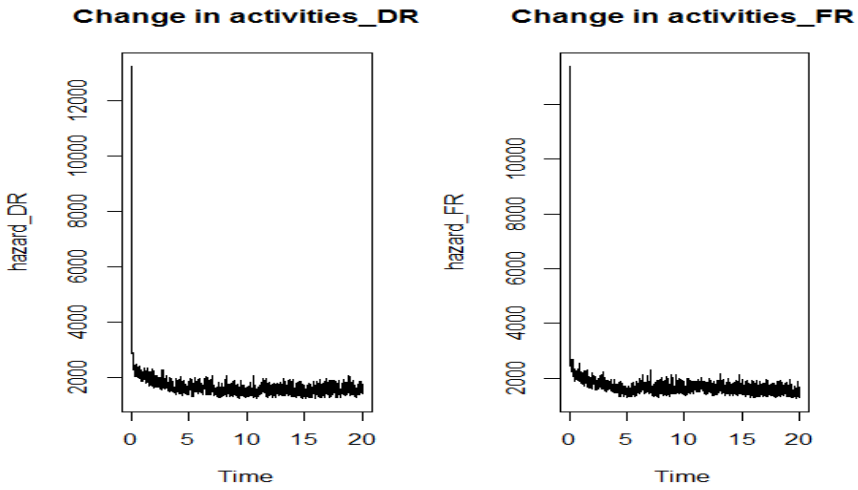

Figure 9: The plots indicate the changes in hazards of the PKC system via *Direct Method (DR)* and *First Reaction Method (FR)* for **the time 20** when the system has no impulse.

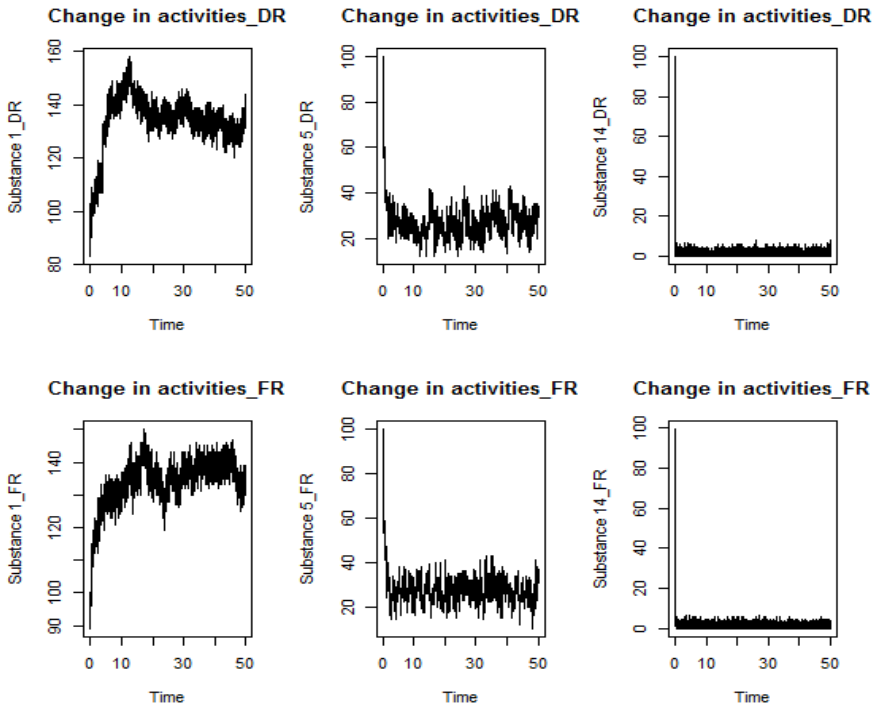

Figure 10: The plots indicate the changes in the number of the Substance 1, Substance 5 and Substance 14 of the PKC system via *Direct Method (DR)* and *First Reaction Method (FR)* for **the time 50** when the system has no impulse.

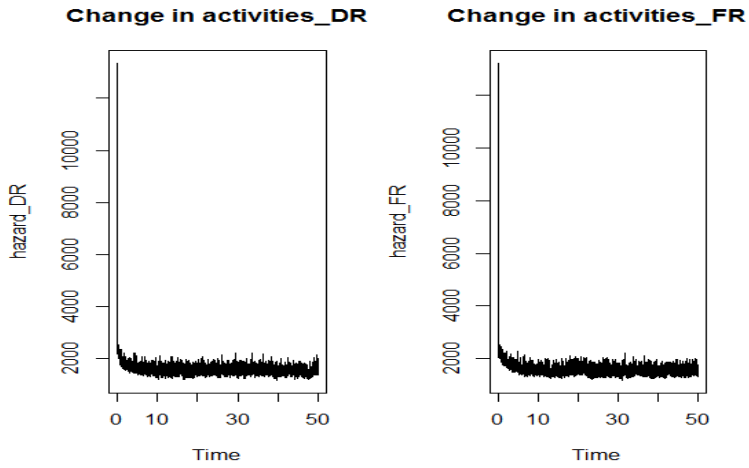

Figure 11: The plots indicate the changes in hazards of the PKC system via *Direct Method (DR)* and *First Reaction Method (FR)* for **the time 50** when the system has no impulse.

In Figures 6-11, we present the changes in the number of substances and the changes in hazard for the PKC system during short and long time. In these analyses, the number of molecules is initialized by 100, as previously implemented, and the stochastic reaction rate constants are set to the entries in the study of Schaechter and Benowitz (1993). From the results it is observed that the outputs of both the Gillespie and the first reaction methods are almost the same apart from the fluctuations due to the stochasticity of the systems.

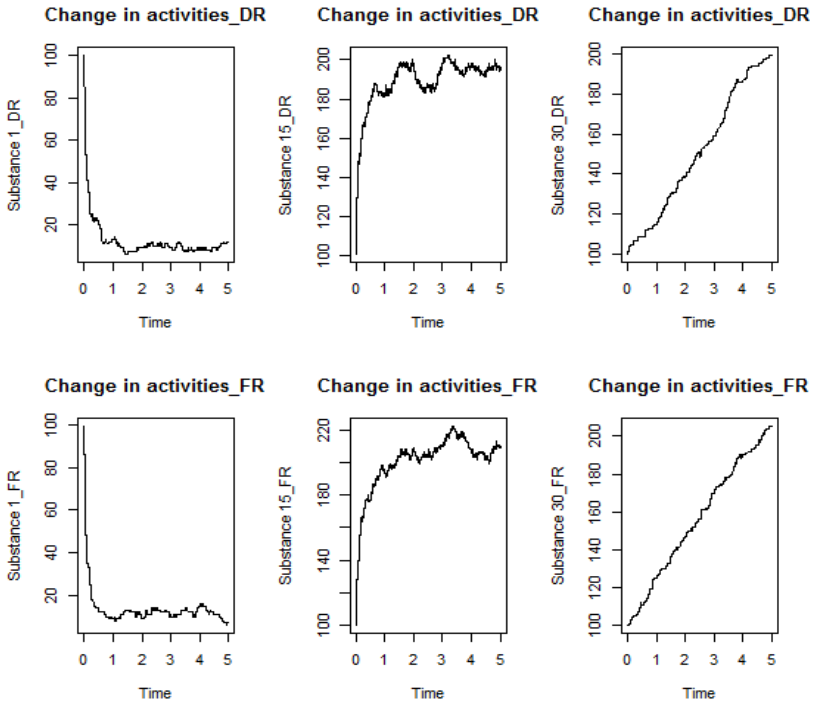

Figure 12: The plots indicate the changes in the number of the substance 1, 15 and 30 of the JAK-STAT system via *Direct Method (DR)* and *First Reaction Method (FR)* for **the time 5** when the system has no impulses.

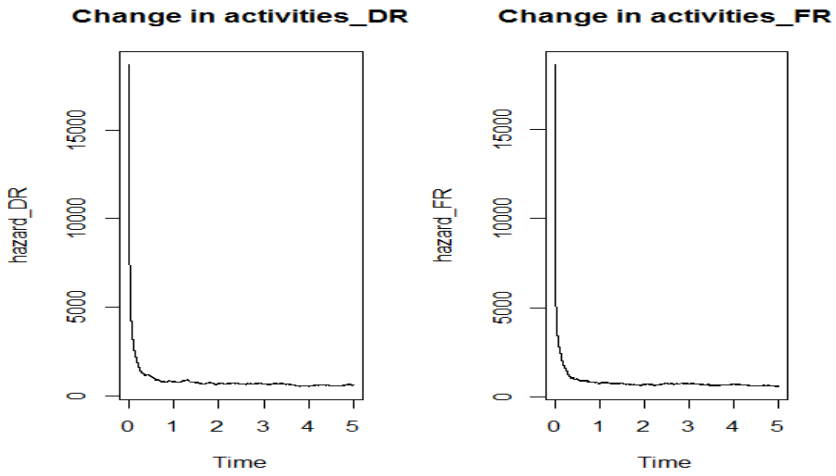

Figure 13: The plots indicate the changes in hazards of the JAK-STAT system via *Direct Method (DR)* and *First Reaction Method (FR)* for **the time 5** when the system has no impulse.

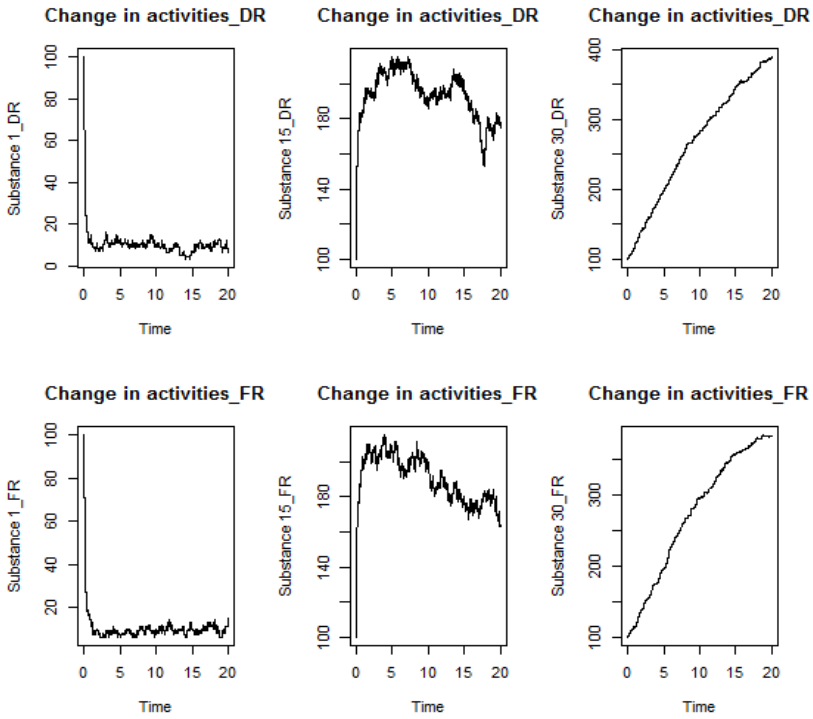

Figure 14: The plots indicate the changes in the number of the Substance 1, 15 and 30 of the JAK-STAT system via *Direct Method (DR)* and *First Reaction Method (FR)* for **the time 20** when the system has no impulse.

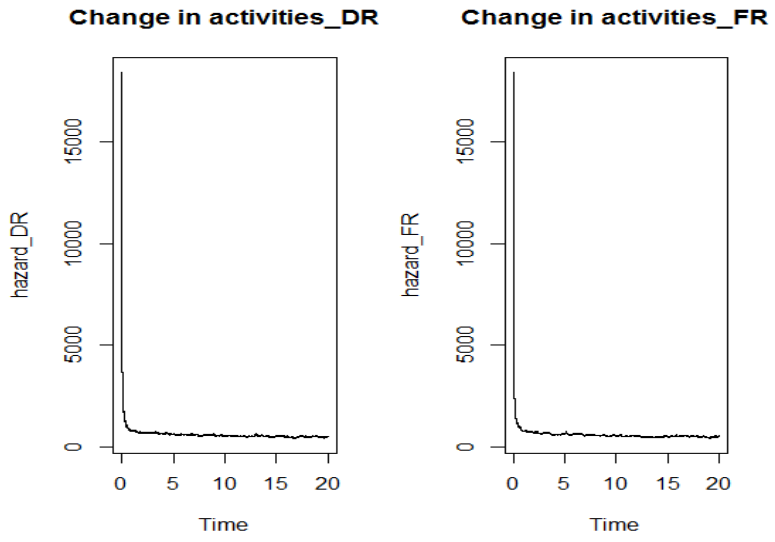

Figure 15: The plots indicate the changes in hazards of the JAK-STAT system via *Direct Method (DR)* and *First Reaction Method (FR)* for **the time 20** when the system has no impulse.

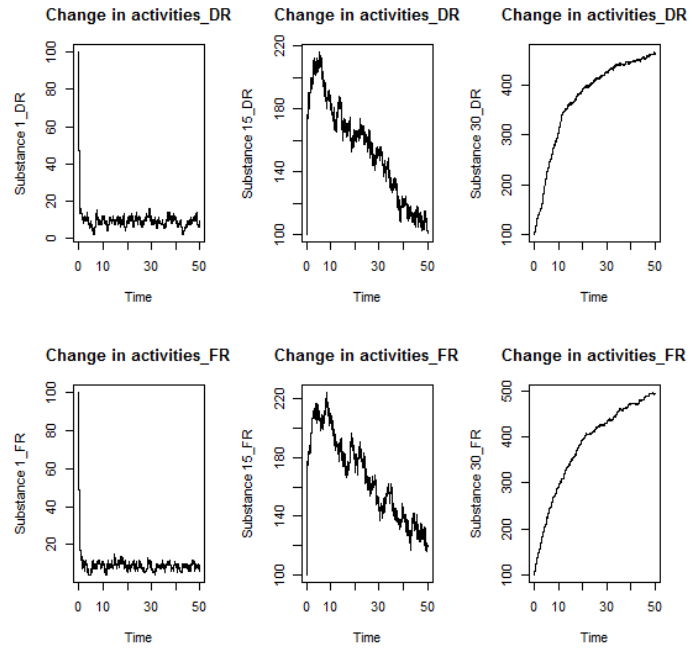

Figure 16: The plots indicate the changes in the number of the Substance 1, 15 and 30 of the JAK-STAT system via *Direct Method (DR)* and *First Reaction Method (FR)* for **the time 50** when the system has no impulse.

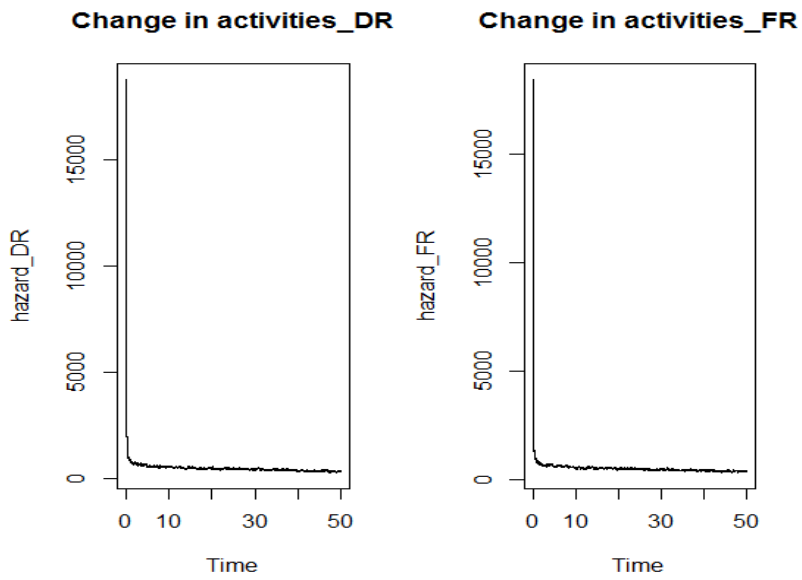

Figure 17: The plots indicate the changes in the number of the hazard of JAK-STAT system via *Direct Method (DR)* and *First Reaction Method (FR)* for **time 50** when the system has no impulse.

