

## ABSTRACT

Title of dissertation:      IMPACT OF STOCHASTICITY ON  
                                  GENE REGULATION NETWORKS

Xin Fang  
Doctor of Philosophy, 2007

Dissertation directed by: Professor Evangelhos Zafiriou  
                                  Department of Chemical and Biomolecular  
                                  Engineering

We studied the impact of stochasticity on gene regulation networks, using the cell-to-cell communication mechanism in *Escherichia coli* as an example.

First we explored signal mediated positive autoregulation networks and their stochastic bistability, in the presence of which an initially homogeneous cell population would evolve into two distinct subpopulations. We proposed the simplification of the full network into one that can be theoretically studied. Simulation results indicate the simplifications retain the bistability and the distribution shapes so that the simplified network can be used to predict the bistable behavior of the full network. Moreover, it was shown that the bistability can be influenced by the signal molecule number, and that stochastic simulation is necessary for bistable systems. The self-promotion network for SdiA protein, with the autoinducer-2 (AI-2) signal molecule, was used as an example. The results further motivate the need for modeling of the AI-2 uptake mechanism.

We next explored cell age distribution in the case where the number of a key

protein for cell division has a stochastic bifurcation. With this bifurcation, the alive probability function (the probability that the cell has not divided) can be written in a double-exponential form. This analytical form allow the use of Laplace transform to calculate an analytical cell age distribution from the population balance model. The computation results indicate that if the key division protein number has a bifurcation, there is likely to be a significant fraction of first-generation cells in the cell population.

Finally, we developed deterministic and stochastic models for the regulation network of the AI-2 uptake in *Escherichia coli*. This network is regulated by a set of *lsr* genes, and we proposed that the LsrD protein needs to reach a threshold for uptake to take place. Based on the deterministic model, kinetic parameter values were estimated by fitting to experimental data from the literature. During the step-by-step fitting procedure, data for mutant cells and effective data for wild type cells were used to avoid the complexity of the full wild-type network. With the estimated parameters, the deterministic simulation results matched experimental data well, except for a steep change and spike. A stochastic model was also developed and simulation results showed a mild change and no spike for the population means. The difference between stochastic means and deterministic paths is due to the LsrD protein number threshold and indicates that stochastic simulation may be necessary for a monostable system if it has a threshold mechanism.

IMPACT OF STOCHASTICITY ON GENE REGULATION  
NETWORKS

by

Xin Fang

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2007

Advisory Committee:

Professor Evangelhos Zafiriou, Chair/Advisor

Professor William E. Bentley

Professor Panagiotis Dimitrakopoulos

Professor Raymond A. Adomaitis

Professor Sreeramamurthy Ankem (Dean's Representative)

© Copyright by  
Xin Fang  
2007

## Dedication

I would like to dedicate this work to my parents, Chunyin Fang and Guilan Diao, and my love, Wei Lu, without whom none of the following would have been possible.

## Acknowledgements

I would like to take this opportunity to gratefully acknowledge all the people who have made this thesis possible.

First of all, I want to thank my advisor, Dr. Evangelhos Zafiriou, for his invaluable time, support, and guidance during all the years of my graduate research. I also wish to thank Dr. William E. Bentley for his helpful discussions and suggestions.

I am thankful to Dr. Jun Li, who gave me great and kind help, suggestions and encouragement. I would also like thank the staff of the Department of Chemical and Biomolecular Engineering for their kind assistance.

This research has been supported by National Science Foundation grant BES-0222687. We are grateful to the Performability Engineering Research Group (PERFORM) at the University of Illinois at Urbana-Champaign for the use of the software package Mobius(<http://www.mobius.uiuc.edu>).

# Table of Contents

List of Tables	vi
List of Figures	vii
1 General Introduction	1
1.1 Stochastic Modeling of Gene Regulation Networks . . . . .	1
1.1.1 Stochasticity in Gene Regulation Networks . . . . .	1
1.1.2 Stochastic Models . . . . .	2
1.1.3 Stochastic Petri Nets . . . . .	5
1.1.4 Summary of Models . . . . .	8
1.2 Quorum Sensing as Case Study . . . . .	9
1.2.1 What is Quorum Sensing . . . . .	9
1.2.2 Quorum Sensing in <i>Escherichia coli</i> . . . . .	11
1.3 Outline of the Thesis . . . . .	15
2 Stochastic Modeling of Gene Positive Autoregulation Networks Involving Signal Molecules	18
2.1 Introduction . . . . .	18
2.2 Network Models . . . . .	20
2.2.1 Full Model . . . . .	21
2.2.2 Simplified Model . . . . .	24
2.3 Analysis of Simplified Model . . . . .	27
2.3.1 Steady-State Distribution from Mathematical formula . . . . .	27
2.3.2 Stochastic Simulation . . . . .	29
2.3.3 Deterministic Simulation . . . . .	32
2.4 Analysis of Full Model . . . . .	34
2.4.1 Stochastic Simulation . . . . .	34
2.4.2 Deterministic Steady-State Analysis . . . . .	37
2.5 Discussion . . . . .	41
3 Cell Age Distribution in Relation to Stochastic Bifurcation of Key Protein Number	45
3.1 Introduction . . . . .	45
3.2 Methods for Obtaining the Alive Probability Function . . . . .	48
3.2.1 Alive Probability Function . . . . .	48
3.2.2 Computation from the Fokker-Planck Equation . . . . .	49
3.2.3 Computation from Stochastic Petri Net Simulation . . . . .	51
3.3 Results for Alive Probability Function . . . . .	53
3.3.1 Results from Fokker-Planck Equation . . . . .	53
3.3.2 Results from Stochastic Petri Nets . . . . .	61
3.3.3 Explanation of the Double-exponential Form in Relation to Bifurcation of Key Protein Number . . . . .	69
3.4 Methods for Obtaining Age Distribution . . . . .	72

3.4.1	Generation Time Distribution . . . . .	72
3.4.2	Total-Based Method to Obtain Age Distribution . . . . .	73
3.4.3	Generation-Based Method to Obtain Age Distribution . . . . .	77
3.5	Results for Age Distribution . . . . .	80
3.5.1	Case without Time delay . . . . .	80
3.5.2	Case with Time Delay . . . . .	83
3.6	Discussion . . . . .	87
4	Deterministic and Stochastic Modeling of Autoinducer-2 Uptake Regulation Network in <i>Escherichia coli</i> . . . . .	95
4.1	Introduction . . . . .	95
4.2	Models . . . . .	99
4.2.1	Kinetic Network . . . . .	99
4.2.2	Deterministic Model . . . . .	101
4.2.3	Optimization Model for Parameter Estimation . . . . .	103
4.2.4	Stochastic Model . . . . .	106
4.3	Results . . . . .	108
4.3.1	Parameter Estimation . . . . .	108
4.3.2	Results of Stochastic Simulation . . . . .	124
4.4	Discussion . . . . .	129
5	Conclusion . . . . .	131
5.1	Protein Number Distribution . . . . .	131
5.2	Age Distribution . . . . .	133
5.3	AI-2 uptake . . . . .	134
5.4	Suggestions for Future work . . . . .	135
	Bibliography . . . . .	137

## List of Tables

1.1	Summary of Stochastic Models. . . . .	9
2.1	Parameters of full model. . . . .	22
2.2	Parameters of simplified model. . . . .	27
2.3	Values for Rescaled Parameters. . . . .	28
2.4	Steady states when $[AI2_{total}] = 600$ . . . . .	41
3.1	Parameter Values of Double-exponential Function. . . . .	70
4.1	ODEs of AI-2 uptake network in <i>E. coli</i> . . . . .	102
4.2	Parameter values adjusted from literature. . . . .	104
4.3	Fitting results for mutant kinetic networks. . . . .	113

## List of Figures

1.1	A simple example for SPN. . . . .	7
1.2	Quorum sensing system in <i>E.coli</i> . . . . .	12
2.1	Full model with signal-mediated positive autoregulation. . . . .	21
2.2	Simplified model with positive autoregulation. . . . .	24
2.3	Distribution of protein monomer (free SdiA) number for the simplified model . . . . .	30
2.4	SPN for simplified model. . . . .	31
2.5	protein monomer (SdiA) number versus time for the simplified model	33
2.6	SPN for full model. . . . .	35
2.7	Distribution of protein monomer (free SdiA) number for the full model.	36
2.8	Total signal molecule number ( $[AI2_{total}]$ ) vs. $z([Dimer])$ . . . . .	40
3.1	General stochastic Petri net to obtain alive probability function. . . . .	52
3.2	Self-promotion network. . . . .	53
3.3	Protein number distribution for the case with bifurcation with distant peaks . . . . .	56
3.4	Alive probability function from Fokker-Planck equation for the case with bifurcation with distant peaks . . . . .	56
3.5	Protein number distribution for the case with bifurcation with partially overlapped peaks . . . . .	57
3.6	Alive probability function from Fokker-Planck equation for the case with bifurcation with partially overlapped peaks . . . . .	58
3.7	Protein number distribution for the case without bifurcation and threshold point is in the peak . . . . .	59
3.8	Alive probability function from Fokker-Planck equation for the case without bifurcation and threshold point is in the peak . . . . .	59

3.9	Protein number distribution for the case without bifurcation and start point is in the peak . . . . .	60
3.10	Alive probability function from Fokker-Planck equation for the case without bifurcation and start point is in the peak . . . . .	61
3.11	Protein number distribution for the case without bifurcation but with double-exponential function . . . . .	62
3.12	Alive probability function from Fokker-Planck equation for the case without bifurcation but with double-exponential function . . . . .	62
3.13	SPN to obtain alive probability function of self-promotion network. . . . .	64
3.14	Protein number distribution of self-promotion network from Stochastic Petri nets (SPN) . . . . .	65
3.15	Alive probability function of self-promotion network from Stochastic Petri nets (SPN) . . . . .	65
3.16	Signal-mediated self-promotion network. . . . .	66
3.17	SPN to obtain alive probability function of signal-mediated self-promotion network. . . . .	67
3.18	Protein number distribution of signal-mediated self-promotion network from Stochastic Petri nets (SPN). . . . .	68
3.19	Alive probability function of signal-mediated self-promotion network from Stochastic Petri nets (SPN) . . . . .	68
3.20	$c$ vs $x_0$ curve. . . . .	71
3.21	Total cell number when alive probability functions $F(t)$ are in two different forms. . . . .	84
3.22	The shape of total cell number for the case without time delay when $\mu = 1, c = \frac{3}{4}$ . . . . .	84
3.23	The shape of total cell number for the case without time delay when $\mu = 1, c = \frac{1}{2}$ . . . . .	85
3.24	The shape of total cell number for the case without time delay when $\mu = 1, c = \frac{1}{4}$ . . . . .	85
3.25	The functions $Q_1(t)$ for the case with time delay. . . . .	88

3.26	The functions $Q_2(t)$ for the case with time delay. . . . .	88
3.27	The functions $Q_3(t)$ for the case with time delay. . . . .	89
3.28	The functions $Q_4(t)$ for the case with time delay. . . . .	89
3.29	The functions $Q_5(t)$ for the case with time delay. . . . .	90
3.30	The functions $Q_6(t)$ for the case with time delay. . . . .	90
3.31	The functions $Q_7(t)$ for the case with time delay. . . . .	91
3.32	The functions $Q_8(t)$ for the case with time delay. . . . .	91
3.33	Total cell number for the case with time delay. . . . .	92
4.1	The regulation networks of AI-2 uptake in <i>Salmonella typhimurium</i> .	96
4.2	The regulation networks of AI-2 uptake in <i>Escherichia coli</i> . . . . .	97
4.3	AI-2 uptake network of wild type <i>E.coli</i> . . . . .	100
4.4	Stochastic Petri net for AI-2 uptake network in <i>E.coli</i> . . . . .	106
4.5	Fitting results for cell growth data . . . . .	108
4.6	Kinetic network of mutant $\Delta lsrR$ to fit $\beta$ -galactosidase units (full transcription). . . . .	110
4.7	Kinetic network of mutant $\Delta lsrACDBFG$ to fit external AI-2 activity (alternative uptake). . . . .	110
4.8	Kinetic network of mutant $\Delta lsrR$ to fit external AI-2 activity (both uptake). . . . .	110
4.9	Kinetic network of mutant $\Delta lsrK$ to fit $\beta$ -galactosidase units (repression). . . . .	111
4.10	Kinetic network of mutant $\Delta lsrB$ to fit $\beta$ -galactosidase units (de-repression). . . . .	112
4.11	Deterministic simulation of mutant network corresponding to the step in Table 4.3 for “no uptake” . . . . .	114
4.12	Deterministic simulation of mutant network corresponding to the step in Table 4.3 for “full transcription1” . . . . .	114

4.13	Deterministic simulation of mutant network corresponding to the step in Table 4.3 for “full transcription2” . . . . .	115
4.14	Deterministic simulation of mutant network corresponding to the step in Table 4.3 for “alternative uptake” . . . . .	115
4.15	Deterministic simulation of mutant network corresponding to the step in Table 4.3 for “both uptake” . . . . .	116
4.16	Deterministic simulation of mutant network corresponding to the step in Table 4.3 for “repression” . . . . .	116
4.17	Deterministic simulation of mutant network corresponding to the step in Table 4.3 for “de-repression” . . . . .	117
4.18	Analysis of lower bound of LsrD threshold . . . . .	118
4.19	Analysis of lower bound of LsrD threshold . . . . .	118
4.20	Kinetic network with effective transcription level. . . . .	120
4.21	Effective transcription level for obtaining $T_D$ . . . . .	120
4.22	Fitting result for obtaining $T_D$ from the network with effective transcription level: $T_D = 297$ . . . . .	121
4.23	Kinetic network with effective AI-2 uptake rate. . . . .	122
4.24	Effective AI-2 uptaken amount for obtaining $k_f$ , $k_b$ and $k_{dAI2}$ . . . . .	123
4.25	Effective AI-2 uptake rate for obtaining $k_f$ , $k_b$ and $k_{dAI2}$ . . . . .	123
4.26	Fitting result for obtaining $k_f$ , $k_b$ and $k_{dAI2}$ from the network with effective AI-2 uptake rate: $k_f = 1.070 \times 10^{-3}$ , $k_b = 5.954 \times 10^{-5}$ and $k_{dAI2} = 1.077 \times 10^{-5}$ . . . . .	124
4.27	Deterministic simulation results of external AI-2 activity of wild type <i>E.coli</i> . . . . .	125
4.28	Deterministic simulation results of $\beta$ -galactosidase units of wild type <i>E.coli</i> . . . . .	125
4.29	Results of external AI-2 activity in the whole time range from deterministic simulation and stochastic simulation. . . . .	127
4.30	Results of $\beta$ -galactosidase units in the whole time range from deterministic simulation and stochastic simulation. . . . .	127

- 4.31 Results of external AI-2 activity in time range between 390 and 490 minutes from deterministic simulation and stochastic simulation. . . . 128
- 4.32 Results of  $\beta$ -galactosidase units in time range between 390 and 490 minutes from deterministic simulation and stochastic simulation. . . . 128

## Chapter 1

### General Introduction

#### 1.1 Stochastic Modeling of Gene Regulation Networks

##### 1.1.1 Stochasticity in Gene Regulation Networks

A gene regulation network is a network with interactions between a set of genes and transcription factors. Transcription factors are proteins that can regulate transcription rates. Transcription is initiated by RNA polymerase (RNAP) binding to the promoter, the regulation region preceding genes on the DNA sequence. When a transcription factor binds the corresponding operator, the binding of RNAP on the promoter may become easier or more difficult, and thus the transcription rate may be increased or decreased. If it is increased, the transcription factor functions as an activator; while the factor acts as a repressor if the transcription rate is decreased. Sometimes signal molecules, small molecules which usually carry environmental information, are necessary for the effectiveness of a transcription factor. Transcription factors are also expressed from their genes and those genes can be affected by other transcription factors or their own gene products. Therefore the interactions between transcription factors and genes can form a network [1, 2]. This network is gene regulation network.

Gene regulation networks have been studied by many researchers and can be

simulated with deterministic ordinary differential equations (ODEs) [3]. However, activities of gene regulation networks are basically stochastic. In a cell, almost all activities involve chemical reactions between molecules, and the chemical reactions are subject to thermal fluctuations. In regular chemical reactions, the molecule numbers are usually around the scale of Avogadro number. The molecule numbers are so large that the effects of fluctuations are not apparent. However, in gene regulation networks, the numbers of related molecules may be just around several hundreds or less. Thus, the effects of fluctuations may be significant. Therefore we should view the activities of gene regulation networks, like transcription, translation, protein binding and other molecular reactions, as stochastic processes [4, 5].

The stochasticity in gene regulation networks may cause behavior that cannot be explained with deterministic models. For example, in a stochastic cell population the experimental data for single cells may differ significantly from each other, although the data measuring population are regular [6]. Also an initially homogeneous stochastic cell population can develop into two distinct subpopulations [7, 8]. Therefore, often stochastic models are necessary to simulate the behavior of gene regulation networks.

### 1.1.2 Stochastic Models

Before the introduction of stochastic models, let us briefly review deterministic models, because stochastic models can be developed from them. A deterministic model describes how concentrations of the various species change with time and is

composed of a set of ODEs:

$$\frac{dx(t)}{dt} = A(x) \quad (1.1)$$

where  $t$  is time. The vector  $x$  represents the concentrations or molecule numbers of the species. In a cell, the species can be DNA, mRNA, protein and other molecules; and in a biochemical process they can be biomass and substrates. The functions  $A(x)$  are the dynamics from rate laws of biochemical reactions. They can be written in the form of first or higher order chemical reaction kinetics, and can also be written as Michaelis-Menten kinetics. The deterministic model may be very detailed and complex and contain many parameters. If all the parameters and initial conditions are known, quantitatively precise prediction can be made for the future evolution of the biochemical system. Some models of this kind can accurately simulate a real system (for example a cell), and match the experimental data well [9].

The Langevin equation considers molecular fluctuations by adding a noise term to the equation (1.1) as follows [10]:

$$\frac{dx(t)}{dt} = A(x) + \sqrt{B(x)}\xi(t) \quad (1.2)$$

where  $A(x)$  and  $B(x)$  are the functions of drift and diffusion coefficients.  $\xi(t)$  is the noise term, which ideally represents white noise. White noise has infinite variance and thus does not really exist. However some real systems can be approximated by this ideal form. The solution of Langevin equation is obtained by solving the stochastic differential equations above. From the equation above we could see that the variables  $x(t)$  must be continuous variables. Each time we solve the same Langevin equation, we may have a different result. To know the probability distri-

bution at a specific time, we need to solve the Langevin equation many times and then statistically average the results.