In the analyses of the JAK-STAT pathway, as used beforehand, we set the time to 5, 20 and 50 in order to see the effects of the short, moderate and long time period. Moreover, as performed in previous analyses, we equate the initial number of molecules of each species to 100 if there is no any biological explanation for this system in the literature. On the other hand, the description of the system and associated reaction rates are taken from the study of Rawlings (2004) which is one of the main biological sources for this pathway. As a result, as seen in Figures 12-17, we observe that both algorithms produce very similar findings in most of species in the system. Whereas, for certain substrates such as Substrate 15, the changes in the number of molecules are more obviously seen. On the other side, if we compare the changes in hazards of both algorithms, we find that they have almost similar time jumps during whole simulation and it means that the first reaction algorithm cannot significantly improve the computational burden of the Gillespie algorithm in realistic systems.

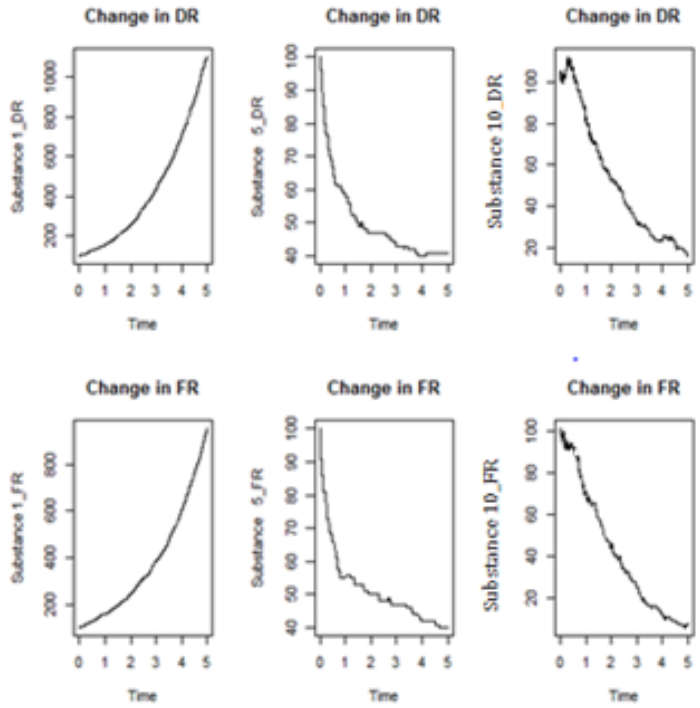

Figure 18: The plots indicate the changes in the number of the Substance 1, 5 and 10 of the Lysis – Lysogeny system via *Direct Method (DR)* and *First Reaction Method (FR)* for **the time 5** when the system has no impulse.

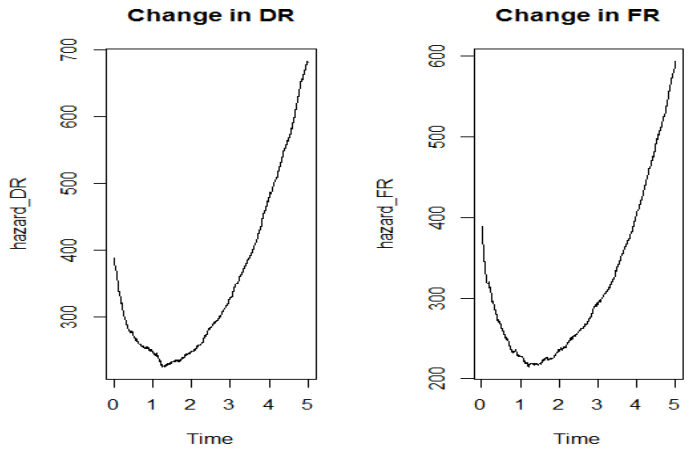

Figure 19: The plots indicate the changes in the number of the hazard of Lysis – Lysogeny system via *Direct Method (DR)* and *First Reaction Method (FR)* for **time 5** when the system has no impulse.

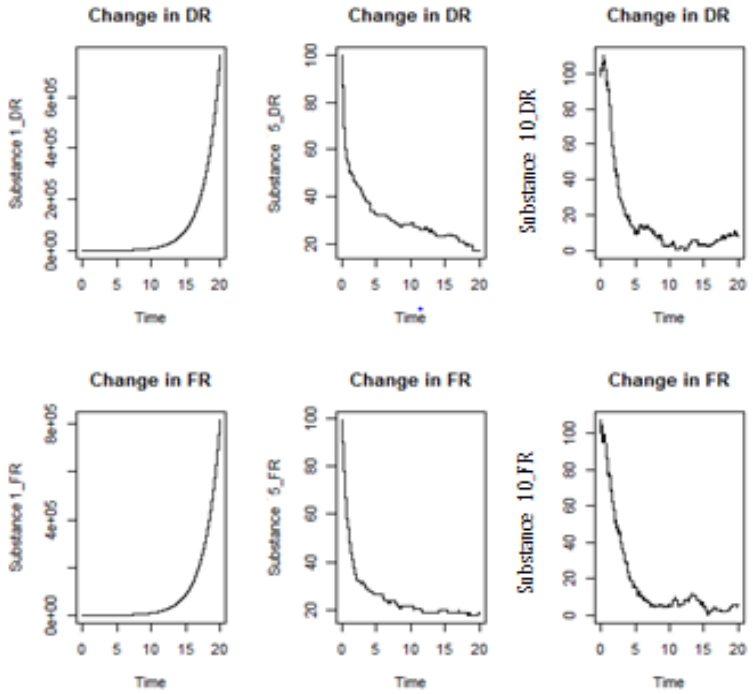

Figure 20: The plots indicate the changes in the number of the Substance 1, 5 and 10 of the Lysis – Lysogeny system via *Direct Method (DR)* and *First Reaction Method (FR)* for the **time 20** when the system has no impulse.

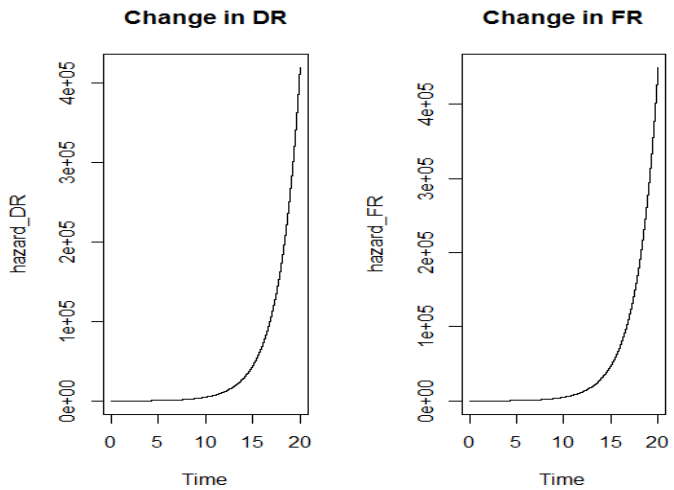

Figure 21: The plots indicate the changes in the number of the hazard of Lysis – Lysogeny system via *Direct Method (DR)* and *First Reaction Method (FR)* for **time 20** when the system has no impulse.

In the analyses of the Lysis-Lysogency system, initial values of all species are taken as 100 and we set the time to 5, 20 and 50, respectively. On the other hand, the description of the system and associated reaction rates are obtained from the study of Arkin (1998) which is one of the main biological sources for this pathway. Hence, as seen in Figures 18-21, we detect almost the same results from the Gillespie and the first reaction algorithms in the simulation of the system. But in both algorithms, we observe a significantly long simulation-time under all conditions. We consider that the reason is dependent on the biological description of the system and the initial values, rather than the performance of the algorithms for this pathway.

Table 3: The results for the Lysis-Lysogenic model for *Direct Method* and *First Reaction Method*. For time 50, we obtain the results by using the ratio of time for 5 and 20 since the time step under time 50 increases by 0.0002 units which leads to computational burden.

LL		**Without Impulse**			
		CPU	**Local Time**	**SD**	**WD**
Direct Method	**t=5**	16.2	352.26	5.49	2.78
	t=20	6677.65	29867.95	5.68	2.99
	t=50	8806532.3	2935560.26	457.51	231.66
First Reaction Method	**t=5**	22.88	388.73	5.64	2.78
	t=20	11153.05	45008.64	6.70	2.99
	t=50	9899785.86	6675700.8	546.5	242.8

According to Table 3, the time effect on CPU is higher compared to other systems like LV or PKC system. The Gillespie and the first reaction methods have little difference. On the other hand, while the time increases, the difference of CPU time between two methods gets larger.

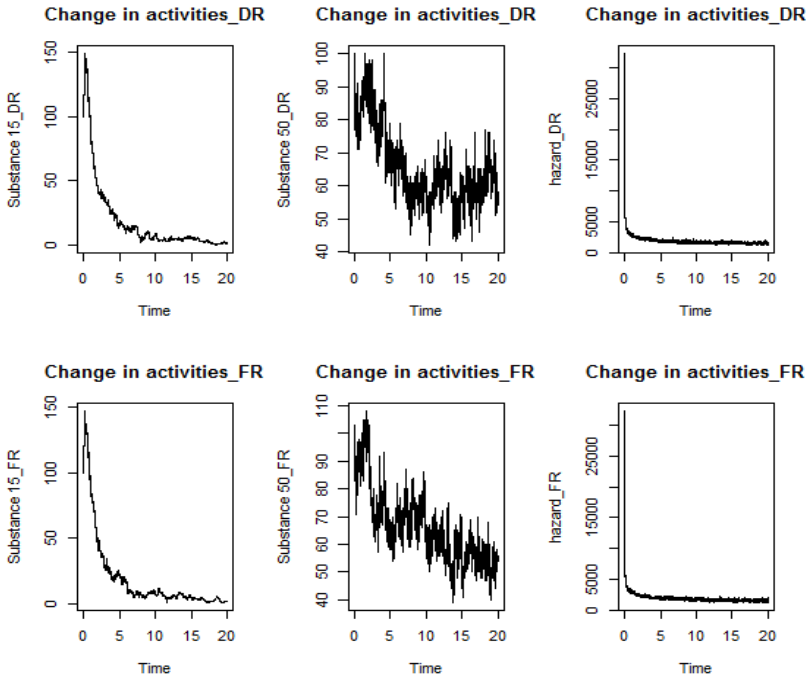

Figure 22: The plots indicate the changes in the number of molecules 15 and 50 and the hazard of the convergence of PKC and JAK-STAT system via *Direct Method (DR)* and *First Reaction Method (FR)* for time 20 when the system has no impulse.

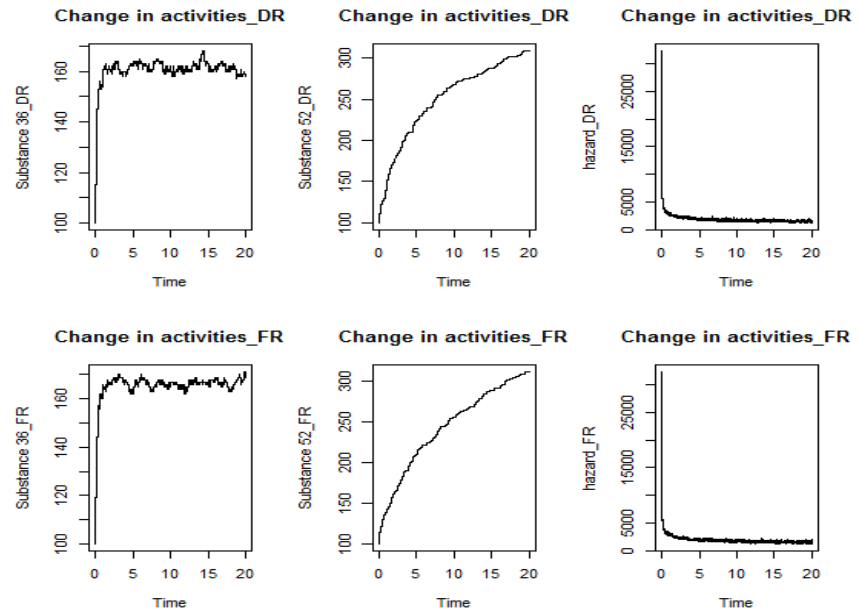

Figure 23: The plots indicate the changes in the number of the molecules 36 and 52 and the hazard of the convergence of PKC and JAK-STAT system via *Direct Method (DR)* and *First Reaction Method (FR)* for **time 20** when the system has no impulse.

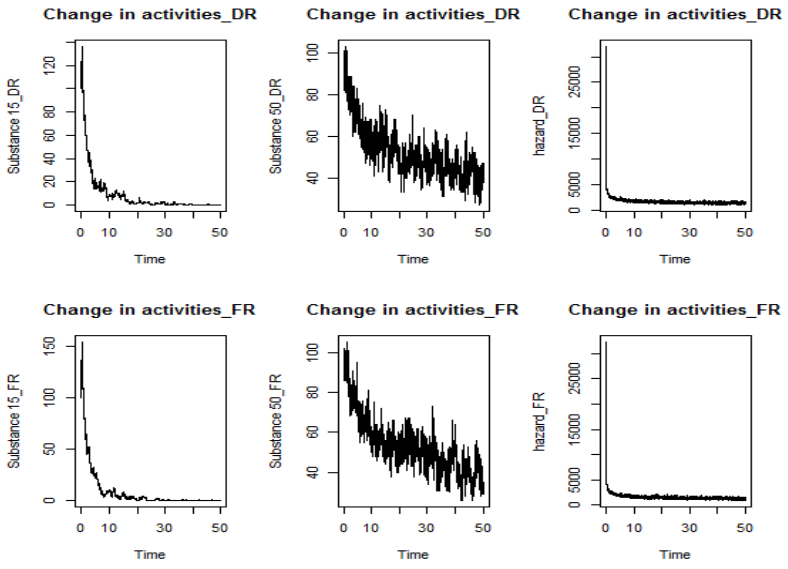

Figure 24: The plots indicate the changes in the number of molecules 15 and 50 and the hazard of the convergence of PKC and JAK-STAT system via *Direct Method (DR)* and *First Reaction Method (FR)* for **time 50** when the system has no impulse.