The Fokker-Planck equation analytically describes the evolution of the probability density function, which avoids the time-consuming repetition of solving the Langevin equation. Fokker-Planck equation is a set of partial differential equations (PDEs) [10]:

$$\frac{dp(x, t)}{dt} = -\nabla \cdot [A(x)p(x, t)] + \frac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} [B(x)p(x, t)] \quad (1.3)$$

Here  $p(x, t)$  denotes the probability density function of  $x$  at time  $t$ . The value for  $p(x, t)\partial x$  is the probability that the values for stochastic variables reside between  $x$  and  $x + \partial x$  at time  $t$ . When Fokker-Planck and Langevin equations describe the same system, they should share the same functions of  $A(x)$  and  $B(x)$ . In Fokker-Planck equation, the stochastic variable  $x$  is continuous and the probability density function  $p(x, t)$  can be obtained by solving the Fokker-Planck equation above.

To describe discrete biochemical networks, the master equation can be obtained. Both the Langevin and Fokker-Planck equations are used to describe continuous systems, thus the stochastic variables in those equations are continuous. However, when we consider activities in a cell, for example gene expression, the numbers of DNA, mRNA and proteins are often very limited. In that case it is not appropriate to consider the numbers as continuous. They should be considered as discrete variables instead. The master equation is developed to calculate the evolution of the probability density function of discrete stochastic variables. The general

form of the master equation can be described as [11]:

$$\frac{dP_k}{dt} = \sum_l T_{kl} P_l \quad (1.4)$$

where  $P_k$  is the probability of state  $k$ ;  $T_{kl}$  is the transition rate constant from state  $l$  to state  $k$ . To illustrate the master equation more clearly, an example is raised [5].

Suppose there is a protein with two states A or B.  $k_1$  and  $k_2$  represent the likelihood of A-to-B and B-to-A transitions. Then the master equation is:

$$\frac{dp(n_a, n_b; t)}{dt} = -(k_1 n_a + k_2 n_b) p(n_a, n_b; t) + k_1 (n_a + 1) p(n_a + 1, n_b - 1; t) + k_2 (n_b + 1) p(n_a - 1, n_b + 1; t) \quad (1.5)$$

In the equation,  $p(n_a, n_b; t)$  is the probability at time  $t$  that  $n_a$  proteins are in state A and  $n_b$  proteins are in state B. Usually, master equations are too difficult to solve analytically, so they are converted to Fokker-Planck equation with some assumptions [12], or to the Langevin type equation [13], or solved with Monte Carlo simulation [14].

### 1.1.3 Stochastic Petri Nets

The methods we introduced before are not very appropriate to describe a complex stochastic and discrete system. The Langevin and Fokker-Planck equations are continuous models. Master equations are discrete models but usually too difficult to solve. An efficient representation and simulation approach is based on Stochastic Petri nets (SPN), which is a mathematical formalism that can realize a simulation of stochastic and discrete events. Hence it is a good tool for the stochastic simulation of molecular biological systems [15]. SPN is based on Petri nets, a graphics-oriented

formalism to analyze discrete events [16].

The structure of an SPN is composed of tokens, transitions and input and output gates. A simple SPN model is shown in Fig. 1.1. The tokens (circles) represent the chemical species such as DNA, mRNA, proteins and other molecules. Each token has a number, which represents the concentration or molecular number of the species. The transitions (vertical lines) represent the chemical reactions, with arrows pointing away from reactant and into products. Here we can consider cell activities like gene expression, molecular binding and unbinding as well as chemical reactions. The input and output gates (triangles) indicate the precondition and effect of the transition firing. If there is no precondition for transition firing and the effect of transition firing on the reactant token is just the token number minus one, the input gate can be canceled. If the effect of transition firing on the product token is just the token number plus one, the output gate can be canceled. The Mobius software is designed to simulate the SPN. This software is available from Performability Engineering Research Group (PERFORM) at the University of Illinois at Urbana-Champaign [17].

In SPN the transition fires in a stochastic manner. The time when the transition fires is determined by a distribution function of  $f_{fire}(t)$ , which is a negative exponential function [17]:

$$f_{fire}(t) = \lambda \exp(-\lambda t) \tag{1.6}$$

where  $\lambda$  is the rate parameter. When the transition represents a chemical reaction,

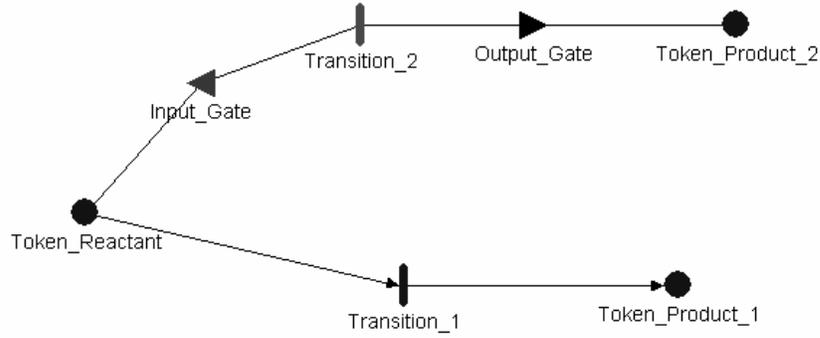


Figure 1.1: A simple example for SPN.

$\lambda$  can be written in the terms of kinetics:

$$\lambda = \beta n \quad (1.7)$$

$$\lambda = \beta n(n - 1) \quad (1.8)$$

$$\lambda = \beta n_A n_B \quad (1.9)$$

where  $n$  is molecule number. Eq. (1.7) represents a reaction of one molecule, Eq. (1.8) represents a reaction between two molecules of the same kind, Eq. 1.9 represents a reaction between two different kinds of molecules. Here  $\beta$  is the stochastic rate constant. The relation between  $\beta$  and the deterministic rate constant  $k$  depends on the situation. When  $k$  is in terms of molar concentration and the transition is a first-order reaction,  $\beta$  is equal to  $k$ . When  $k$  is still in terms of molar concentration and the reaction is second-order, the relation between stochastic and deterministic rates is [18]:

$$\beta = \frac{k}{VN_A} \quad (1.10)$$

where  $V$  is the cell volume and  $N_A$  is Avogadro's number. When  $k$  is in terms of molecule number, for both the first-order and second-order reactions, the stochastic rate constant  $\beta$  has the same value and unit as the deterministic rate constant  $k$ .

#### 1.1.4 Summary of Models

The stochastic models introduced above can be classified into four categories as shown in Table 1.1. Master equations and stochastic Petri nets are discrete models, in which the variable values are in terms of actual integer molecule numbers. Langevin and Fokker-Planck equations are continuous models, meaning that molecule numbers are considered as continuous variables. Discrete models are more accurate than continuous models, but are also more difficult to deal with. In similar manner, master equations and Fokker-Planck equations are probability distribution models. The solutions from these models are the probability distribution of molecule numbers. Langevin equations and stochastic Petri nets are Monte Carlo models. To obtain probability distribution, stochastic simulations need to be repeated many times. The probability distribution of relatively simple systems can be quickly obtained with probability distribution models. However the probability distribution models describing very complex models are difficult to solve, and sometimes even to be written in the explicit forms. Monte Carlo models will be applied in these systems.

The stochastic models are also interchangeable. For continuous models, Langevin and Fokker-Planck equations can be transformed to each other because they have

	Monte Carlo Model	Probability Distribution Model
Continuous Model	Langevin Equation	Fokker-Planck Equation
Discrete Model	Stochastic Petri Nets	Master Equation

Table 1.1: Summary of Stochastic Models.

the same functions  $A(x)$  and  $B(x)$  when representing the same system [10]. For discrete models, master equations for molecule interactions can be obtained by writing the Kolmogorov equations for stochastic Petri nets [15]. Discrete models can also be changed into continuous models when the molecule numbers are very large. It is possible to obtain Langevin equations and Fokker-Planck equations from master equations with large molecule numbers [13]. If there are some steps which are much faster than others, quasi-equilibrium can be assumed. Then the Fokker-Planck equation derived from master equations can have low dimension and be solvable [12].

## 1.2 Quorum Sensing as Case Study

### 1.2.1 What is Quorum Sensing

In this study we choose the quorum sensing system in *Escherichia coli* as an important application. Quorum sensing is a fairly innovative concept which overthrows our notion that bacterial cells are isolated each other. On the contrary, quorum sensing describes how cells communicate with each other and form a cooperating society. Quorum sensing has industrial value because it can influence intracellular gene expression and affect cell growth. Then in biochemical processes, quorum sensing can

influence the productivity of products. There are still many open questions in this field. How cells signal each other still has many unclear points, which make quorum sensing a very active topic and attract many researchers' attention.

Quorum sensing is a phenomenon which relates the cell activities to the density of cell population. The most typical quorum sensing system is the bioluminescence system in *Vibrio fischeri*. When this kind of bacteria accumulate together and the density of bacterial population becomes higher than a specific threshold, light is emitted from the bacteria [19]. Some eukaryotic hosts provide nutrient-rich environment in which *V. fischeri* can grow to high density, and then achieve their own purpose with light from the parasites. The reason we use the term of quorum sensing here, is that the bioluminescence activity depends on whether the cell density reaches the threshold. Quorum sensing indicates that cells are not isolated individuals, but able to perform cooperatively [20].