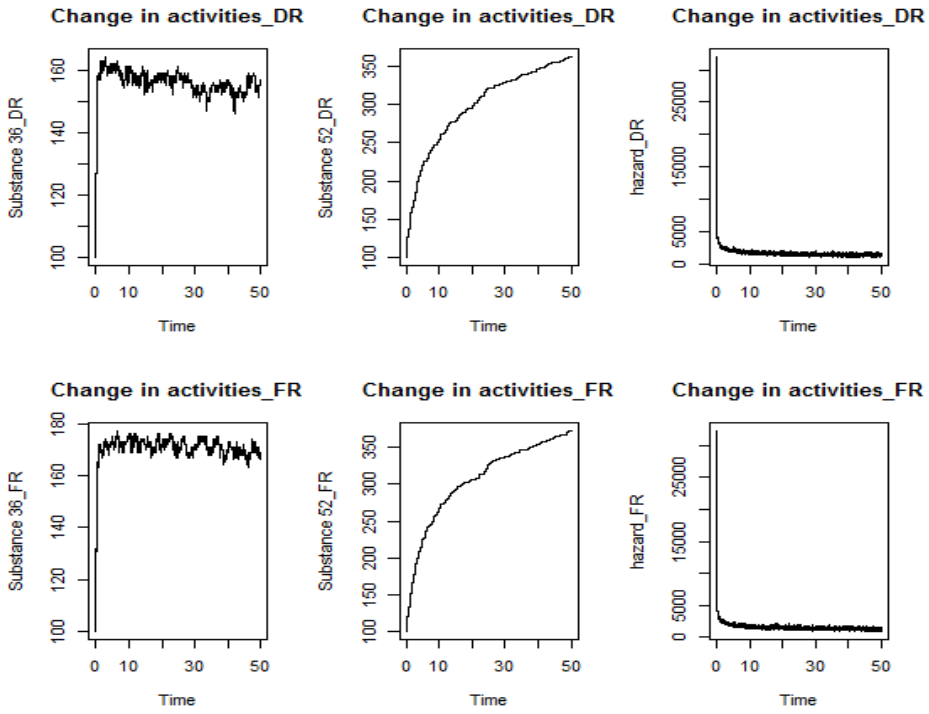

Figure 25: The plots indicate the changes in the number of molecules 36 and 52 and the hazard of the convergence of PKC and JAK-STAT system via *Direct Method (DR)* and *First Reaction Method (FR)* for **time 50** when the system has no impulse.

Finally, we investigate the convergence of the JAK-STAT and the PKC systems to observe the behavior of the largest system for the DR and FR methods. Similar to the previous findings, we obtain same results in the sense that there is no significant difference between two methods as seen in Table 4. On the other hand, we can visually detect more obvious stochasticity under the Gillespie methods.

Table 4: The results for the convergence of JAK-STAT and PKC model for *Direct Method* and *First Reaction Method*.

The Convergence of JAK-STAT and PKC		**Without Impulse**			
		CPU	Local Time	SD	WD
Direct Method	**t=20**	228.3	1300.09	61.14	6.06
	t=50	1703.23	8015.76	62.62	5.87
First Reaction Method	**t=20**	486.58	2290.65	61.50	6.07
	t=50	1081.41	5585.89	62.84	5.83

3.2.2 Comparison for long and short time with impulses

In this part, we investigate the impulsive effects on LV, PKC, JAK-STAT, Lysis–Lysogeny and the convergence of the JAK-STAT and the PKC systems by using the Gillespie and the first reaction methods for short and long runs. To see the underlying effects, we consider two different scenarios for each system. In the initialization of all systems, we set the number of molecules of every species to 100 if there is no any other biological explanation about the systems. Furthermore, we equate the reaction rate constants for the Lotka Voltera (LV) system as given in Wilkinson (2006), for the PKC pathway as presented in Schaechter and Benowitz (1993), for the complex JAK-STAT pathway as listed in Rawlings (1994), for the Lysis-Lysogeny system as described in Arkin et al (1998), and finally for the convergence of the JAK-STAT and the PKC systems as presented in Wang (2005).

For the LV system, the first scenario is to detect the effect of the changes in the number of molecules at a decreasing direction. Hereby, if the number of the prey becomes greater than 60 in simulations, we decrease it 20 units. On the other side, in Scenario 2, we consider to decrease the number of preditors at 20 units if it exceeds 60 molecules in the simulations.

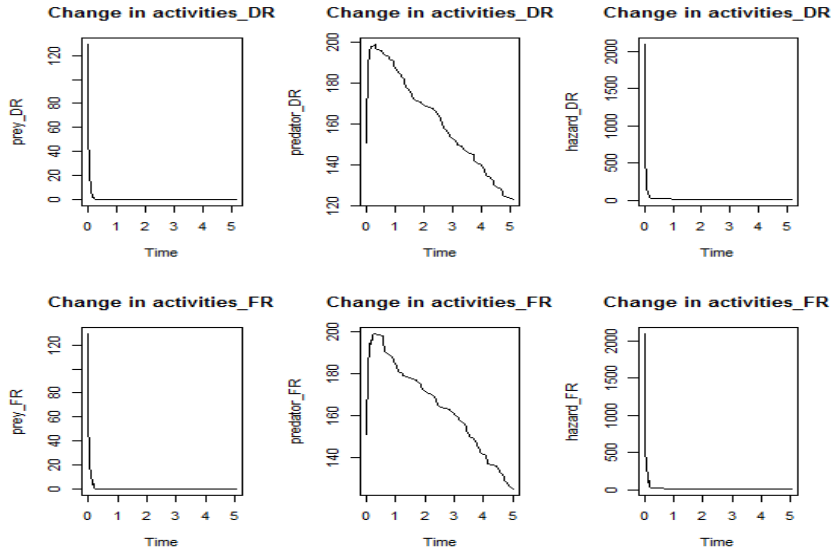

Figure 26: (Scenario 1) The plots indicate the changes in the number of preys, predators and the changes in hazards of the LV Model for *Direct Method (DR)* and *First Reaction Method (FR)* for the **fixed time 5** when the system has an impulsive effect.

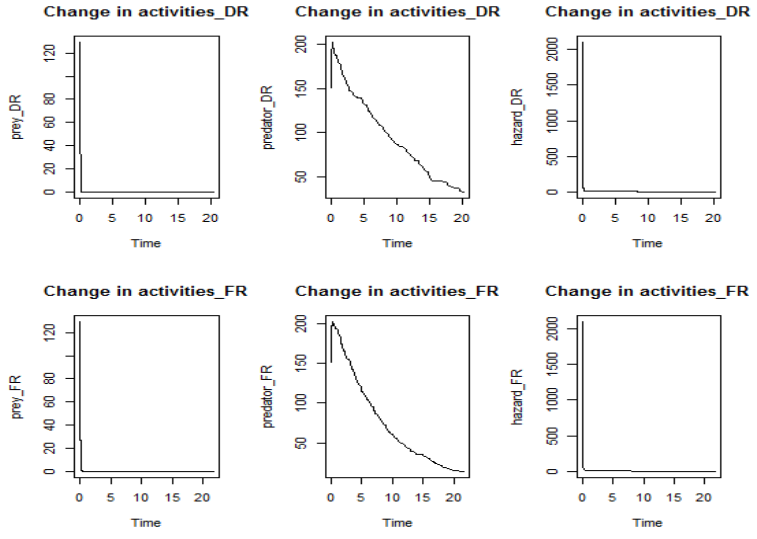

Figure 27: (Scenario 1) The plots indicate the changes in the number of preys, predators and the changes in hazards of the LV Model for *Direct Method (DR)* and *First Reaction Method (FR)* for the **fixed time 20** when the system has an impulsive effect.

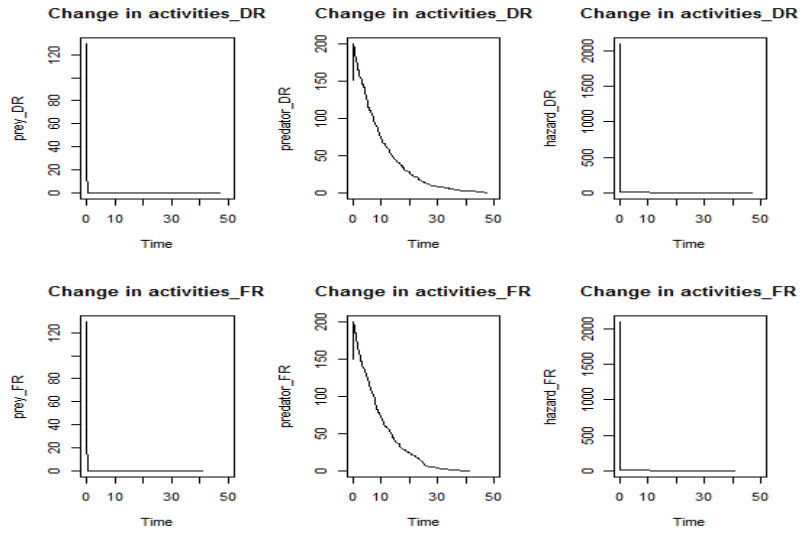

Figure 28: (Scenario 1) The plots indicate the changes in the number of preys, predators and the changes in hazards of the LV Model for *Direct Method (DR)* and *First Reaction Method (FR)* for the **fixed time 50** when the system has an impulsive effect.

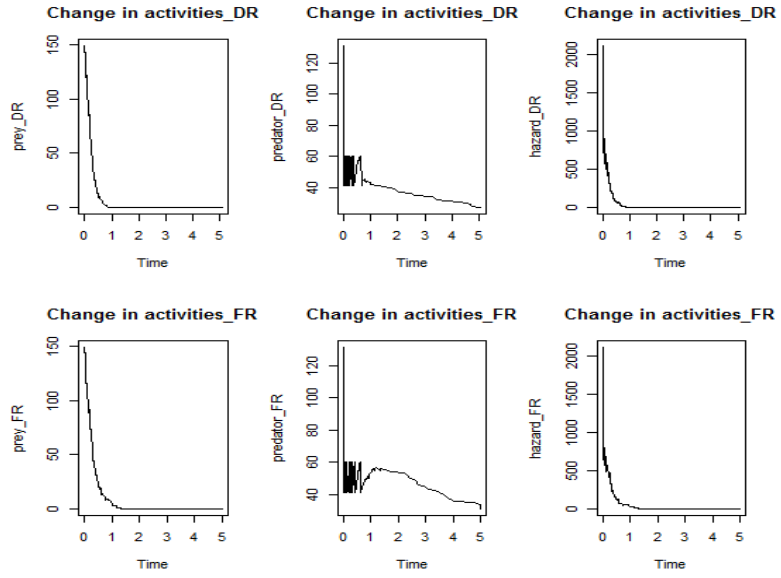

Figure 29: (Scenario 2) The plots indicate the changes in the number of preys, predators and the changes in hazards of the LV Model for *Direct Method (DR)* and *First Reaction Method (FR)* for the **fixed time 5** when the system has an impulsive effect.

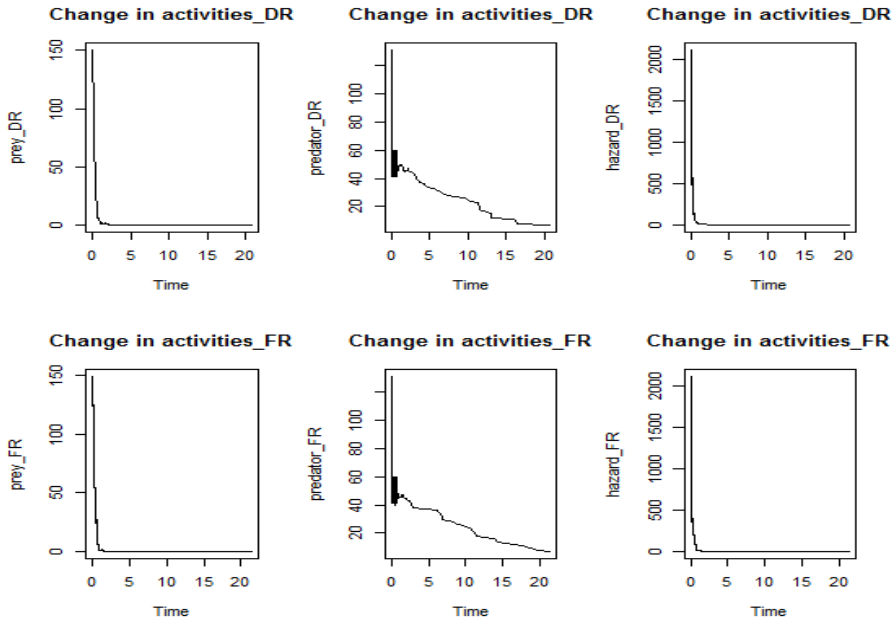

Figure 30: (Scenario 2) The plots indicate the changes in the number of the prey, the predator and the changes in for *Direct Method (DR)* and *First Reaction Method (FR)* for the **fixed time 20** when the system has an impulsive effect.

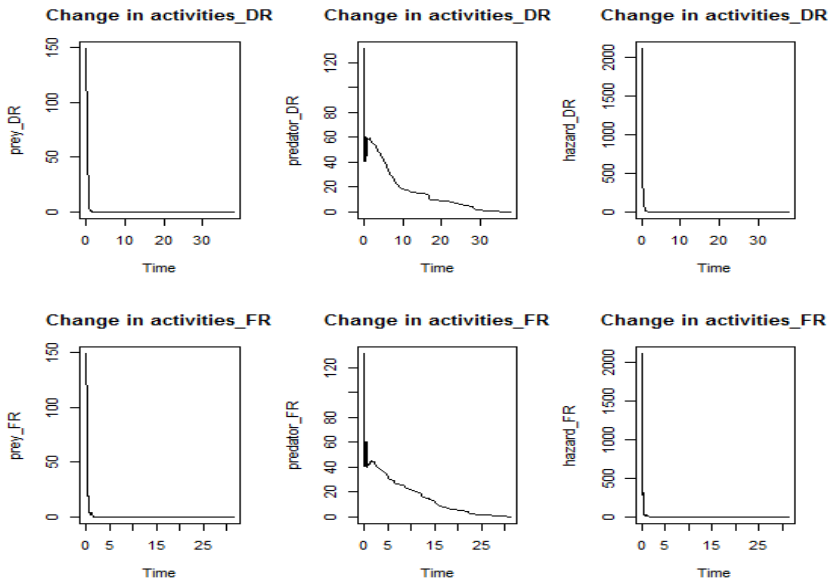

Figure 31: (Scenario 2) The plots indicate the changes in the number of the prey, the predator and the changes in for *Direct Method (DR)* and *First Reaction Method (FR)* for the **fixed time 50** when the system has an impulsive effect.

Figures 26-31 show the impulsive effects of the LV system under different simulation periods. In these plots we observe that the system dies when there is the extinction in one species. Therefore, we cannot run the simulations until the time 50 for both scenarios and algorithms. Furthermore, we observe that the changes in predators are more visible than the changes in preys under impulses and the die of the predators can be more crucial with respect to the die of the preys for the system. We explain this situation as follows: As the number of predators is always higher than the number of preys in the equilibrium states, a sharp decrease in the predator can cause a sharp change in the system and can lead to the underlying extinction.

On the other hand, from Table 5, we detect that there is no significant difference in the computational demand of the direct, i.e. Gillespie, and first reaction methods. In general, the impulsive effects cause higher CPU and local time according to Scenario 1 and 2.

In the PKC system, we construct our scenarios based on three substances, namely, Substances 1, 5 and 14, which are chosen arbitrarily. In this part, the effect of impulses for Scenario 1 depends on the sudden increase of the Substance 1. This scheme considers that if the number of molecules for Substance 1 becomes greater than 60, it decreases 20 units. On the other hand, under the Scenario 2, we think the same plan for the Substance 5.