Quorum sensing is realized by a system that enables cells to communicate with each other through signal molecules. A quorum sensing system is composed of three steps: First signal molecules are synthesized and then sent out of the cell. When the concentration of the signal molecules reaches a threshold, cells detect the molecules and uptake them through a mechanism. Finally proteins bind the molecules and regulate the expression of target genes, such as letting *V. fischeri* emit light. [21, 22]

There are three different kinds of quorum sensing systems. In gram-negative bacteria, the signal molecules, autoinducers, are acylated homoserine lactone (AHL). The autoinducers are synthesized by LuxI-type enzymes. The LuxR-type protein binds the autoinducers and controls the transcription of target genes [21]. In gram-

positive bacteria, the signal molecules are oligopeptides. The oligopeptides are synthesized by processing and secreting precursor peptides [23]. Signal molecules are transduced by a two-component system to regulate target gene transcription [24]. There is also a kind of hybrid quorum sensing system. The signal molecules are AHL autoinducers, which are synthesized by a system like in gram-negative bacteria. But a two-component system, similar to that in gram-positive bacteria, make signal molecules able to regulate gene expression [22].

Signal molecules for interspecies bacterial communication can be involved in quorum sensing. In the quorum sensing system of *Vibrio harveyi*, there are two kinds of signal molecules, i.e., autoinducers [25]. Only when these two kinds of autoinducers, AI-1 and AI-2, are present together, the target gene can be expressed after a series of regulation [26]. AI-1 is a normal AHL and is only used in introspecies communication [27, 28]. However, AI-2 is suggested to be used in interspecies communication, because its synthase, LuxS, exists in many kinds of bacteria, and AI-2 can be detected by many different bacteria [22, 28, 29]. Therefore bacterial activity is possible to be affected by other species through quorum sensing.

### 1.2.2 Quorum Sensing in *Escherichia coli*

Quorum sensing in *E.coli* can be described in Fig. 1.2. AI-2, an important type of signal molecule in *E.coli*, is synthesized from S-adenosylmethionine (SAM). As a methyl donor, SAM can be converted to S-adenosylhomocysteine (SAH) with methyl transferase enzymes. SAH is toxic to the cell so it is degraded into S-

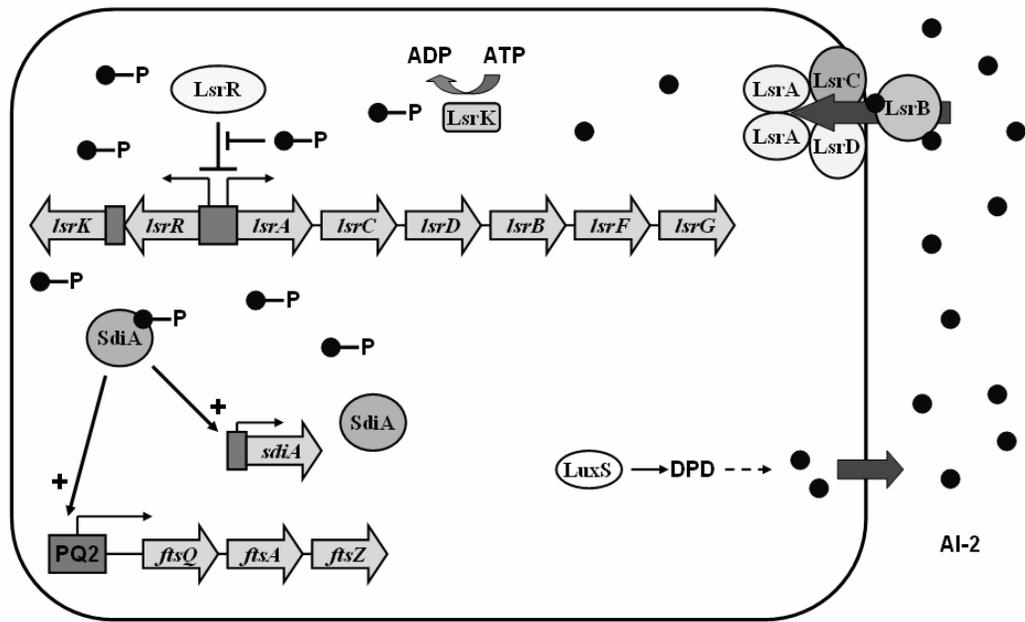


Figure 1.2: Quorum sensing system in *E. coli*. Adapted and revised from [36].

ribosylhomocysteine (SRH) with pfs enzyme. Then LuxS can convert SRH into AI-2 and other materials [30]. The regulation of AI-2 production is still not very clear. It was found that AI-2 is produced during the exponential phase and is eliminated from culture during the stationary phase [31, 32], and that the AI-2 production prefers low pH, high osmolarity and presence of glucose [31, 33, 34, 35]. However, the details about how these factors affect AI-2 production are still unknown. Recent study discovered that glucose influence AI-2 production and uptake through cyclic AMP and cAMP receptor protein [32].

The uptake of AI-2 into *E.coli* cells at least involves two pathways [32, 36]. One pathway is the uptake mediated by *lsr* gene products. In this pathway, AI-2 is first transported into cells by the transporter apparatus encoded by *lsrACDB*. Then AI-2 is phosphorylated by the kinase LsrK. LsrR can repress the transcription of *lsr* genes but the LsrR repression can be derepressed by the phospho-AI-2. LsrR and phospho-AI-2 can regulate the *lsr*-mediated uptake. Finally, the protein LsrF and LsrG involve the degradation of AI-2. The detail for another pathway is still unknown.

SdiA, the only LuxR-type receptor in *E. coli*, is involved in three independent gene regulation cases [37]. The first regulation is the positive effect of SdiA on cell division. Overexpression of SdiA can directly or indirectly activate the PQ2 promoter, which is located in the upstream of the cell division gene *ftsQAZ*. The products from *ftsQAZ* are necessary for cell division. FtsZ is used during the Z-ring formation and FtsA is required for the formation of cross wall [38]. The second regulation is that SdiA can result in resistance to mitomycin C (MMC) [39]. The

third regulation is that SdiA also has an effect of resistance to quinolones [40]. These regulations suggest the potential role of quorum sensing in *E. coli*.

AI-2 can be involved in SdiA-regulated gene expression. First, it was reported that the PQ2 promoter is activated in a high cell density environment [41], which suggests a relation between the SdiA-activated gene expression and quorum sensing. Second, the molecular structures are similar to the LuxI-LuxR quorum sensing system in gram-negative bacteria. LuxS, the final synthase of AI-2, is a LuxI-type protein in that it is responsible for synthesizing an autoinducer [30]. SdiA is a LuxR-type protein [37]. It was also found that the N-terminal domain of SdiA can bind to an autoinducer [42]. Third, experiment results show that gene expression from PQ2 can be enhanced in the condition medium with AI-2, compared with that in the condition medium without AI-2. It was recently discovered that in chemostat culture the signaling of AI-2 to SdiA is independently of cell density, glucose concentration and growth rate [34]. All the facts above suggest that AI-2 molecules regulate gene expression through SdiA protein.

AI-2 and SdiA may regulate the production of SdiA itself. In *V. fischeri*, the R protein LuxR is involved in the down-regulation of its own expression by the extracellular factor [43, 44], suggesting that LuxR-type protein can regulate its own expression with signal molecules. As for SdiA, it was reported that an extracellular factor, the derivative of homoserine lactone (HSL), can down-regulate expression of *sdiA*. But it seems that the down-regulation by this signal molecule is not mediated by SdiA itself [45]. Appearance of AI-2 can increase the expression of *sdiA* [46]. Although it is still unknown whether SdiA itself is related to this regulation, self-

regulation of LuxR suggests it is very possible that with the extracellular signal molecule AI-2, SdiA up-regulates its own expression.

### 1.3 Outline of the Thesis

The research in this thesis studies the impact of stochasticity on gene regulation networks, with networks involved in the quorum sensing mechanism of *E.coli* used as a case study. Chopp [47] and Viretta and Fussenegger [48] used deterministic models to study the quorum sensing of *Pseudomonas aeruginosa*. Work on stochastic modeling of quorum sensing is still rare and in previous work the value for stochastic models is not apparent because stochastic means follow deterministic results [49, 50, 51]. Therefore further work on stochastic modeling is still needed.

In Chap. 2 we study positive autoregulation in gene regulation networks, which has been shown in the past to exhibit stochastic behavior, including stochastic bistability, in which an initially uniform cell population develops into two distinct subpopulations. However, positive autoregulation is often mediated by signal molecules, which have not been considered in prior stochastic analysis of these networks. Here we propose both a full model of such a network that includes a signal molecule, and a simplified model in which the signal molecules have been eliminated through the use of two simplifications. The simplified model is amenable to direct mathematical analysis that shows that stochastic bistability is possible. We use stochastic Petri networks for simulating both types of models. The simulation results show that (i) the stochastic behavior of the two models is similar, and (ii) that the analytical

steady state distribution of the simplified model matches well the transient results at times equal to that of a cell generation. A discussion of the simplifications we used in the context of the results, indicates the importance of the signal molecule number as a factor determining the presence of bistability. This is further supported from a deterministic steady state analysis of the full model that is shown to be a useful indicator of potential stochastic bistability. We use the regulation of SdiA in *Escherichia coli* as an example, due to the importance of this protein and of the signal molecule, AI-2, that is involved. However, the use of kinetic parameter values representing typical cellular activities make the conclusions applicable to other signal-mediated positive autoregulation networks as well.