Table 5: The results for LV for *Direct Method* and *First Reaction Method.*

LV		Without Impulse				With Impulse (scenario 1)				With Impulse (scenario 2)			
		CPU	Local Time	SD	WD	CPU	Local Time	SD	WD	CPU	Local Time	SD	WD
Direct Method	t=5	20.98	3648.11	2.37	2.55	1474.64	12959.4	2.59	2.39	1444.55	12089.78	1.97	2.75
	t=20	24.69	3916.6	2.59	2.35	1482.68	13274.68	2.72	2.23	1440.7	11789.94	2.04	2.66
	t=50	28.89	4127.22	2.62	2.33	1485.81	13511.13	2.79	2.20	1447.36	12280.33	2.06	2.64
First Reaction Method	t=5	22.24	3651.37	2.38	2.57	1475.25	12961.56	2.55	2.39	1445.47	12095.05	1.95	2.72
	t=20	26.42	3922.02	2.58	2.36	1483.75	13278.22	2.73	2.22	1441.96	11798.44	2.04	2.69
	t=50	30.52	4132.7	2.63	2.33	1486.68	13517.28	2.76	2.20	1485.81	13511.13	2.07	2.68

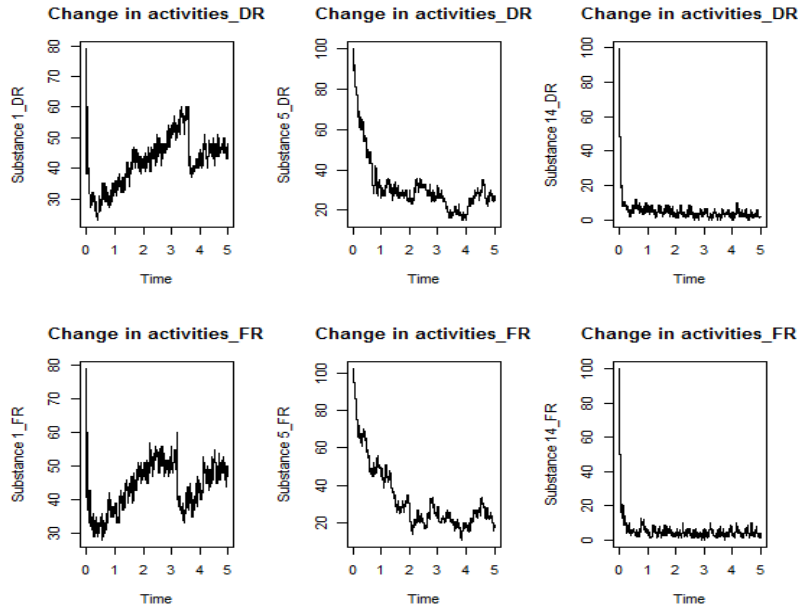

Figure 32: (Scenario 1) The plots indicate the changes in the number of Substance 1, Substance 5 and hazard of PKC Model for *Direct Method* and *First Reaction Method* for the fixed time 5, respectively, when the system has impulsive effect.

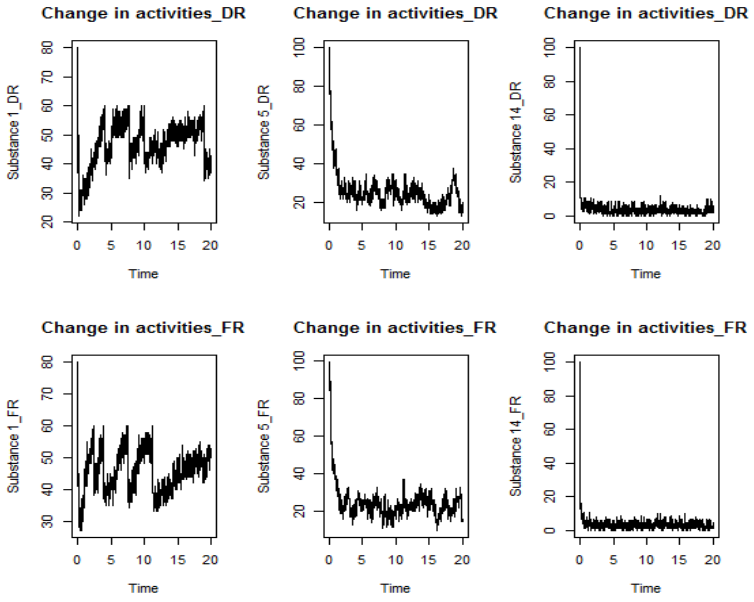

Figure 33: (Scenario 1) The plots indicate the changes in the number of Substance 1, Substance 5 and hazard of PKC Model for *Direct Method* and *First Reaction Method* for the **fixed time 20**, respectively, when the system has impulsive effect.

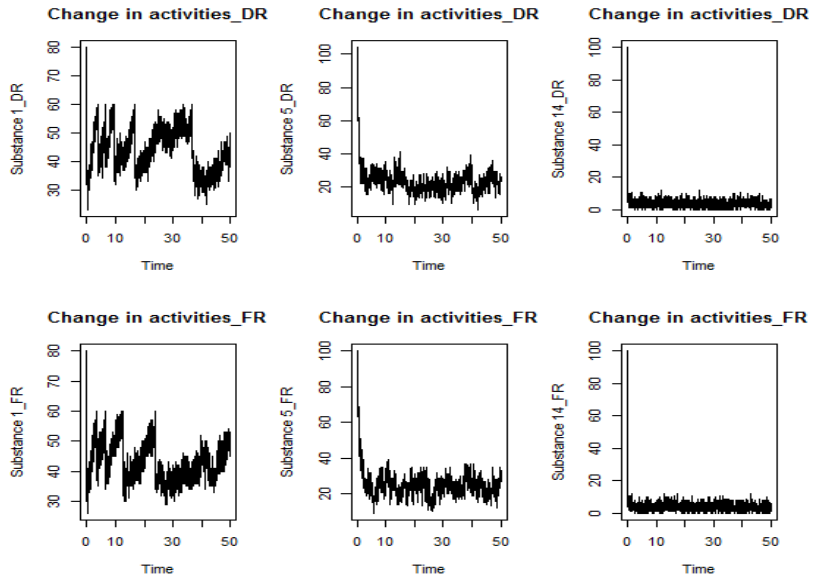

Figure 34: (Scenario 1) The plots indicate the changes in the number of Substance 1, Substance 5 and hazard of PKC Model for *Direct Method* and *First Reaction Method* for **the fixed time 50**, respectively, when the system has impulsive effect.

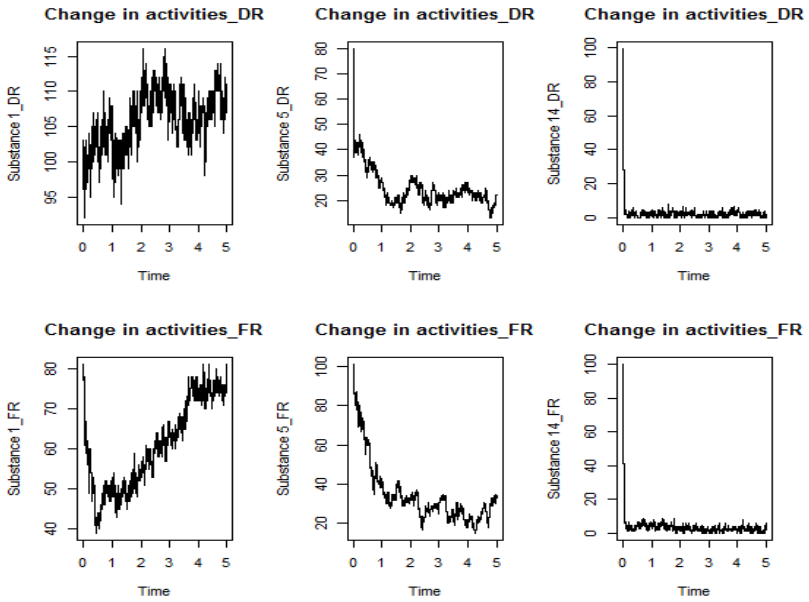

Figure 35: (Scenario 2) The plots indicate the changes in the number of Substance 1, Substance 5 and hazard of PKC Model for *Direct Method* and *First Reaction Method* for **the fixed time 5**, respectively, when the system has impulsive effect.

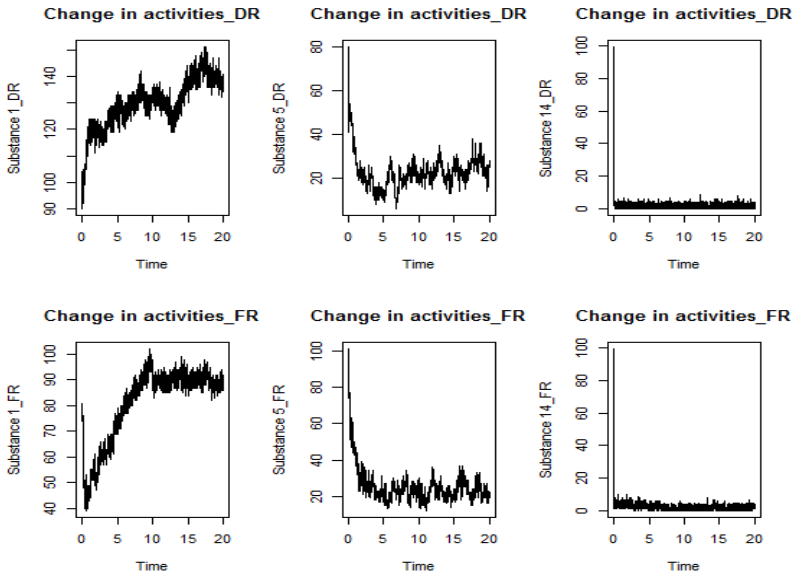

Figure 36: (Scenario 2) The plots indicate the changes in the number of Substance 1, Substance 5 and hazard of PKC Model for *Direct Method* and *First Reaction Method* for the **fixed time 20**, respectively, when the system has impulsive effect.

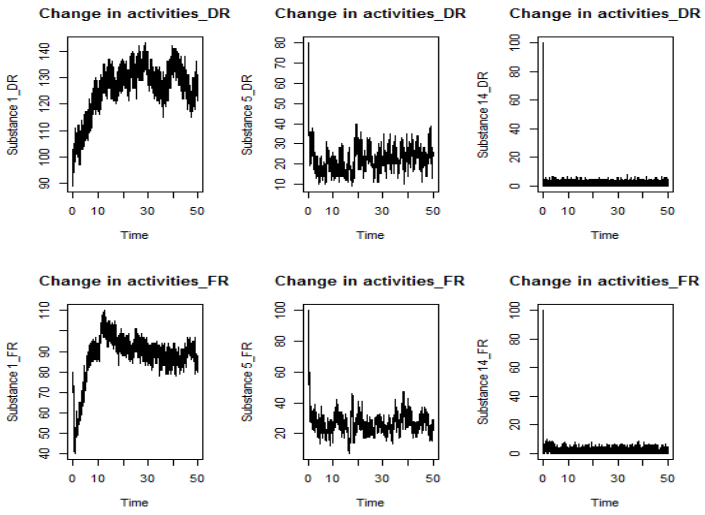

Figure 37: (Scenario 2) The plots indicate the changes in the number of Substance 1, Substance 5 and hazard of PKC Model for *Direct Method* and *First Reaction Method* for **the fixed time 50**, respectively, when the system has impulsive effect.

Figures 32-37 show the impulsive effects of the PKC system under different simulation periods. In these plots we indicate that the number of the Substance 1 increases while the number of Substance 5 decreases. Like the LV system, we cannot detect impulse effect clearly due to the sharp decrease. In addition, the impulse effect is more visible for Substance 1, especially, for long run simulations.

From Table 6, there is no significant difference between two methods. However, PKC has higher difference compared to LV system. In addition, the difference increases when system has long run.

Substances 1, 15 and 30 are selected arbitrarily for impulsive effects in the JAK-STAT system. For this system, the impulsive effect for the first scenario depends on the sudden increase of the Substance 15 and the second scenario relies on the impulsive effect for the Substance 30, meaning that if its number of molecules becomes greater than 60, this number decreases by 20 units.

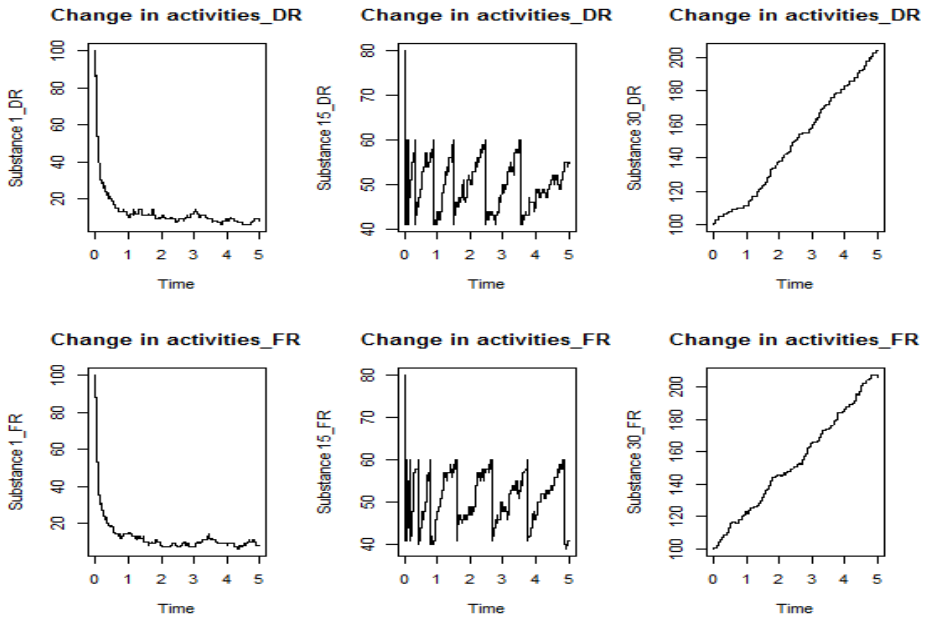

Figure 38: (Scenario 1) The plots indicate the changes in the number of the Substance 1, Substance 15 and Substance 30 of JAK-STAT Model for *Direct Method* and *First Reaction Method* for the **fixed time 5**, respectively, when the system has impulsive effect.

Table 6: The results for PKC for Direct Method and First Reaction Method.

PKC		Without Impulse				With Impulse (scenario 1)				With Impulse (scenario 2)			
		CPU	Local Time	S	D	CPU	Local Time	SD	WD	CPU	Local Time	SD	WD
Direct Method	t=5	133.55	1348.92	10.21	5.87	3807.95	187167.2	10.85	5.94	3894.72	187801.2	9.94	5.87
	t=20	323.88	2526.77	9.39	5.73	3405.7	185175.6	10.47	5.87	4073.76	188761	9.38	5.80
	t=50	851.9	4730.27	9.29	5.72	3151.68	184121.8	10.47	5.85	4536.39	190665	9.19	5.75
First Reaction Method	t=5	170.35	1471.29	10.24	5.91	3848.27	187302.8	10.97	5.96	3934.43	187927.7	10.54	5.88
	t=20	478.39	3026.31	9.53	5.75	3511.34	185542.4	10.50	5.86	4201.15	189185.1	9.98	5.81
	t=50	1221	6299.52	9.28	5.70	3282.69	184537.8	10.46	5.85	4877.44	191787.9	9.75	5.76

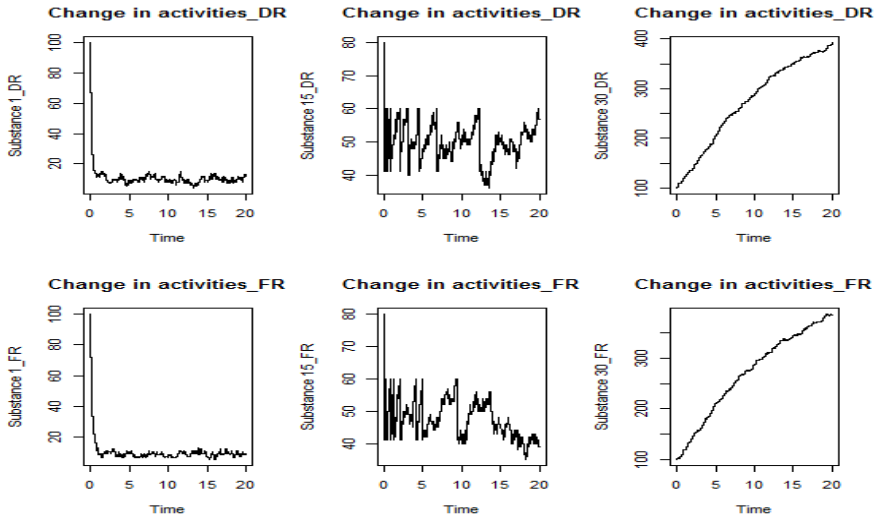

Figure 39: (Scenario 1) The plots indicate the changes in the number of the Substance 1, Substance 15 and Substance 30 of JAK-STAT Model for *Direct Method* and *First Reaction Method* for the **fixed time 20**, respectively, when the system has impulsive effect.

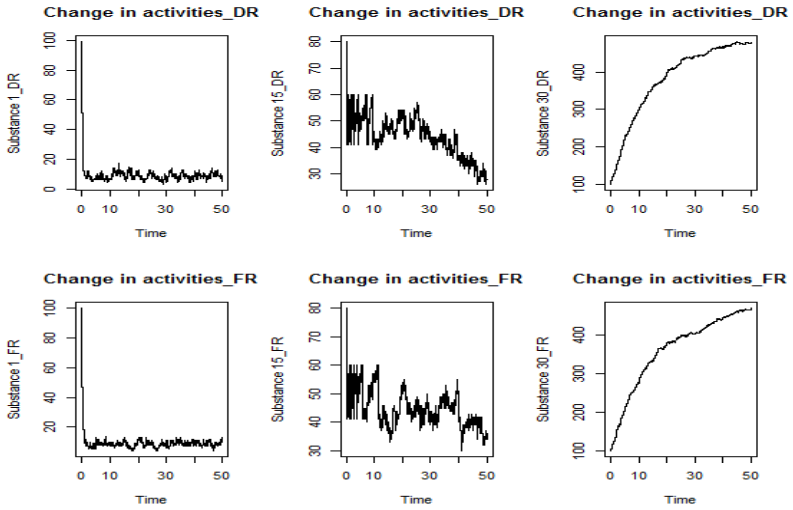

Figure 40: (Scenario 1) The plots indicate the changes in the number of the Substance 1, Substance 15 and Substance 30 of JAK-STAT Model for *Direct Method* and *First Reaction Method* for the **fixed time 50**, respectively, when the system has impulsive effect.

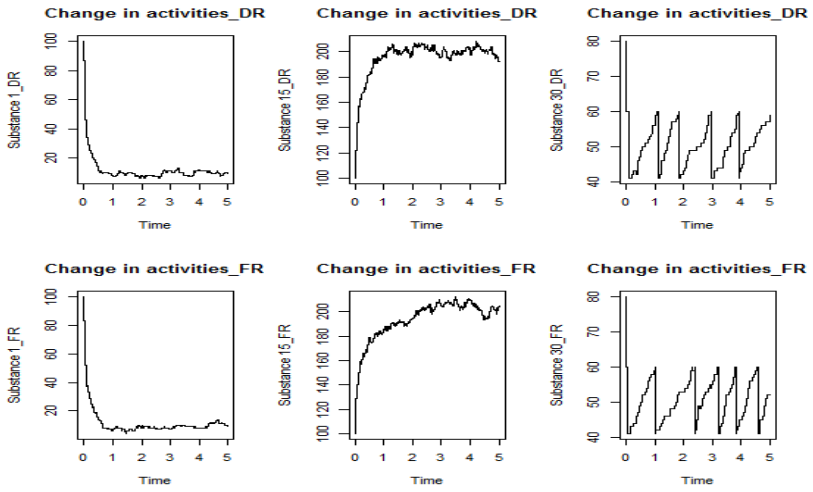

Figure 41: (Scenario 2) The plots indicate the changes in the number of the Substance 1, Substance 15 and Substance 30 of JAK-STAT Model for *Direct Method* and *First Reaction Method* for the **fixed time 5**, respectively, when the system has impulsive effect.

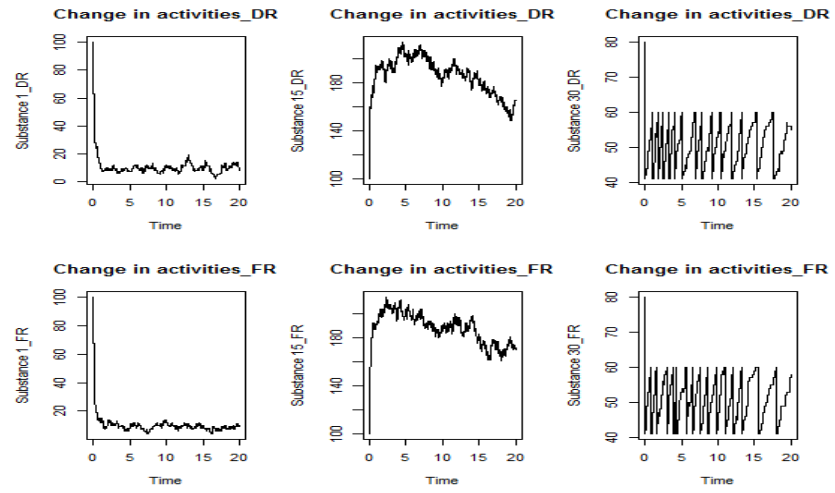

Figure 42: (Scenario 2) The plots indicate the changes in the number of the Substance 1, Substance 15 and Substance 30 of JAK-STAT Model for *Direct Method* and *First Reaction Method* for the **fixed time 20**, respectively, when the system has impulsive effect.

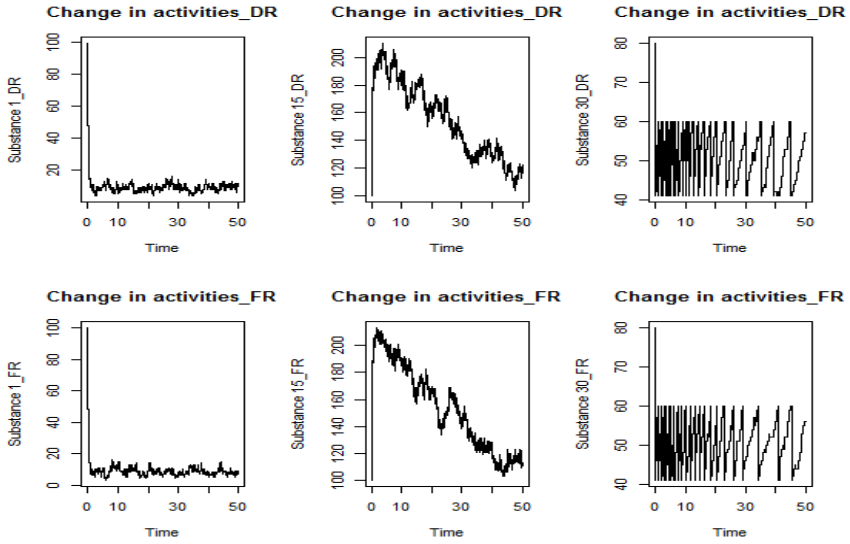

Figure 43: (Scenario 2) The plots indicate the changes in the number of the Substance 1, Substance 15 and Substance 30 of JAK-STAT Model for *Direct Method* and *First Reaction Method* for the **fixed time 50**, respectively, when the system has impulsive effect.

In Figures 38-43, there are impulsive effects of the JAK-STAT system under different simulation periods. We can see this effect for the underlying more obvious than others. Furthermore, we find that the Substance 30 is not affected by the impulses as seen for the Substance 15. Moreover, the sudden changes in the Substance 15 also affect the changes of the Substance 30. Finally, similar to other systems, this system also does not indicate any important difference in the outcomes of the Gillespie and the first reaction methods.

On the other side, from Table 7, as expected, CPU and local time increase under impulsive effects. Indeed, the JAK-STAT system has higher dimension than the ones of PKC and LV. Thus, the difference between the two exact methods is more clearly seen in this system, although the underlying difference is not significant, compared to the findings of the other systems.

In order to clearly observe the impulsive effects in particular for large systems, we further investigate the model which is composed of the convergence of the JAK-

STAT and PKC systems. As seen in Figure 44, there is no significant difference between the two exact methods for non-impulsive systems. Hereby, similar to other systems, the stochasticity is observed easily when the simulation is conducted via the Gillespie Method. In addition, to detect the impulsive effects more often in whole simulations, we change our impulsive regimes for the lower number of molecules in the sense that we increase the number of the AII hormone via 5 units when the number of this species becomes greater than 10 units. In the analysis we select this hormone since it is the growth factor which affects the complete activation of the system and the level of the AII hormone can cause the apoptosis for the system, resulting in crucial role in the flow of signals. In Figures 45-46, we show the changes in the number of the Substances 15 and 30, as used for the sole description of the JAK-STAT pathway, for the comparative purpose. The plots indicate that the effect of impulses under this system and new regime causes continuous fluctuations.

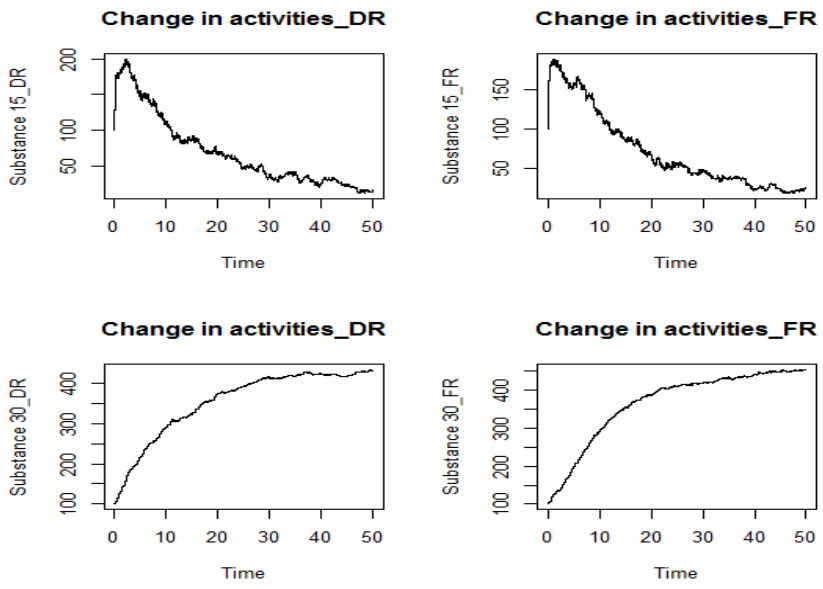

Figure 44: The plots indicate the changes in the number of molecules 15 and 30 of the convergence of PKC and JAK-STAT system via Direct Method (DR) and First Reaction Method (FR) for time 50 when the system has **no impulse**.

Table 7: The results for JAK-STAT model for *Direct Method* and *First Reaction Method*.

JAK-STAT PATHWAY		Without Impulse				With Impulse (scenario 1)				With Impulse (scenario 2)			
		CPU	Local Time	SD	WD	CPU	Local Time	SD	WD	CPU	Local Time	SD	WD
Direct Method	t=5	1259.2	8245.74	22.06	7.18	1862.15	95208.14	21.93	7.11	2294.29	97610.94	21.61	7.18
	t=20	1368.56	8870.04	23.24	7.18	1959.43	96219.07	22.43	7.17	2386.79	98328.78	22.95	7.22
	t=50	1608.52	9945.06	24.22	7.03	2135.85	96976.58	23.38	6.99	2600.19	99525.47	22.78	7.28
First Reaction Method	t=5	1286.39	8329.5	22.34	7.18	1884.21	95327.5	21.50	7.15	2316.78	97707.35	22.07	7.18
	t=20	1444.1	9107.82	23.01	7.17	2028.8	96482.65	22.48	7.12	2458.54	98606.32	22.92	7.18
	t=50	1776.67	10587.56	23.76	7.11	2262.23	97368.3	23.27	7.00	2777.31	182215.3	22.56	7.25

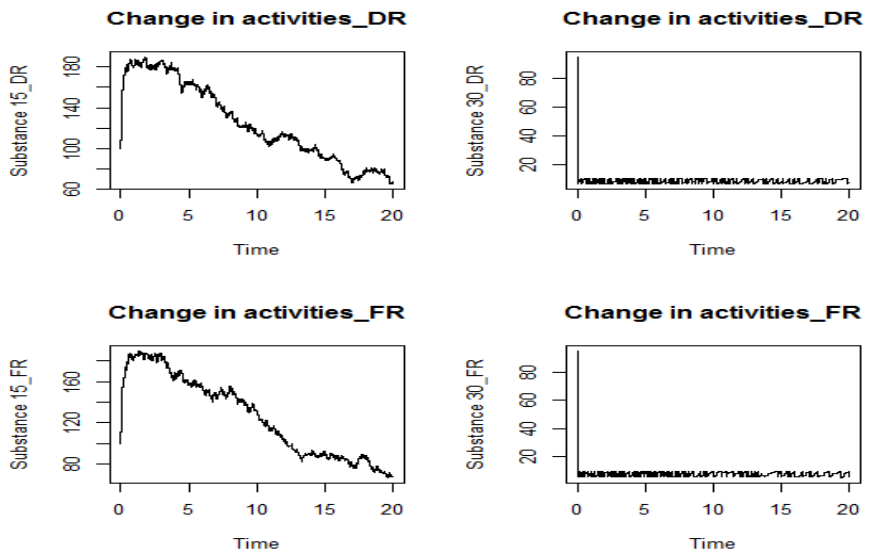

Figure 45: The plots indicate the changes in the number of molecules 15 and 30 of the convergence of PKC and JAK-STAT system via Direct Method (DR) and First Reaction Method (FR) for time 20 when the system has **impulsive effect**.

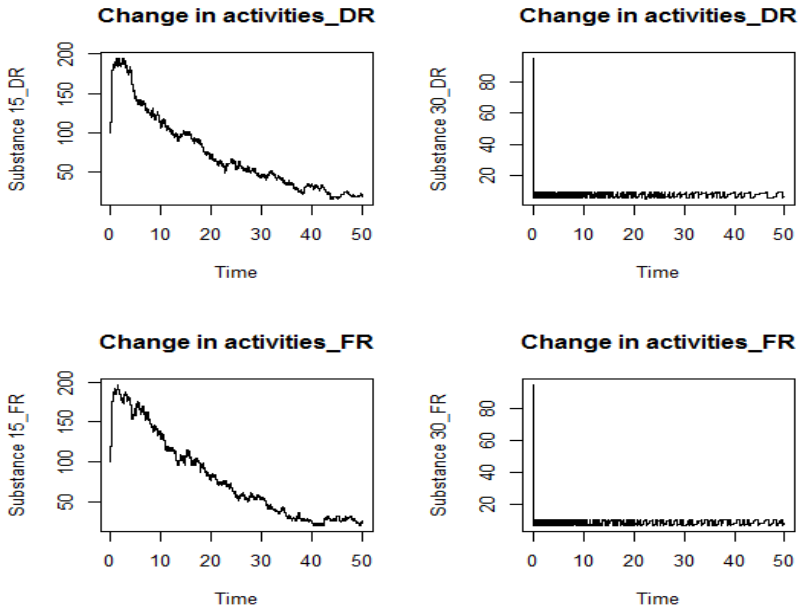

Figure 46: The plots indicate the changes in the number of molecules 45 of the convergence of PKC and JAK-STAT system via *Direct Method* (DR) and *First Reaction Method* (FR) for time 50 when the system has **impulsive effect**.

According to Table 8, the results for both systems are very close to each other, apart from the outcomes under the simulation time 50. The impulses affect, in particular, the CPU time of the calculation.

Table 8: The results for the convergence of JAK-STAT and PKC systems for *Direct Method* and *First Reaction Method*.