In Chap. 3 cell age distribution is mathematically explored in the case that the number of key proteins for cell division exhibit stochastic bifurcation. It is proposed that in this case the alive probability functions from both Fokker-Planck equations and Stochastic Petri nets can be written in the form of a double-exponential function. The double-exponential function can be used in a population balance model, from which analytical forms of age distribution for this cell population can be obtained from the population balance model by using Laplace transformation. The calculated results show that if the number of the key proteins for cell division has a bifurcation, even after a relatively long time there would still be a significant fraction of first-generation cells in the population.

In Chap. 4 we focus on the uptake of autoinducer-2 (AI-2) molecules into *Escherichia coli*, which is regulated by a gene regulation network. Deterministic and stochastic models for this network are developed. Based on the deterministic

model, parameters in the network are estimated by fitting to experimental data from the literature. The full network in wild-type cells is very complex, and thus the network is simplified by considering mutant cells, in which the networks are relatively simple, and by using effective transcription levels and AI-2 uptake rates. With these simplified networks, parameter values for the regulation network of AI-2 uptake are obtained step by step from experimental data. These parameter values are used for the model for the full network and the simulation results match well with experimental data of wild-type cells, except for a steep drop of external AI-2 activity and spikes in transcription level, which one does not expect to see in experimental data. A stochastic model is also developed, and the simulation results indicate that when there is threshold mechanism inside the cells, deterministic and stochastic simulations show significant differences. Deterministic simulation predicts a step change at the time a threshold is reached, while in stochastic simulation each cell reaches threshold at different time so that the overall change is mild. Hence the stochastic model better represents the type of measurements we can expect from a cell population.

Chap. 5 provides the conclusion to this thesis.

## Chapter 2

### Stochastic Modeling of Gene Positive Autoregulation Networks

#### Involving Signal Molecules

##### 2.1 Introduction

Gene expression is subject to intrinsic noise [52], which may come from thermal fluctuations of molecular events that constitute the process of gene expression [5]. When the numbers of some molecules in a cell are relatively low, the effect of fluctuations can be apparent and then even cloned cells in the same environments can evolve into different phenotypes [4]. In such situations, gene regulation networks should be viewed as stochastic processes. Several gene regulation networks have been stochastically modeled. MacAdams and Arkin [7] performed stochastic simulation for a simple regulation link of two genes. The expression of pyelonephritis-associated pili in uropathogenic *E. coli* has been modeled as a Markov chain [53]. A stochastic model for the quorum sensing mechanism in *E. coli* was developed and used to obtain new insights on synthesis pathways of autoinducer-2 (AI-2) [51].

Gene regulation networks with positive autoregulation are strong candidates for consideration of stochastic behavior because positive autoregulation may cause bistability and let an initially homogeneous cell population develop into two distinct subpopulations. Arkin *et al.* [8] developed a stochastic model for the developmental

pathway in phage  $\lambda$ -infected *E.coli* and simulation results predicted a bifurcation in the cell population, which was experimentally verified. In this model, although the  $\lambda$  switch is essential for the selection for lysis/lysogeny pathways, the positive autoregulation of CI expression is necessary for the bifurcation. Positive feedback modules are present in *E.coli* cells and experimental results indicate bistability of protein amounts, consistent with predictions from a stochastic model [54]. In another experiment, stochastic bifurcation can even be directly observed under the microscope [55]. Positive autoregulation networks have also been mathematically analyzed to obtain insight in bistability. Ferrell [56] discussed how cellular systems with positive feedback loops can exhibit bistability and reviewed some typical bistable systems. Kepler and Elston [12] analytically obtained the bistable probability distribution of protein number in a simple self-promotion network.

However, positive autoregulation in gene regulation networks is often mediated through signal molecules. Let us consider two examples in quorum sensing systems, a cell-to-cell communication mechanism, in which extracellular factors function as signal molecules. In *Vibrio fischeri*, when DNA sequences in operon<sub>R</sub> are deleted or autoinducers (AI) are at low levels, LsrR protein can have a positive regulation on *lsrR* gene in the presence of AI [44]. In *E.coli*, the appearance of another kind of extracellular factor, autoinducer-2 (AI-2), can increase the expression of *sdiA* gene [46]. Although it is still uncertain that the SdiA protein is related to this regulation, there are some supporting facts. First, SdiA has a LuxR-type sequence [21, 37] and its N-terminal domain can bind to a factor [42], which suggests that SdiA may regulate genes by binding some factor. Second, AI-2 can enhance some

SdiA-regulated gene expression [34], which suggests that AI-2 may be the factor. Therefore, it is likely that SdiA protein upregulates its own expression by binding AI-2 molecules [49].

Work on the stochastic modeling of such signal-mediated networks has been limited. Through the study of the mathematical relation between the signal-mediated networks and the pure positive autoregulation networks that have been more extensively studied, conclusions from the latter networks, such as the property of bistability, may be carried to the signal-mediated networks. Here we present one network, in which positive autoregulation is mediated through signal molecules, as the full model, and another network, in which there is only pure positive autoregulation, as the simplified model. With some assumptions the simplified model can be obtained from the full model. Mathematical analysis and simulations for both models are performed stochastically and deterministically, so that the bistable properties of the two models can be studied in relation to each other, in order to explore under what circumstances the conclusions from the simplified model can be applied to the full model.

## 2.2 Network Models

We have selected the regulation of SdiA in *E.coli* as our example network. As discussed in the previous section, it is considered likely that the signal molecule AI-2 may be involved in the SdiA protein upregulation of its own expression by binding to it. It should be noted, however, that the network structures that we study are

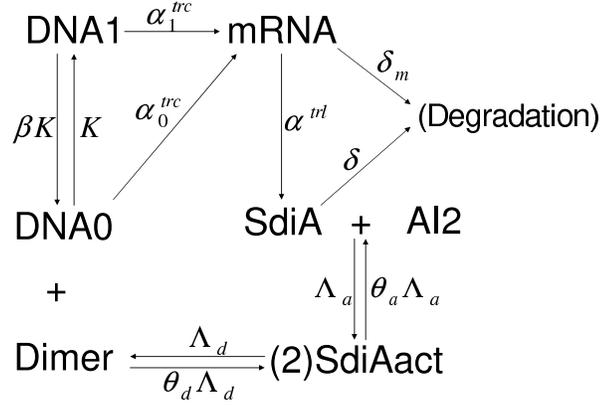


Figure 2.1: Full model with signal-mediated positive autoregulation.

more general and several kinetic parameters in the simulations are assigned values representing typical cellular activities. Therefore we expect that conclusions reached from this study on the stochastic behavior of the network can be applicable to other signal-mediated positive autoregulation networks in other bacteria.

### 2.2.1 Full Model

The signal-mediated positive autoregulation network, termed the full model, is shown in Fig. 2.1. The gene (DNA) has two states. Transcription rates under the two states are different, so mRNA can accumulate at different rates. After transcription, protein monomers (SdiA) are translated from mRNA and then form activated protein complexes (SdiAact) by binding signal molecules (AI2). Two activated complexes bind each other and become a dimer. The dimers determine which state the gene is in. When the dimer binds to the gene, the gene is in occupied state (DNA1) and the transcription is much faster than that of the unoccupied gene (DNA0). So we have a positive autoregulation which is dependent on signal

Parameters	Values	References
$\alpha_0^{trc}$	$1.2 \text{ min}^{-1}$	based on transcription rate range: $10^{-4}$ - $1 \text{ s}^{-1}$ [58]
$\alpha_1^{trc}$	$60 \text{ min}^{-1}$	set to maximum rate 1 transcription per sec [58]
$\delta_m$	$0.36 \text{ min}^{-1}$	based on mRNA half life range: $40 \text{ s}$ - $20 \text{ min}$ [59]
$\alpha^{trl}$	$6 \text{ min}^{-1}$	same order of magnitude as Chen <i>et al.</i> [61]
$\beta$	28	same order of magnitude as Kepler and Elston [12]
$K$	$1 \text{ min}^{-1}$	same order of magnitude as Kepler and Elston [12]
$\delta$	$2.5 \text{ min}^{-1}$	corresponds to half life of $17 \text{ s}$ within the range for proteins [60]
$\theta_a$	1000	selected to be higher than $\theta_d$
$\Lambda_a$	$1 \text{ min}^{-1}$	adjusted based on Kepler and Elston [12] within same order of magnitude
$\theta_d$	250	adjusted based on Kepler and Elston [12]
$\Lambda_d$	$1 \text{ min}^{-1}$	adjusted based on Kepler and Elston [12] within same order of magnitude

Table 2.1: Parameters of full model.

molecules. Here we assume it is the dimer that binds to the gene because the regulatory proteins often function as dimers or higher-order oligomers [12, 57].

The parameter values for Fig. 2.1, i.e., the rate constants of the reactions, are listed in Table 2.1. The transcription rate,  $\alpha_0^{trc}$ , for the unoccupied gene is selected near the middle point on a logarithmic range for *E. coli* [58] and it corresponds to 200 transcripts per generation. The rate for the occupied gene,  $\alpha_1^{trc}$ , is set to the maximum of the range. The mRNA decay rate,  $\delta_m$ , corresponds to a half-life

of 2 minutes, which is within the range of mRNA half lives in *E. coli* [59]. The protein decay rate,  $\delta$ , corresponds to a half life of 17 s (half lives of proteins in cells vary widely [60]). The translation rate,  $\alpha^{trl}$ , is based on Chen *et al.* [61], and  $\beta$ ,  $K$  on Kepler and Elston [12] with some adjustments within the same order of magnitude. The parameters for the formation of the complex and the dimerization were selected to result in parameter values for the simplified model, developed in the next section, close to those used in Kepler and Elston [12], while keeping the rate for the formation of the complex higher than the dimerization rate.