		Without Impulse				**With Impulse**			
		CPU	Local Time	S	D	CPU	Local Time	S	D
Direct Method	**t=20**	228.3	1300.09	61.14	6.06	1241.78	140733.7	54.37	5.70
	t=50	1703.23	8015.76	62.62	5.87	3190.51	146804.3	58.28	5.48
First Reaction Method	**t=20**	486.58	2290.65	61.50	6.07	1484.69	141549.3	54.39	5.74
	t=50	1081.41	5585.89	62.84	5.83	3714.02	148288.5	57.87	5.53

CHAPTER 4

CONCLUSION

In this study we have dealt with the stochastic simulation of the biochemical systems by performing two major algorithms in this field. These are the Gillespie and the first reaction methods. They are exact algorithms in the sense that they are based on the chemical master equation. Here, we have evaluated them under different dimensional systems and impulsive scenarios. The impulses refer to the sudden changes in the system and are typically used to explain the sudden increasing/decreasing effects of a particular species in the system and can cause diseases such as cancer or diabetes. Thereby, their roles in biological systems are very crucial. In this study, we have considered to include them in the two well-known exact stochastic simulation algorithms in order to compare their performances in terms of the computational demand. For the measure of comparison, we have selected the central processing unit, local time, average search depth and the average weighted degree.

As a result, from the analyses, we have used small and large systems, namely, Lotka Volterra, PKC, JAK-STAT, Lysis-Lysogeny and the combination of JAK-STAT and PKC pathways. Then we have performed two underlying methods for time t= 5, 20 and 50 so that the behavior of the systems and algorithms can be evaluated under both short and long time periods. From the findings, we have observed that there is no significant difference between the Gillespie and the first reaction methods in most of the pathways and simulation times. But a slight difference is seen when the system gets complicated. Moreover, we have detected that the effect of impulses can be observable for large systems and can be hardly seen under small or moderate dimensional systems. Furthermore, if the impulses are put on the initial species which can trigger whole activation of the systems such as the growth factor or external stimulus, the activations of all remaining species alter too as expected and the impulses becomes visible in the outputs of the simulations. On the other hand, if

the underlying effects, either in increasing or decreasing direction in the number of molecules, are frequently fired, they cannot be observed clearly. Whereas, such effects have always caused an increase in CPU and local time as expected and these differences become obvious while the systems have complex structures, i.e. higher dimensions.

As a future work, we consider to include other exact stochastic simulation algorithms in the analyses under different dimensional systems and distinct impulsive scenarios. Furthermore the stochastic bifurcation analysis can be investigated for these systems since the impulses can cause unstable position for the systems and the bifurcation is detected under such instability. Finally, we think to extend this study for approximate stochastic simulation algorithms under various conditions of impulses.

REFERENCES

- Altıntan, Derya, Purutçuoğlu, Vilda and Uğur, Ömür (2014). Impulsive Expression in Stochastic Simulation Algorithms (Under review).

- Arkin, Adam, Ross John and McAdams Harley (1998). Stoc-hastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage λ-Infected Escherichia coli Cells. Genetic, 149 (4), 1633-1648.

- Belle, Van, Taylor, Herrath Von (2009). Mouse Models for Type 1 Diabetes. Drug Discovery Today Disease Models, 6 (2), 41–45.

- Bolouri, Hamid (2008). Computational Modeling of Gene Regulatory Networks. Imperial College Press: 2008.

- Bower, James M. and Bolouri, Hamid (2001). Computational Modeling of Genetic and Biochemical Networks, MIT Press.

- Cao, Yang, Hong, Li and Petzold, Linda (2004). Efficient Formulation Of The Stochastic Simulation Algorithm For Chemically Reacting Systems. Journal Of Chemical Physics, 121 (9).

- Gibson, Michael A. and Bruck, Jehoshua (2000). Efficient exact stochastic simulation of chemical systems with many species and many channels. Journal of Physical Chemistry, 104 (9), 1876-1889.

- Gillespie, Daniel T. (1977). Exact Stochastic Simulation of Coupled Chemical Reactions. Journal of Physical Chemistry, 81 (25), 2340–2361.

- Gillespie, Daniel T. (1992). A Rigorous Derivation of the Chemical Master Equation. Physica A, 188, 404-425.

- Haley, John, Gullick, William John (2008). EGFR Signalling Networks in Cancer Therapy. Humana Press, Network.

- Hornberg, Jorrit (2005). Towards integrative tumor cell biology control of MAP kinase signalling. PhD thesis, Vrije Universiteit, Amsterdam.

- InvivoGen (2005). Type I IFN Production and Signaling Review, http://www.invivogen.com/review-type1-ifn-production [Last visited on 12 October 2015].

- Jason, Rosler, Kriston and Harrison, Douglas (2014). The JAK/STAT Signaling Pathway, http://jcs.biologists.org/content/117/8/1281.full [Last visited on 1 October 2014].

- Jong, Hidde De (2002). Modeling and Simulation of Genetic Regulatory Systems: A Literature Review. Journal of Computational Biology, 9, 67-103.

- Jonge, Peter, Alonso, Jordi, Stein, Dan, Kiejna, Andrzej, Aguilar Gaxiola, Sergio, Viana, Maria Carmen et al. (2014). Associations between DSM-IV mental disorders and diabetes mellitus: a role for impulse control disorders and depression, Diabetologia, 57 (4), 699–709.

- Jun Wang (2005). Convergence of Protein Kinase C and JAK-STAT Signaling on Transcription Factor GATA-4. Molecular and Cellular Biology, 25 (22), 9829–9844.

- Kolch, Walter (2000). Meaningful Relationships: The Regulation of the Ras/Raf/MEK/ERK Pathway by Protein Interactions. Biochemical Journal, 351, 289–305.

- Kolch, Walter, Calder, M., Gilbert, D. (2005). When Kinases Meet Mathematics: The Systems Biology of MAPK Signalling. FEBS Letters, 579, 1891–1895.

- Maiwald, Tim, Schneider, Annette, Busch, Hauke, Sahle, Sven, Gretz, Norbert, Weiss, Thomas S.,Kummer, Ursula, and Klingmüller, Ursula. (2010). Combining Theoretical Analysis and Experimental Data Generation Reveals IRF9 as a Crucial Factor for Accelerating Interferon Alpha-Induced Early Anti viral Signalling. The FEBS Journal, 277 (22), 4741–54.

- Manninen, Tiina, Makiraatikka Eeva (2006). Discrete Stochastic Simulation of Cell Signaling: Comparison of Computational Tools. Conference Proceedings IEEE Engineering Medicine and Biology Society, 1, 2013-2016.

- McCollum, J.M., Peterson, J.D. (2006). The Sorting Direct Method for Stochastic Simulation of Biochemical Systems with Varying Reaction Execution Behavior. Computational Biology and Chemistry, 30 (1), 39-49.

- McMahon, Mary (2014). What is RNA?, http://www.wisegeek.com/what-is-rna.htm

- National Human Genome Research Institute (2014). Deoxyribonucleic Acid (DNA), http://www.genome.gov/25520880 [Last visited on 7 August 2014].

- Purutçuoğlu, Vilda (2010). Survey on Stochastic Simulations for Biochemical Reactions (Biyokimyasal Reaksiyonlar İçin Stokastik Simulasyon Algoritmalarına Genel Bir Bakış). Journal of Statistical Research, 7, 70-82.

- Purutçuoğlu, Vilda, Ayyıldız, Ezgi (2014). Statistics for Bioinformatics (Biyoinformatik Alanında İstatistik). Nobel Press. Ankara

- Rosen, Olov (2013). Parallelization of Stochastic Estimation Algorithms on Multicore Computational Platforms, Depart-ment of Information Technology Uppsala University.

- Schaechter, J.D and Benowitz, L.I. (1993). Activation of Protein Kinase C by Arachidonic Acid Selectively Enhances the Phosphorylation of GAP-43 in Nerve Terminal Membranes. Journal of Neuroscience, 13 (10), 4361-4371.

- Selice Wordexpress (2011). Interferon Regulatory Factor(IRF) Family. https://selice.wordpress.com/2011/01/ [Last visited on 6 February 2015].

- The RNA Society (2014). What is RNA?, http://www.rnasociety.org/about/what-is-rna/ [Last visited on 7 August 2015].

- Turner, Thomas E., Schnell, Santiago and Burrage, Kevin. (2004). Stochastic Approaches for Modelling in Vivo Reactions. Computational Biology and Chemistry, 28, 165–178.

- U.S. National Library of Medicine (2014). What is DNA?, http://www.ghr.nlm.nih.gov/handbook/basics/dna, 2014 [Last visited on 12 June 2014].

APPENDIX

Appendix A

The reaction list of the Lotka Volterra system (Wilkinson, 2006) is presented below. In this system, there are 2 substrates and 3 reactions.

1. $Y_1 \xrightarrow{k_1} 2Y_1$
2. $Y_1 + Y_2 \xrightarrow{k_2} 2Y_2$
3. $Y_2 \xrightarrow{k_3} \emptyset$

The differential equations of the system as the mathematical description, is described as below:

1. $\frac{d_{y_1}}{d_t} = k_1.[Y_1] - k_2.[Y_1].[Y_2]$
2. $\frac{d_{y_2}}{d_t} = k_2.[Y_1].[Y_2] - k_3.[Y_2]$

Appendix B

The reaction list of the PKC pathway (Schaechter and Benowitz, 1993) is presented below. There are 14 substrates and 10 reactions in this system.

1. $PKC - cytosolic + AA \overset{k_1}{\rightarrow} PKC.AA$
2. $PKC - cytosolic + Ca^{2+} \overset{k_2}{\rightarrow} PKC.Ca^{2+}$
3. $PKC.Ca^{2+} + AA \overset{k_3}{\rightarrow} PKC.Ca^{2+}.AA^{*}$
4. $DAG + PKC.Ca^{2+} \overset{k_4}{\rightarrow} PKC.Ca^{2+}.DAG$
5. $PKC.DAG.AA \overset{k_5}{\rightarrow} PKC.DAG.AA^{*}$
6. $PKC - cytosolic \overset{k_6}{\rightarrow} PKC - basal$
7. $PKC.Ca^{2+} \overset{k_7}{\rightarrow} PKC.Ca^{2+}.memb$
8. $PKC.Ca^{2+}.DAG \overset{k_8}{\rightarrow} PKC.DAG.memb$
9. $PKC - cytosolic.DAG \overset{k_9}{\rightarrow} PKC.DAG$
10. $PKC.DAG + AA \overset{k_{10}}{\longrightarrow} PKC.DAG.AA$

The differential equations of the system as the mathematical description, is described as below:

1. $\frac{d_{PKC-cytosolic}}{d_t} = -k_1[PKC - cytosolic]\,[AA] - k_2\,[PKC - cytosolic].[Ca^{2+}] - k_6\,[PKC - cytosolic]$
2. $\frac{d_{AA}}{d_t} = -k_1[PKC - cytosolic][AA] - k_3\,[PKC.Ca^{2+}].[AA] - k_{10}\,[PKC.DAG].[AA]$
3. $\frac{d_{PKC.AA}}{d_t} = k_1[PKC - cytosolic][AA]$
4. $\frac{d_{PKC.Ca^{2+}}}{d_t} = k_2[PKC - cytosolic].[Ca^{2+}] - k_3\,[PKC.Ca^{2+}].[AA] - k_4\,[DAG][PKC.Ca^{2+}] - k_7[PKC.Ca^{2+}]$

5. $\frac{d_{PKC.Ca^{2+}.AA^*}}{d_t} = k_3[PKC.Ca^{2+}][AA]$

6. $\frac{d_{DAG}}{d_t} = -k_4\,[DAG][PKC.Ca^{2+}]$

7. $\frac{d_{PKC.Ca^{2+}.DAG}}{d_t} = k_4\,[DAG][PKC.Ca^{2+}] - k_8[PKC.Ca^{2+}.DAG]$

8. $\frac{d_{PKC.DAG.AA}}{d_t} = k_{10}\,[PKC.DAG][AA] - k_5[PKC.DAG.AA]$

9. $\frac{d_{PKC.DAG.AA^*}}{d_t} = k_5[PKC.DAG.AA]$

10. $\frac{d_{PKC-basal}}{d_t} = k_6[PKC - cytosolic]$

11. $\frac{d_{PKC.Ca^{2+}.memb}}{d_t} = k_7[PKC.Ca^{2+}]$

12. $\frac{d_{PKC.DAG.memb}}{d_t} = k_8[PKC.Ca^{2+}.DAG]$

13. $\frac{d_{PKC-cytosolic.DAG}}{d_t} = k_9[PKC - cytosolic.DAG]$

14. $\frac{d_{PKC.DAG}}{d_t} = k_{10}[PKC.DAG][AA] - k_9[PKC - cytosolic.DAG]$

Appendix C

Below, we present the reactions list of the JAK-SAT pathway as applied in the study of Maiwald et al. (2010). The system is defined via 38 substrates and 66 reactions. The description of each substrate can be found in the same study.