There are two physical meanings for these rate constants. In the deterministic model (ODE model), the constants are kinetic rate constants. In the stochastic model, they are stochastic rate constants, which denote the probability of the reaction occurrence [5]. When the deterministic rate constants are in terms of concentrations, the stochastic rate constants are equal to them for first-order reactions. For second-order reactions, the relation between stochastic rate and deterministic rate is [15]:

$$\beta = \frac{k}{VN_A} \quad (2.1)$$

Here  $\beta$  is the stochastic rate constant,  $k$  is the deterministic rate constant in terms of concentrations,  $V$  is the cell volume,  $N_A$  is Avogadro's number. In the case where the deterministic rate constants are in terms of molecule numbers, for both the first-order and second-order reactions, the deterministic rate constants have the same values and units as the stochastic rate constants. Therefore the values in Table 2.1 can be used in both cases.

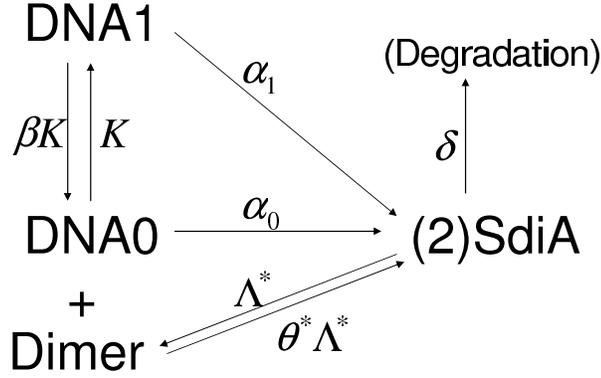


Figure 2.2: Simplified model with positive autoregulation.

### 2.2.2 Simplified Model

The simplified model is shown in Fig. 2.2. This model is a network in which positive autoregulation does not involve signal molecules. To obtain it from the full model, two simplifications are involved: First, transcription and translation, the two steps of gene expression in the full model, are merged into one reaction:



In the simplified model, mRNA is eliminated so we do not need to consider its activity. The second simplification is that activation and dimerization of protein monomers (SdiA) are simplified into a single reversible reaction:



After this simplification, the activated protein complex (SdiAact) and signal molecules (AI2) are eliminated from the equations. The simplified model has the same structure as the general model examined by Kepler and Elston [12], where an analytical expression for the probability distribution was obtained.

These two simplifications are based on three assumptions: (i) The mRNA number is assumed at steady state. (ii) The number of free signal molecules is considered constant; (iii) Activation of protein is much faster than dimerization of the activated complex. Next we show how to derive the model from the assumptions.

The first simplification is based on the first assumption. In the full model from Fig. 2.1 the translation kinetics are:

$$r_{SdiA} = \alpha^{trl}[mRNA] \quad (2.5)$$

The notation  $[X]$  represents the molecule number of species X in the cell. With the assumption that mRNA is always at steady state, the mRNA number in the full model (Fig. 2.1) is obtained:

$$[mRNA] = \frac{\alpha_0^{trc}}{\delta_m}[DNA0] + \frac{\alpha_1^{trc}}{\delta_m}[DNA1] \quad (2.6)$$

By applying eq. (2.6) into (2.5) a simplified rate function for protein monomer (SdiA) production is obtained:

$$r_{SdiA} = \alpha_0[DNA0] + \alpha_1[DNA1] \quad (2.7)$$

where

$$\alpha_0 = \alpha_0^{trl} \alpha^{trc} / \delta_m \quad (2.8)$$

$$\alpha_1 = \alpha_1^{trl} \alpha^{trc} / \delta_m \quad (2.9)$$

Equation (2.7) is just the kinetic rate function for reactions (2.2) and (2.3).

The second simplification is obtained from the last two assumptions. It is assumed that activation is much faster than dimerization. Hence, compared with

dimerization, activation can be considered at quasi-equilibrium:

$$[SdiAact] = \frac{[SdiA][AI2]}{\theta_a} \quad (2.10)$$

On the other hand, in those two steps dimerization is the determining step. Then for the overall reaction, the forward reaction rate is

$$r_f = \Lambda_d [SdiAact]^2 = \Lambda^* [SdiA]^2 \quad (2.11)$$

where

$$\Lambda^* = \frac{\Lambda_d [AI2]^2}{\theta_a^2} \quad (2.12)$$

The overall reverse reaction rate is:

$$r_r = \theta_d \Lambda_d [Dimer] = \theta^* \Lambda^* [Dimer] \quad (2.13)$$

where

$$\theta^* = \frac{\theta_a^2 \theta_d}{[AI2]^2} \quad (2.14)$$

Equations (2.11) and (2.13) describe the kinetics for the overall reaction (2.4) for the simplified model. Hence the second simplification is realized. From eqs. (2.12) and (2.14) it can be concluded that  $\theta^*$  and  $\Lambda^*$  can be used as effective rate constants because of the second assumption that the free signal molecule (AI2) number is constant.

Equations (2.8), (2.9), (2.12) and (2.14) indicate the relation between the parameters of the simplified and full models. With these equations and the constant signal molecule (AI2) number equal to 500, values for parameters of the simplified model are calculated and shown in Table 2.2.

Parameters	Values
$\alpha_1$	$1000min^{-1}$
$\alpha_0$	$20min^{-1}$
$\delta$	$2.5min^{-1}$
$\beta$	28
$K$	$1min^{-1}$
$\theta^*$	1000
$\Lambda^*$	$0.25min^{-1}$

Table 2.2: Parameters of simplified model.

## 2.3 Analysis of Simplified Model

### 2.3.1 Steady-State Distribution from Mathematical formula

For networks with the structure of the simplified model, an analytical expression for the stochastic distribution under steady state has been obtained by Kepler and Elston [12]. This distribution is based on two approximations. The first approximation is the small noise approximation. When the protein abundance is large, the number of protein monomer (SdiA) molecules can be expressed by defining a continuous variable instead of a discrete one. The second is the fast noise approximation, which means the transitions between the two gene states (DNA0 and DNA1) are fast but finite.

With these approximations, the steady-state density function of protein monomer

Parameters	Values
$a_0$	0.02
$b$	0.175
$\kappa$	64
$m_o$	400

Table 2.3: Values for Rescaled Parameters.

number is:

$$\bar{\rho}(x) = \frac{\lambda}{B(x)} \exp\left(2 \int_0^x \frac{A(x')}{B(x')} dx'\right) \quad (2.15)$$

where

$$A(x) = \frac{ba_0 + x^2}{b + x^2} - x - \frac{2xb(a_0 - 1)[((a_0 - 2) + x)x^2 + b(x - a_0)]}{\kappa(b + x^2)^4} \quad (2.16)$$

$$B(x) = \frac{1}{m_o} \left( \frac{b(a_0 + x) + x^2(1 + x)}{b + x^2} \right) + \frac{2bx^2(a_0 - 1)^2}{\kappa(b + x^2)^3} \quad (2.17)$$

(Note that there are a number of typographical errors in the equations in Kepler and Elston [12] corresponding to (2.15) and (2.17) above. Results shown in Fig. 6 in Kepler and Elston [12] match eqs. (2.15) and (2.17).) There are four dimensionless rescaled parameters in the equations above:  $a_0 = \alpha_0/\alpha_1$ ,  $b = \beta\theta\delta^2/\alpha_1^2$ ,  $\kappa = K\alpha_1^2/(\theta\delta^3)$ ,  $m_o = \alpha_1/\delta$ . With the parameter values in Table 2.2, the values for the rescaled parameters are calculated as shown in Table 2.3.  $x$  is the stochastic variable denoting the number of protein monomer.  $x$  is also dimensionless with the relation  $x = [SdiA]/m_o$ . The parameter  $\lambda$  is used to normalize the distribution  $\bar{\rho}(x)$ .

The distribution of protein monomer number obtained from eq. 2.15 is shown in Fig. 2.3. In this figure the plot of the density function (solid curve) is a two-peak

curve, which means the pathway of protein production has a bifurcation. The bifurcation separates the bacterial cells into two subpopulations. In one subpopulation, cells produce protein in a slower rate. There is a lower number of protein monomers in the cells and thus a lower number of activated protein complexes and the dimers. Then, in most of the time, the gene is in the unoccupied state and protein production rate becomes even slower. As time goes on, this group of cells is now at steady state with a lower protein number. The other subpopulation is the opposite. Faster protein production causes higher number of dimers and puts the gene in the occupied state more frequently, resulting in much faster protein production rate. Cells in this group are at steady state with a higher protein number. The distribution discussed here is a steady-state distribution.

### 2.3.2 Stochastic Simulation

In this work, stochastic simulation is realized with stochastic Petri networks (SPN). An SPN is a mathematical formalism for representing and simulating events that are stochastic and discrete. Goss and Peccoud [15] showed that they are a very good tool for the stochastic simulation of molecular biological systems and since then they have been used more extensively for this purpose. The SPN for the simplified model is constructed and shown in Fig. 2.4. Cell activities like gene expression, molecular binding and unbinding are all described as chemical reactions. In the SPN the tokens (circles) represent the chemical species including DNA, mRNA, proteins and signal molecules. The transitions (vertical lines) represent the chem-

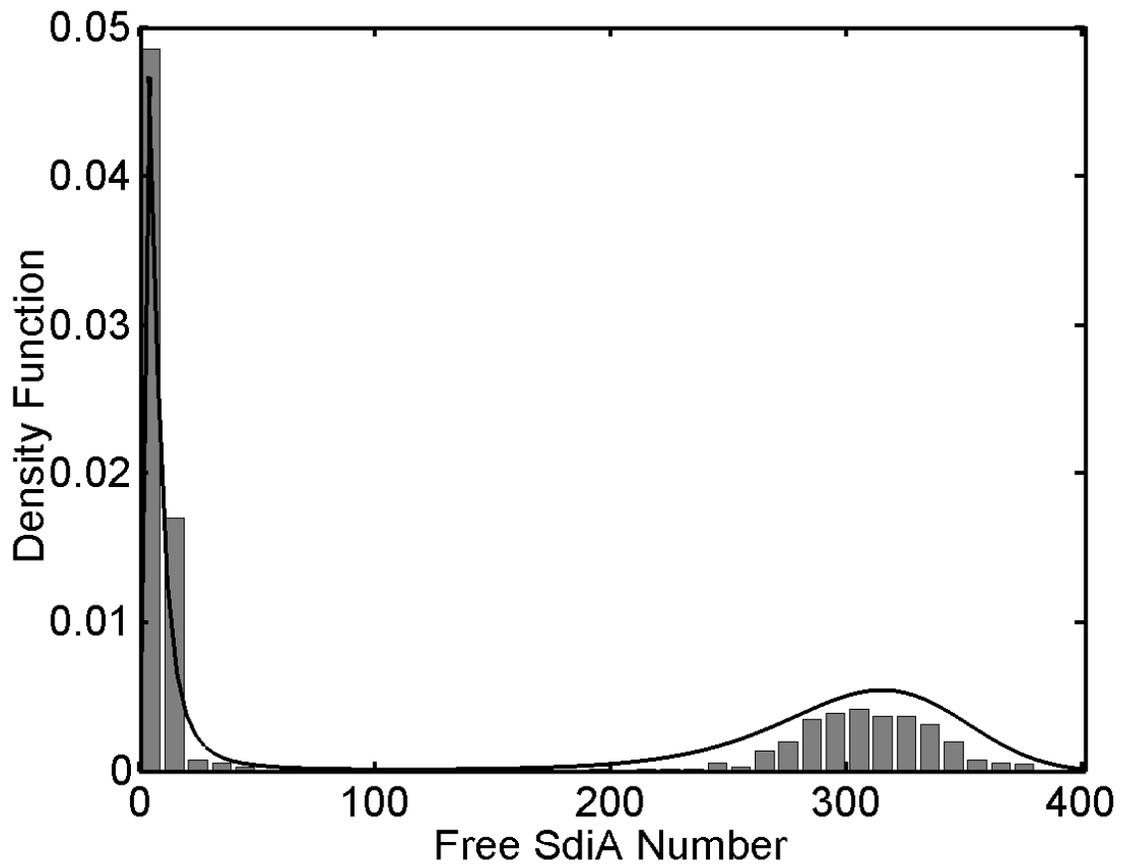


Figure 2.3: Distribution of protein monomer (free SdiA) number for the simplified model. Solid curve: analytical steady-state distribution of monomer number. Bar graph: distribution obtained from SPN simulation after 50-minutes.

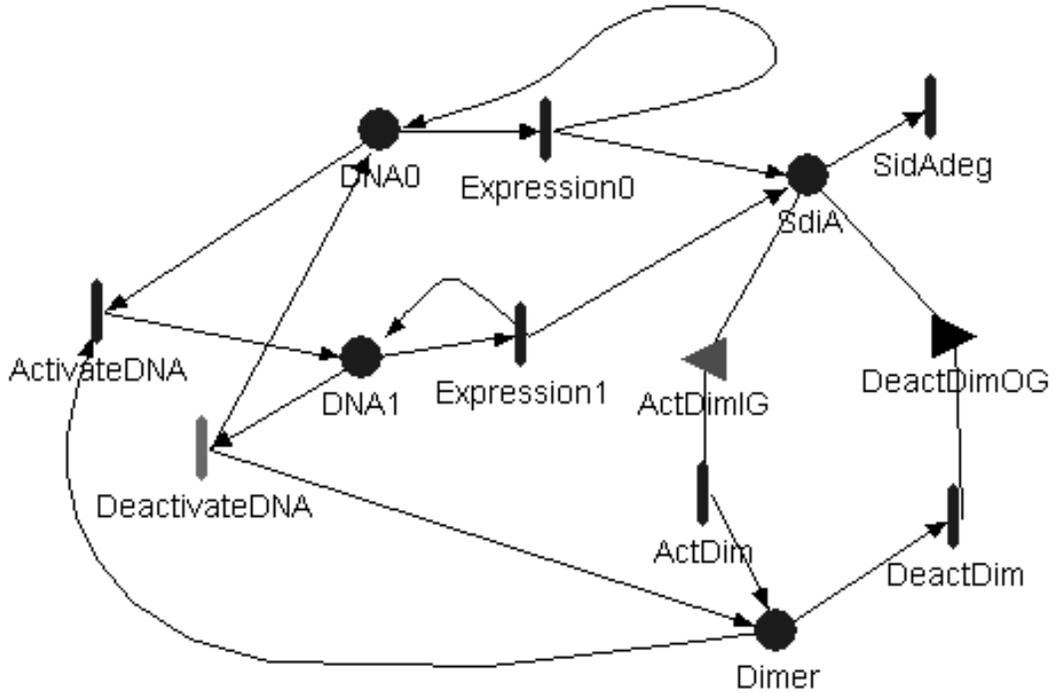


Figure 2.4: SPN for simplified model.

ical reactions, with arrows pointing from reactants and into products. For second order reactions, there are input and output gates (triangles) in Fig. 2.4, indicating preconditions and effects of the transition firing, because reactants and products of the corresponding reactions involve two molecules. The Mobius software package is used to simulate the SPN as we mentioned in Chap. 1.

The result of the SPN stochastic simulation after 50 minutes is the bar graphic in Fig. 2.3, which matches the analytically obtained steady state distribution (solid curve) very well. The analytical steady-state distribution is derived from master equations with the two approximations mentioned above [12]. Both master equations and SPN suppose that the reactions are Markov chains in which the probability of reaction depends on current molecule numbers [5, 15]. Hence, since the two methods

are based on the same stochastic foundations, one would expect the results to match well.

The analytical steady-state distribution can predict the long-run trend of stochastic simulation. The distribution from the stochastic simulation is not a steady-state distribution. Unlike the steady-state distribution, which is independent of initial condition, stochastic simulations with different initial conditions may result in different distributions. However, as long as the system evolves for long enough time, stochastic simulation can reach the same steady-state distribution from any initial condition. In Fig. 2.3 the two distributions match very well because we choose a proper initial condition. From this initial condition the cell can almost reach steady state within generation time. (The time length of 50 minutes is an approximate generation time of *E. coli cells*.)

### 2.3.3 Deterministic Simulation

With small-noise and fast-noise approximations, the deterministic ODE for the simplified model is [12]:

$$\frac{dx}{d(\delta t)} = \frac{ba_0 + x^2}{b + x^2} - x \quad (2.18)$$

Note that “scaled time”  $\delta t$  should be used here, while in Kepler and Elston [12] the corresponding equation uses  $t$ . The ODE solution is shown in Fig. 2.5.

Comparison between the deterministic and the stochastic simulation indicates the existence of bifurcation. For monostable systems, the stochastic mean is close to the deterministic path, so such systems can be adequately described in deterministic