1. Receptor IFNAR 1 + TYK $\xrightarrow{k_1}$ Receptor TYK Complex

2. Receptor TYK Complex $\xrightarrow{k_2}$ Receptor IFNAR 1 + TYK

3. Receptor IFNAR 2 + JAK $\xrightarrow{k_3}$ Receptor JAK Complex

4. Receptor JAK Complex $\xrightarrow{k_4}$ Receptor IFNAR 2 + JAK

5. Receptor JAK Complex + Receptor TYK Complex + IFN_free $\xrightarrow{k_5}$ IFNAR dimer

6. IFNAR dimer $\xrightarrow{k_6}$ Receptor JAK Complex + Receptor TYK Complex + IFN_free

7. IFNAR dimer $\xrightarrow{k_7}$ Active Receptor Complex

8. STAT2c_IRF9 + Active Receptor Complex $\xrightarrow{k_8}$ Active Receptor Complex_STAT2c + IRF9c

9. STAT2c + Active Receptor Complex $\xrightarrow{k_9}$ Active Receptor Complex_STAT2c

10. Active Receptor Complex_STAT2c $\xrightarrow{k_{10}}$ STAT2c + Active Receptor Complex

11. STAT1c + Active Receptor Complex_STAT2c $\xrightarrow{k_{11}}$ Active Receptor Complex_STAT2c_STAT1c

12. Active Receptor Complex_STAT2c_STAT1c $\xrightarrow{k_{12}}$ STAT1c + Active Receptor Complex_STAT2c

13. Active Receptor Complex_STAT2c_STAT1c $\xrightarrow{k_{13}}$ Active Receptor Complex + STAT1c*_STAT2c*

14. IRF9c + STAT1c*_STAT2c*$\xrightarrow{k_{14}}$ ISGF-3c

15. ISGF-3c $\xrightarrow{k_{15}}$ IRF9c + STAT1c*_STAT2c*

16. ISGF-3c $\xrightarrow{k_{16}}$ ISGF-3n

17. ISGF-3n $\xrightarrow{k_{17}}$ ISGF-3c

18. STAT1c*_STAT2c* $\xrightarrow{k_{18}}$ STAT1n*_STAT2n*

19. STAT1n*_STAT2n* $\xrightarrow{k_{19}}$ STAT1c*_STAT2c*

20. STAT1n*_STAT2n* + IRF9n $\xrightarrow{k_{20}}$ ISGF-3n

21. ISGF-3n $\xrightarrow{k_{21}}$ STAT1n*_STAT2n* + IRF9n

22. ISGF-3n + Free transcription factor binding site (TFBS) $\xrightarrow{k_{22}}$ Occupied TFBS

23. Occupied TFBS $\xrightarrow{k_{23}}$ ISGF-3n + Free transcription factor binding site (TFBS)

24. $\emptyset \xrightarrow{k_{24}}$mRNAn

25. mRNAn$\xrightarrow{k_{25}}$ mRNAc

26. mRNAc$\xrightarrow{k_{26}}$ mRNAn

27. $\emptyset \xrightarrow{k_{27}}$ IRF9c

28. IRF9c $\xrightarrow{k_{28}}$ $\emptyset$

29. $\emptyset \xrightarrow{k_{29}}$ SOCS

30. SOCS $\xrightarrow{k_{30}}$ $\emptyset$

31. Active Receptor Complex $\xrightarrow{k_{31}}$ Receptor IFNAR1 +Receptor IFNAR2 + JAK + TYK

32. Active Receptor Complex $\xrightarrow{k_{32}}$ IFNAR dimer

33. IRF9n $\xrightarrow{k_{33}}$ ∅

34. STAT2c_IRF9 $\rightarrow\xrightarrow{k_{34}}$ STAT2c

35. STAT2n_IRF9 $\xrightarrow{k_{35}}$ STAT2n

36. ISGF-3c + Cytoplasmic phosphatase (CP) $\xrightarrow{k_{36}}$ ISGF-3c_CP

37. ISGF-3c_CP $\xrightarrow{k_{37}}$ ISGF-3c + Cytoplasmic phosphatase (CP)

38. ISGF-3c_CP $\xrightarrow{k_{38}}$ STAT1c +STAT2c + CP + IRF9c

39. STAT1c*_STAT2c* + CP $\xrightarrow{k_{39}}$ STAT1c*_STAT2c*_CP

40. STAT1c*_STAT2c*_CP $\xrightarrow{k_{40}}$ STAT1c*_STAT2c* + CP

41. STAT1c*_STAT2c*_CP $\xrightarrow{k_{41}}$ STAT1c + STAT2c + CP

42. STAT1n*_STAT2n* + Nuclear phosphatase (NP) $\xrightarrow{k_{42}}$ STAT1n*_STAT2n*_NP

43. STAT1n*_STAT2n*_NP $\xrightarrow{k_{43}}$ STAT1n*_STAT2n* + Nuclear phosphatase (NP)

44. STAT1n*_STAT2n*_NP $\xrightarrow{k_{44}}$ STAT1n + STAT2n + NP

45. ISGF-3n + NP $\xrightarrow{k_{45}}$ ISGF-3n_NP

46. ISGF-3n_NP $\xrightarrow{k_{46}}$ ISGF-3n + NP

47. ISGF-3n_NP $\xrightarrow{k_{47}}$ STAT1n + STAT2n + NP + IRF9n

48. Occupied TFBS + NP $\xrightarrow{k_{48}}$ Occupied TFBS_NP

49. Occupied TFBS_NP $\xrightarrow{k_{49}}$ Occupied TFBS + NP

50. Occupied TFBS_NP $\xrightarrow{k_{50}}$ STAT1n + STAT2n + Free TFBS +IRF9n

51. PIAS + ISGF-3n $\xrightarrow{k_{51}}$ PIAS_ISGF-3n

52. PIAS_ISGF-3n $\xrightarrow{k_{52}}$ PIAS + ISGF-3n

53. mRNAc$\xrightarrow{k_{53}}$ ∅

54. STAT1c $\xrightarrow{k_{54}}$ STAT1n

55. STAT1n $\xrightarrow{k_{55}}$ STAT1c

56. STAT2c $\xrightarrow{k_{56}}$ STAT2n

57. STAT2n $\xrightarrow{k_{57}}$ STAT2c

58. STAT2c + IRF9c $\xrightarrow{k_{58}}$ STAT2c_IRF9

59. STAT2c_IRF9 $\xrightarrow{k_{59}}$ STAT2c + IRF9c

60. STAT2n + IRF9n $\xrightarrow{k_{60}}$ STAT2n_IRF9

61. STAT2n_IRF9 $\xrightarrow{k_{61}}$ STAT2n + IRF9n

62. STAT2c_IRF9 $\xrightarrow{k_{62}}$ STAT2n_IRF9

63. STAT2n_IRF9 $\xrightarrow{k_{63}}$ STAT2c_IRF9

64. IRF9c $\xrightarrow{k_{64}}$ IRF9n

65. IRF9n $\xrightarrow{k_{65}}$ IRF9c

66. IFN_influx$\xrightarrow{k_{66}}$ IFN_free

The differential equations of the system as the mathematical description, is described as below:

1. $\frac{d_{Receptor\ IFNAR\ 1}}{d_t} =$

   $k_2 Receptor\ TYK\ Complex - k_1[Receptor\ IFNAR\ 1][TYK] +$

   $k_{31}[Active\ Receptor\ Complex]$
2. $\frac{d_{TYK}}{d_t} = k_2 Receptor\ TYK\ Complex - k_1[Receptor\ IFNAR\ 1][TYK] +$

   $k_{31}[Active\ Receptor\ Complex]$
3. $\frac{d_{Receptor\ TYK\ Complex}}{d_t} =$

   $k_1[Receptor\ IFNAR\ 1][TYK] - k_2 Receptor\ TYK\ Complex -$

   $k_5[Receptor\ JAK\ Complex][Receptor\ TYK\ Complex][IFN_{free}] +$

   $k_6[Receptor\ JAK\ Complex][Receptor\ TYK\ Complex][IFN_{free}]$
4. $\frac{d_{Receptor\ IFNAR\ 2}}{d_t} =$

   $k_4[Receptor\ IFNAR\ 2][JAK] - k_3[Receptor\ IFNAR\ 2][JAK] +$

   $k_{31}[Active\ Receptor\ Complex]$
5. $\frac{d_{JAK}}{d_t} = k_4[Receptor\ IFNAR\ 2][JAK] - k_3[Receptor\ IFNAR\ 2][JAK] +$

   $k_{31}[Active\ Receptor\ Complex]$
6. $\frac{d_{Receptor\ JAK\ Complex}}{d_t} =$

   $k_3[Receptor\ IFNAR\ 2][JAK] - k_4[Receptor\ IFNAR\ 2][JAK] -$

   $k_5[Receptor\ JAK\ Complex][Receptor\ TYK\ Complex]\left[IFN_{free}\right] +$

   $k_6[Receptor\ JAK\ Complex][Receptor\ TYK\ Complex][IFN_{free}]$
7. $\frac{d_{IFN_{free}}}{d_t} =$

   $-k_5[Receptor\ JAK\ Complex][Receptor\ TYK\ Complex]\left[IFN_{free}\right] +$

   $k_6[Receptor\ JAK\ Complex][Receptor\ TYK\ Complex]\left[IFN_{free}\right] +$

   $k_{66}\ [IFN\_influx]$
8. $\frac{d_{IFNAR\ dimer}}{d_t} =$

   $k_5[Receptor\ JAK\ Complex][Receptor\ TYK\ Complex]\left[IFN_{free}\right] -$

$k_6[Receptor\ JAK\ Complex][Receptor\ TYK\ Complex][IFN_{free}] -$

$k_7 IFNAR\ dimer + k_{32}[Active\ Receptor\ Complex]$

9. $\frac{d_{STAT2c\_IRF9}}{d_t} =$

$k_{58}\ [STAT2c][IRF9c] - k_8[STAT2c_{IRF9}][Active\ Receptor\ Complex] -$

$k_{59}\ [STAT2c\_IRF9]$

10. $\frac{d_{Active\ Receptor\ Complex}}{d_t} =$

$k_{13}[Active\ Receptor\ Complex_{STAT2c_{STAT1c}}] + k_{10}[Active\ Receptor\ Complex_{STAT2c}] +$

$k_7[IFNAR\ dimer] -$

$k_8[STAT2c_{IRF9}][Active\ Receptor\ Complex] - k_9[STAT2c]\ [Active\ Receptor\ Complex] -$

$k_{31}[Active\ Receptor\ Complex] - k_{32}[Active\ Receptor\ Complex]$

11. $\frac{d_{Active\ Receptor\ Complex\_STAT2c}}{d_t} =$

$k_{12}[Active\ Receptor\ Complex_{STAT2c_{STAT1c}}] +$

$k_8[STAT2c_{IRF9}][Active\ Receptor\ Complex] + k_9[STAT2c][Active\ Receptor\ Complex] -$

$k_{10}[Active\ Receptor\ Complex\_STAT2c\ ] - k_{11}[STAT1c][Active\ Receptor\ Complex\_STAT2c]$

12. $\frac{d_{IRF9c}}{d_t} =$

$k_8[STAT2c_{IRF9}][Active\ Receptor\ Complex] -$

$k_{14}[IRF9c][\ STAT1c\ *_{STAT2c}*] + k_{15}[IRF9c][\ STAT1c\ *_{STAT2c}*] +$

$k_{38}[ISGF - 3c_{CP}] - k_{58}\ [STAT2c][IRF9c] + k_{59}\ [STAT2c_{IRF9}] -$

$k_{64}\ [IRF9c] + k_{65}\ [IRF9n]$ -$k_{28}\ [IRF9c]$

13. $\frac{d_{STAT2c}}{d_t} = k_{10}[Active\ Receptor\ Complex_{STAT2c}] - k_9[STAT2c][Active\ Receptor\ Complex] +$

$k_{34}[STAT2c\_IRF9\ ] + k_{38}[ISGF - 3c\_CP] + k_{41}[STAT1c\ *_{STAT2c}*\ \_CP] - k_{56}\ [STAT2c] +$

$k_{57}\ [STAT2n] - k_{58}\ [STAT2c][IRF9c] + k_{59}\ [STAT2c\_IRF9]$

14. $\frac{d_{STAT1c}}{d_t} =$

$k_{12}[Active\ Receptor\ Complex\_STAT2c\_STAT1c] -$

$k_{11}[STAT1c][Active\ Receptor\ Complex_{STAT2c}] +$

$k_{38}[ISGF - 3c\_CP] + k_{41}[STAT1c\ *_{STAT2c}*\ \_CP] - k_{54}[STAT1c] +$

$k_{55}\ [STAT1n]$

15. $\frac{d_{Active\ Receptor\ Complex\_STAT2c\_STAT1c}}{d_t} =$

$k_{11}[STAT1c][Active\ Receptor\ Complex_{STAT2c}] -$

$k_{12}[Active\ Receptor\ Complex_{STAT2c_{STAT1c}}] -$

$k_{13}[Active\ Receptor\ Complex\_STAT2c\_STAT1c]$

16. $\frac{d_{STAT1c*\_STAT2c*}}{d_t} =$

$k_{19}[\ STAT1n *_{STAT2n}*] + k_{13}[Active\ Receptor\ Complex_{STAT2c_{STAT1c}}] -$

$k_{14}[IRF9c][\ STAT1c *_{STAT2c}*] + k_{15}[IRF9c][\ STAT1c *_{STAT2c}*] -$

$k_{18}[\ STAT1c *_{STAT2c}*] + k_{40}[STAT1c *_{STAT2c}* \_CP] - k_{39}[STAT1c *_{STAT2c}*$

$][CP]$

17. $\frac{d_{ISGF-3c}}{d_t} = k_{14}[IRF9c][\ STAT1c *_{STAT2c}*] - k_{15}[IRF9c][\ STAT1c *_{STAT2c}*] -$

$k_{16}[ISGF - 3c\ ] + k_{17}[ISGF - 3n\ ] -$

$k_{36}[ISGF - 3c]\ [Cytoplasmic\ phosphatase\ (CP)] + k_{37}[ISGF - 3c\_CP]$

18. $\frac{d_{ISGF-3n}}{d_t} =$

$k_{16}[ISGF - 3c\ ] - k_{17}[ISGF - 3n\ ] + k_{20}[\ STAT1n *_{STAT2n}*][IRF9n] -$

$k_{21}[\ STAT1n *_{STAT2n}*][IRF9n] - k_{22}[\ ISGF - 3n][TFBS] +$

$k_{23}[\ Occupied\ TFBS] - k_{45}[ISGF - 3n][NP] + k_{46}[ISGF - 3n_{NP}] -$

$k_{51}\ [PIAS]\ [ISGF - 3n] + k_{52}\ [PIAS\_ISGF - 3n]$

19. $\frac{d_{STAT1n*\_STAT2n*}}{d_t} = k_{18}[\ STAT1c *_{STAT2c}*] - k_{19}[\ STAT1n *_{STAT2n}*] -$

$k_{20}[\ STAT1n *_{STAT2n}*][IRF9n] + k_{21}[\ STAT1n *_{STAT2n}*][IRF9n] -$

$k_{42}[STAT1n *_{STAT2n}*][Nuclear\ phosphatase\ (NP)] + k_{43}[STAT1n *$

$\_STAT2n * \_NP]$

20. $\frac{d_{IRF9n}}{d_t} =$

$-k_{20}[\ STAT1n *_{STAT2n}*][IRF9n] + k_{21}[\ STAT1n *_{STAT2n}*][IRF9n] +$

$k_{47}[ISGF - 3n_{NP}] + k_{50}\ [Occupied\ TFBS_{NP}] - k_{60}\ [STAT2n][IRF9n] +$

$k_{61}\ [STAT2n_{IRF9}] + k_{64}\ [IRF9c] - k_{65}\ [IRF9n]$

21. $\frac{d_{TFBS}}{d_t} =$

$k_{23}[\ Occupied\ TFBS] - k_{22}[\ ISGF - 3n][TFBS] + k_{50}\ [Occupied\ TFBS\_NP]$

22. $\frac{d_{Occupied\ TFBS}}{d_t} = k_{22}[\ ISGF - 3n][TFBS] -\ k_{23}[\ Occupied\ TFBS] -$

$k_{48}\ [Occupied\ TFBS][NP] + k_{49}\ [Occupied\ TFBS\_NP]$

23. $\frac{d_{mRNAn}}{d_t} = k_{26}[mRNAc] - k_{25}[mRNAn]$

24. $\frac{d_{mRNAc}}{d_t} = k_{25}[mRNAn] - k_{26}[mRNAc] - k_{53}[mRNAc]$

25. $\frac{d_{STAT2c\_IRF9}}{d_t} =$

$k_{63}\,[STAT2n\_IRF9] - k_{34}[STAT2c\_IRF9\,] - k_{62}\,[STAT2c\_IRF9\,]$

26. $\frac{d_{STAT2n\_IRF9}}{d_t} =$

$k_{62}\,[STAT2c\_IRF9\,] - k_{35}[STAT2n\_IRF9\,] + k_{60}\,[STAT2n][IRF9n] -$

$k_{61}\,[STAT2n_{IRF9}] - k_{63}\,[STAT2n\_IRF9]$

27. $\frac{d_{STAT2n}}{d_t} = k_{35}[STAT2n\_IRF9\,] + k_{44}[STAT1n\,*_{STAT2n}*_{NP}] + k_{47}[ISGF -$

$3n_{NP}] + k_{50}\,[Occupied\ TFBS\_NP] + k_{56}\,[STAT2c] - k_{57}\,[STAT2n] -$

$k_{60}\,[STAT2n][IRF9n] + k_{61}\,[STAT2n_{IRF9}]$

28. $\frac{d_{CP}}{d_t} =$

$k_{37}[ISGF - 3c\_CP] - k_{36}[ISGF - 3c]\,[Cytoplasmic\ phosphatase\ (CP)]\ +$

$k_{38}[ISGF - 3c_{CP}] -\ k_{39}[STAT1c\,*_{STAT2c}*][CP] + k_{40}[STAT1c\,*_{STAT2c}*$

$\_CP] + k_{41}[STAT1c\,*_{STAT2c}*\,\_CP]$

29. $\frac{d_{ISGF-3c\_CP}}{d_t} = k_{36}[ISGF - 3c]\,[Cytoplasmic\ phosphatase\ (CP)] -$

$k_{37}[ISGF - 3c_{CP}] - k_{38}[ISGF - 3c\_CP]$

30. $\frac{d_{STAT1c*STAT2c*CP}}{d_t} =$

$k_{39}[STAT1c\,*_{STAT2c}*$

$][CP] - k_{40}[STAT1c\,*_{STAT2c}*_{CP}] - k_{41}[STAT1c\,*_{STAT2c}*_{CP}]$

31. $\frac{d_{NP}}{d_t} =$

$k_{43}[STAT1n\,*_{STAT2n}*_{NP}] -$

$k_{42}\,[STAT1n\,*_{STAT2n}*]\,[Nuclear\ phosphatase\ (NP)]\ +$

$k_{44}[STAT1n\,*_{STAT2n}*_{NP}] - k_{45}[ISGF - 3n][NP] +\ k_{46}[ISGF - 3n_{NP}] +$

$+k_{47}[ISGF - 3n_{NP}] -\ k_{48}\,[Occupied\ TFBS][NP] +$

$k_{49}\,[Occupied\ TFBS\_NP]$

32. $\frac{d_{STAT1n*\_STAT2n*\_NP}}{d_t} = k_{42}[STAT1n\,*_{STAT2n}*][Nuclear\ phosphatase\ (NP)] -$

$k_{43}[STAT1n\,*_{STAT2n}*_{NP}] - k_{44}[STAT1n * \_STAT2n * \_NP]$

33. $\frac{d_{STAT1n}}{d_t} =$

$k_{44}[STAT1n *_{STAT2n} *_{NP}] + k_{47}[ISGF - 3n_{NP}] + k_{50}\,[Occupied\ TFBS\_NP] +$

$k_{54}[STAT1c] - k_{55}\,[STAT1n]$

34. $\frac{d_{ISGF-3n_{NP}}}{d_t} = k_{44}[ISGF - 3n_{NP}] - k_{46}[ISGF - 3n_{NP}] - k_{47}[ISGF - 3n_{NP}]$

35. $\frac{d_{Occupied\ TFBS\_NP}}{d_t} = k_{48}\,[Occupied\ TFBS][NP] - k_{49}\,[Occupied\ TFBS_{NP}] -$

$k_{50}\,[Occupied\ TFBS\_NP]$

36. $\frac{d_{PIAS}}{d_t} = k_{52}\,[PIAS\_ISGF - 3n] - k_{51}\,[PIAS]\,[ISGF - 3n]$

37. $\frac{d_{PIAS_{ISGF-3n}}}{d_t} = k_{51}\,[PIAS]\,[ISGF - 3n] - k_{52}\,[PIAS_{ISGF} - 3n]$

38. $\frac{d_{IFN_{influx}}}{d_t} = k_{66}\,\left[IFN_{influx}\right]$

Appendix D

Below, we present the reactions list of the Lysis-Lysogency system as applied in the study of Arkin et al. (1998). The system is defined by 12 substrates and 16 reactions. The description of each substrate can be found in the same study.

1. $CI \overset{k_1}{\rightarrow} \emptyset$
2. $2CI \overset{k_2}{\rightarrow} CI_2$
3. $CI_2 \overset{k_3}{\rightarrow} 2CI$
4. $Cro \overset{k_4}{\rightarrow} \emptyset$
5. $2Cro \overset{k_5}{\rightarrow} Cro_2$
6. $Cro_2 \overset{k_6}{\rightarrow} 2Cro$
7. $N \overset{k_7}{\rightarrow} \emptyset$
8. $CII + P1 \overset{k_8}{\rightarrow} P1.CII$
9. $P1.CII \overset{k_9}{\rightarrow} CII + P1$
10. $P1.CII \overset{k_{10}}{\longrightarrow} P1$
11. $CII + P2 \overset{k_{11}}{\longrightarrow} P2.CII$
12. $P2.CII \overset{k_{12}}{\longrightarrow} CII + P2$
13. $P2.CII \overset{k_{13}}{\longrightarrow} P2$
14. $CIII + P2 \overset{k_{14}}{\longrightarrow} P2.CIII$
15. $P2.CIII \overset{k_{15}}{\longrightarrow} CIII + P2$
16. $P2.CIII \overset{k_{16}}{\longrightarrow} P2$

The differential equations of the system as the mathematical description, is described as below:

1. $\frac{d_{\mathrm{CI}}}{d_t} = -k_1\,[\mathrm{CI}] - 2k_2[CI]^2 + 2k_3[CI]^2$

2. $\frac{d_{Cl_2}}{d_t} = k_2[CI]^2 - k_3[CI]^2$

3. $\frac{d_{Cro}}{d_t} = k_4[Cro] - k_5 2[Cro]^2 + 2k_6[Cro_2]^2$

4. $\frac{d_{Cro_2}}{d_t} = k_5[Cro]^2 - k_6[Cro_2]^2$

5. $\frac{d_N}{d_t} = -k_7\ [\mathrm{N}]$

6. $\frac{d_{\mathrm{CII}}}{d_t} = k_9\ [P1.CII] - k_8\ [CII][P1] - k_{11}\ [P2][CII]$

7. $\frac{d_{\mathrm{P1}}}{d_t} = k_9\ [P1.CII] - k_8\ [CII][P1] + k_{10}\ [P1.CII]$

8. $\frac{d_{\mathrm{P1.CII}}}{d_t} = k_8\ [CII][P1] - k_9\ [P1.CII] - k_{10}\ [P1.CII] + k_{12}\ [P2.CII]$

9. $\frac{d_{\mathrm{P2}}}{d_t} = k_{12}\ [P2.CII] - k_{11}\ [P2][CII] + k_{13}\ [P2.CII] - k_{14}\ [CIII][P2] +$

   $k_{15}\ [P2.CIII] + k_{16}\ [P2.CIII]$

10. $\frac{d_{\mathrm{P2CII}}}{d_t} = k_{11}\ [P2][CII] - k_{12}\ [P2.CII] - k_{13}\ [P2.CII]$

11. $\frac{d_{\mathrm{CIII}}}{d_t} = k_{15}\ [P2.CIII] - k_{14}\ [CIII][P2]$

12. $\frac{d_{\mathrm{P2CIII}}}{d_t} = k_{14}\ [CIII][P2] - k_{15}\ [P2.CIII] - k_{16}\ [P2.CIII]$

Appendix E

Below, we present the reactions list of the convergence of the JAK-SAT and PKC pathway as applied in the study of Wang (2005). The system is defined via 56 substrates and 87 reactions. The description of each substrate can be found in the same study.

1. Receptor IFNAR 1 + TYK $\xrightarrow{k_1}$ Receptor TYK Complex
2. Receptor TYK Complex $\xrightarrow{k_2}$ Receptor IFNAR 1 + TYK
3. Receptor IFNAR 2 + JAK $\xrightarrow{k_3}$ Receptor JAK Complex
4. Receptor JAK Complex $\xrightarrow{k_4}$ Receptor IFNAR 2 + JAK
5. Receptor JAK Complex + Receptor TYK Complex + IFN_free $\xrightarrow{k_5}$ IFNAR dimer
6. IFNAR dimer $\xrightarrow{k_6}$ Receptor JAK Complex + Receptor TYK Complex + IFN_free
7. IFNAR dimer $\xrightarrow{k_7}$ Active Receptor Complex
8. STAT2c_IRF9 + Active Receptor Complex $\xrightarrow{k_8}$ Active Receptor Complex_STAT2c + IRF9c
9. STAT2c + Active Receptor Complex $\xrightarrow{k_9}$ Active Receptor Complex_STAT2c
10. Active Receptor Complex_STAT2c $\xrightarrow{k_{10}}$ STAT2c + Active Receptor Complex
11. STAT1c + Active Receptor Complex_STAT2c $\xrightarrow{k_{11}}$ Active Receptor Complex_STAT2c_STAT1c
12. Active Receptor Complex_STAT2c_STAT1c $\xrightarrow{k_{12}}$ STAT1c + Active Receptor Complex_STAT2c
13. Active Receptor Complex_STAT2c_STAT1c $\xrightarrow{k_{13}}$ Active Receptor Complex + STAT1c*_STAT2c*
14. IRF9c + STAT1c*_STAT2c* $\xrightarrow{k_{14}}$ ISGF-3c
15. ISGF-3c $\xrightarrow{k_{15}}$ IRF9c + STAT1c*_STAT2c*
16. ISGF-3c $\xrightarrow{k_{16}}$ ISGF-3n

17. ISGF-3n $\xrightarrow{k_{17}}$ ISGF-3c

18. STAT1c*_STAT2c* $\xrightarrow{k_{18}}$ STAT1n*_STAT2n*

19. STAT1n*_STAT2n* $\xrightarrow{k_{19}}$ STAT1c*_STAT2c*

20. STAT1n*_STAT2n* + IRF9n $\xrightarrow{k_{20}}$ ISGF-3n

21. ISGF-3n $\xrightarrow{k_{21}}$ STAT1n*_STAT2n* + IRF9n

22. ISGF-3n + Free transcription factor binding site (TFBS) $\xrightarrow{k_{22}}$ Occupied TFBS

23. Occupied TFBS $\xrightarrow{k_{23}}$ ISGF-3n + Free transcription factor binding site (TFBS)

24. $\emptyset \xrightarrow{k_{24}}$mRNAn

25. mRNAn$\xrightarrow{k_{25}}$ mRNAc

26. mRNAc$\xrightarrow{k_{26}}$ mRNAn

27. $\emptyset \xrightarrow{k_{27}}$ IRF9c

28. IRF9c $\xrightarrow{k_{28}}$ $\emptyset$

29. $\emptyset \xrightarrow{k_{29}}$ SOCS

30. SOCS $\xrightarrow{k_{30}}$ $\emptyset$

31. Active Receptor Complex $\xrightarrow{k_{31}}$ Receptor IFNAR1 +Receptor IFNAR2 + JAK + TYK

32. Active Receptor Complex $\xrightarrow{k_{32}}$ IFNAR dimer

33. IRF9n $\xrightarrow{k_{33}}$ $\emptyset$

34. STAT2c_IRF9 $\rightarrow\xrightarrow{k_{34}}$ STAT2c

35. STAT2n_IRF9 $\xrightarrow{k_{35}}$ STAT2n

36. ISGF-3c + Cytoplasmic phosphatase (CP) $\xrightarrow{k_{36}}$ ISGF-3c_CP

37. ISGF-3c_CP $\xrightarrow{k_{37}}$ ISGF-3c + Cytoplasmic phosphatase (CP)

38. ISGF-3c_CP $\xrightarrow{k_{38}}$ STAT1c +STAT2c + CP + IRF9c

39. STAT1c*_STAT2c* + CP $\xrightarrow{k_{39}}$ STAT1c*_STAT2c*_CP

40. STAT1c*_STAT2c*_CP $\xrightarrow{k_{40}}$ STAT1c*_STAT2c* + CP

41. STAT1c*_STAT2c*_CP $\xrightarrow{k_{41}}$ STAT1c + STAT2c + CP

42. STAT1n*_STAT2n* + Nuclear phosphatase (NP) $\xrightarrow{k_{42}}$ STAT1n*_STAT2n*_NP

43. STAT1n*_STAT2n*_NP $\xrightarrow{k_{43}}$ STAT1n*_STAT2n* + Nuclear phosphatase (NP)

44. STAT1n*_STAT2n*_NP $\xrightarrow{k_{44}}$ STAT1n + STAT2n + NP

45. ISGF-3n + NP $\xrightarrow{k_{45}}$ ISGF-3n_NP

46. ISGF-3n_NP $\xrightarrow{k_{46}}$ ISGF-3n + NP

47. ISGF-3n_NP $\xrightarrow{k_{47}}$ STAT1n + STAT2n + NP + IRF9n

48. Occupied TFBS + NP $\xrightarrow{k_{48}}$ Occupied TFBS_NP

49. Occupied TFBS_NP $\xrightarrow{k_{49}}$ Occupied TFBS + NP

50. Occupied TFBS_NP $\xrightarrow{k_{50}}$ STAT1n + STAT2n + Free TFBS +IRF9n

51. PIAS + ISGF-3n $\xrightarrow{k_{51}}$ PIAS_ISGF-3n

52. PIAS_ISGF-3n $\xrightarrow{k_{52}}$ PIAS + ISGF-3n

53. mRNAc$\xrightarrow{k_{53}}$ ∅

54. STAT1c $\xrightarrow{k_{54}}$ STAT1n

55. STAT1n $\xrightarrow{k_{55}}$ STAT1c

56. STAT2c $\xrightarrow{k_{56}}$ STAT2n

57. STAT2n $\xrightarrow{k_{57}}$ STAT2c

58. STAT2c + IRF9c $\xrightarrow{k_{58}}$ STAT2c_IRF9

59. STAT2c_IRF9 $\xrightarrow{k_{59}}$ STAT2c + IRF9c

60. STAT2n + IRF9n $\xrightarrow{k_{60}}$ STAT2n_IRF9

61. STAT2n_IRF9 $\xrightarrow{k_{61}}$ STAT2n + IRF9n

62. STAT2c_IRF9 $\xrightarrow{k_{62}}$ STAT2n_IRF9

63. STAT2n_IRF9 $\xrightarrow{k_{63}}$ STAT2c_IRF9

64. IRF9c $\xrightarrow{k_{64}}$ IRF9n

65. IRF9n $\xrightarrow{k_{65}}$ IRF9c

66. $\text{IFN\_influx} \xrightarrow{k_{66}} \text{IFN\_free}$

67. $\text{STAT1} + \text{GATA4} \xrightarrow{k_{67}} \text{AII}$

68. $\text{AII} + \text{AT1R} \xrightarrow{k_{68}} \text{ANF}$

69. $\text{PKC} - \text{cytosolic} + \text{AA} \xrightarrow{k_1} \text{PKC.AA}$

70. $\text{PKC} - \text{cytosolic} + \text{Ca}^{2+} \xrightarrow{k_2} \text{PKC.Ca}^{2+}$

71. $\text{PKC.Ca}^{2+} + \text{AA} \xrightarrow{k_3} \text{PKC.Ca}^{2+}.\text{AA}^{*}$

72. $\text{DAG} + \text{PKC.Ca}^{2+} \xrightarrow{k_4} \text{PKC.Ca}^{2+}.\text{DAG}$

73. $\text{PKC.DAG.AA} \xrightarrow{k_5} \text{PKC.DAG.AA}^{*}$

74. $\text{PKC} - \text{cytosolic} \xrightarrow{k_6} \text{PKC} - \text{basal}$

75. $\text{PKC.Ca}^{2+} \xrightarrow{k_7} \text{PKC.Ca}^{2+}.\text{memb}$

76. $\text{PKC.Ca}^{2+}.\text{DAG} \xrightarrow{k_8} \text{PKC.DAG.memb}$

77. $\text{PKC} - \text{cytosolic.DAG} \xrightarrow{k_9} \text{PKC.DAG}$

78. $\text{PKC.DAG} + \text{AA} \xrightarrow{k_{10}} \text{PKC.DAG.AA}$

In the list of the differential equations of this system as the mathematical description, we only present the newly added reactions to combine the JAK-STAT and PKC pathways. For the remaining differential equations, we use the same expressions given in the previous appendices about the associated pathways.

1. $\frac{d_{\text{GATA4}}}{d_t} = -k_{67}\,[STAT1][GATA4]$
2. $\frac{d_{\text{AII}}}{d_t} = k_{67}\,[STAT1][GATA4] - k_{68}\,[AII][AT1R]$
3. $\frac{d_{\text{AT1R}}}{d_t} = -k_{68}\,[AII][AT1R]$
4. $\frac{d_{\text{ANF}}}{d_t} = k_{68}\,[AII][AT1R]$