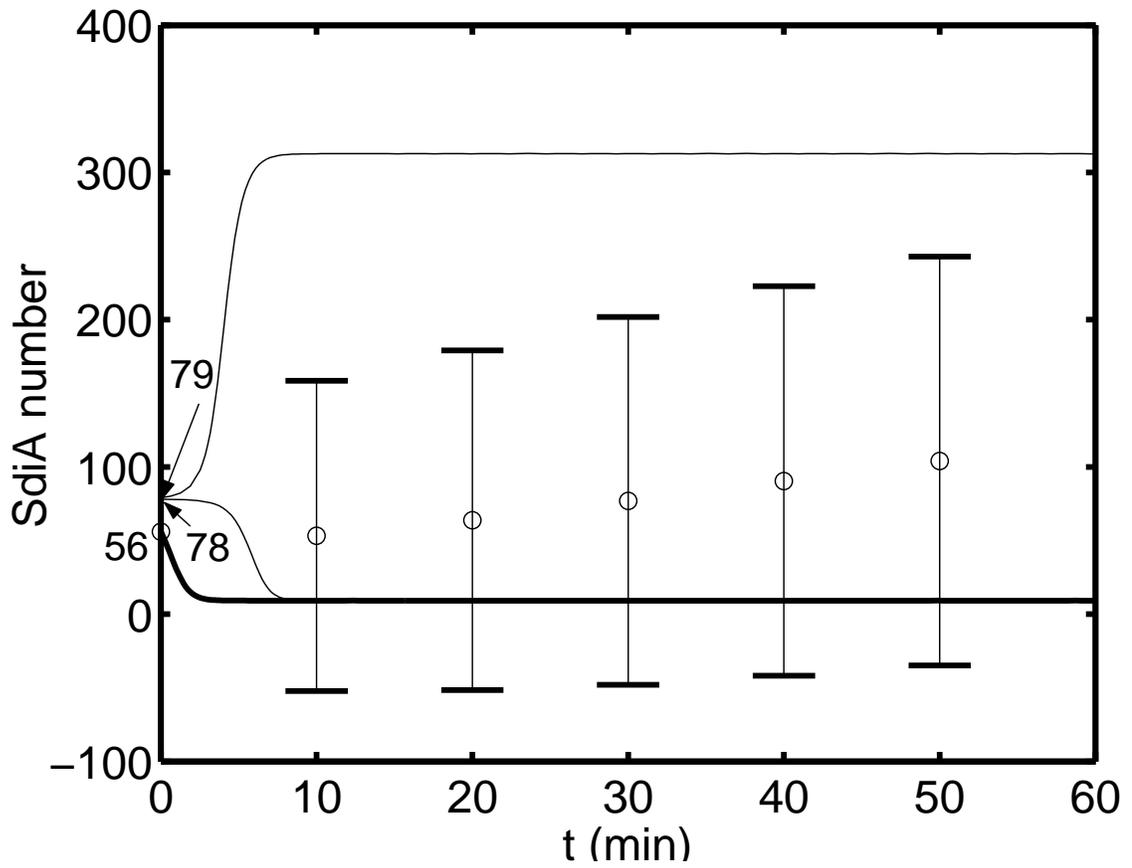


Figure 2.5: Protein monomer (free SdiA) number versus time for the simplified model. Circles: mean of stochastic simulation; bar graphs show the standard deviation. Bold solid curve: deterministic path from 56 initial protein monomers, which is the same initial condition as that of stochastic simulation. Thin solid curves: deterministic paths from the initial conditions above and below the switch point.

form. However, the system we are examining is a bistable system. Therefore, from the same initial condition with protein monomer number equal to 56, the stochastic mean (circles in Fig. 2.5) does not match the deterministic path (bold solid curve in Fig. 2.5). We can also see that the standard deviation for the stochastic simulation (bar graph in Fig. 2.5) is very large relative to noise, which indicates the bifurcation of the cell population.

The deterministic model has a switch point between 78 and 79 molecules in Fig. 2.5. The two deterministic paths starting above and below the switch point (thin solid curves) verify the existence of bifurcation. Initial number of protein monomer higher than the switch point leads to the higher stable steady state. And lower initial protein monomer number leads to the lower stable steady state. Of course, for the stochastic simulation, although the initial condition is below the deterministic switch point, a cell subpopulation at high SdiA number is also obtained, as clearly shown in Fig. 2.3.

## 2.4 Analysis of Full Model

### 2.4.1 Stochastic Simulation

The SPN of the full model is shown in Fig. 2.6. As we mentioned before, gene expression is separated into transcription and translation. Activation and dimerization are not considered as one step here. The result shown in Fig. 2.7 is the distribution of protein monomer number in cells with 50-minute age for the full model.

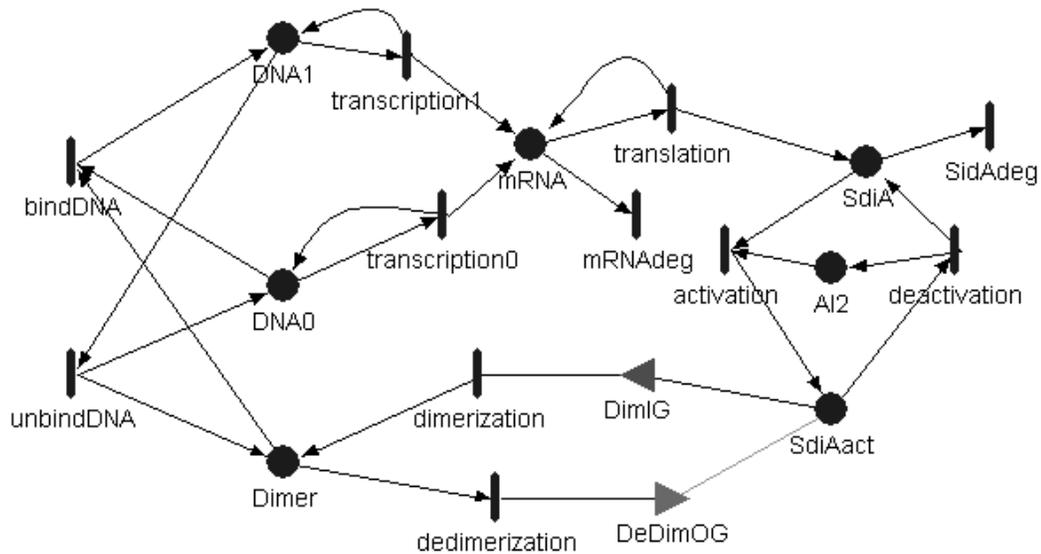


Figure 2.6: SPN for full model.

The distribution in Fig. 2.7 is very similar to that of the simplified model. First, like the stochastic simulation result for the simplified model, there is bistability in the distribution for the full model. So the full model also predicts the stochastic bifurcation for the positive autoregulation network, which verifies the conclusions from the simplified model. Second, the shapes of the distributions from the two models are quite similar. For both the simplified and the full model, the peaks at the lower number of protein monomer are tall and thin, which indicates that in this group of cells the fluctuations of protein number are small. The peaks at the higher protein number are more flat, so the fluctuations in those cells are relatively large.

The higher-number peaks in Figs. 2.3 and 2.7 are not in exactly the same position. This is because of the second assumption for the simplification from the

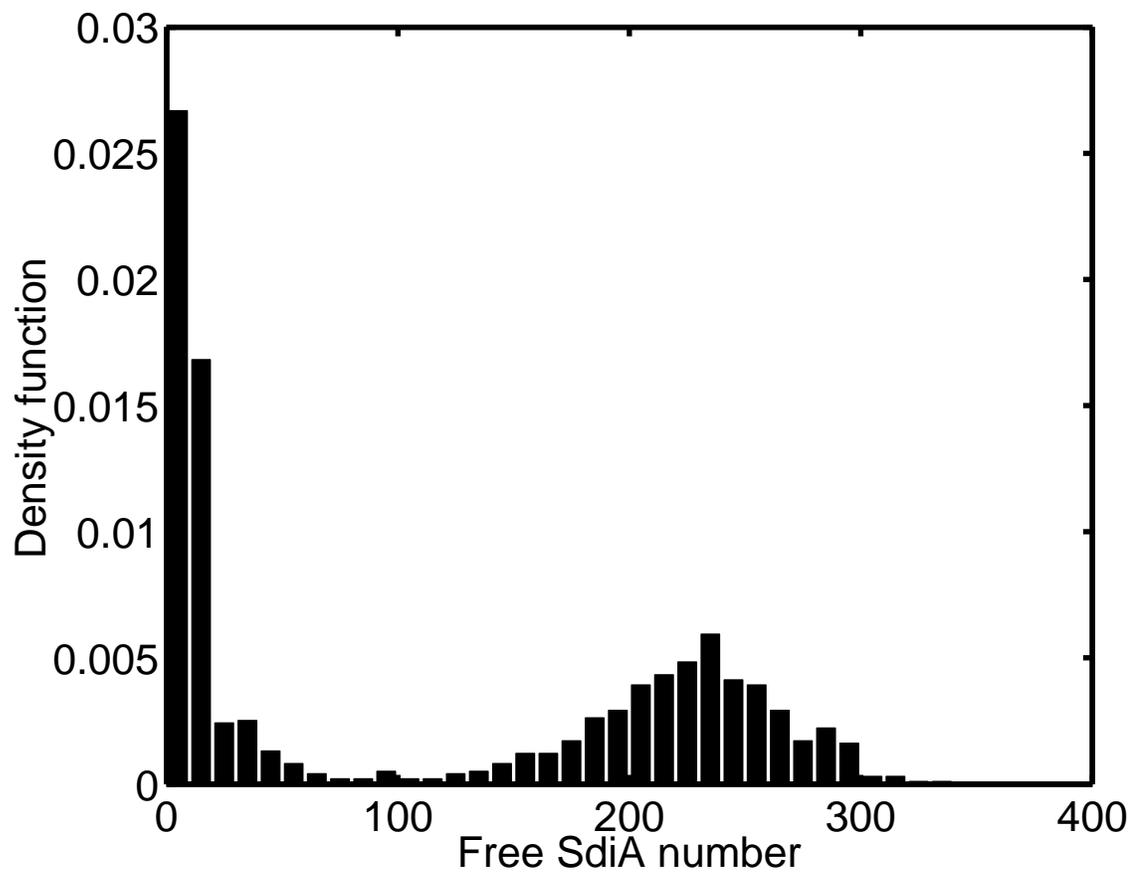


Figure 2.7: Distribution of protein monomer (free SdiA) number for the full model.

full model to the simplified model. In the simplified model, the free signal molecule number is assumed to be equal to 500. In the full model, we assume the total number of all forms of signal molecules, including free signal molecules and signal molecules in the form of activated protein complexes and dimers, to be a constant equal to 600. However the difference between the two peak values is not significant (around 50). Therefore the simplification of the full model does not change the peak positions much.

An important reason for using the simplification is that for the full model we cannot obtain an analytical distribution as we did for the simplified model. We have to simplify the full model first, and obtain a mathematical description for the simplified model. Such a description of the distribution allows a detailed analysis of the effect of different parameters on the presence or not of stochastic bistability [12].

As we discussed earlier, the analytical distribution matches the simulation distribution for the simplified model very well, and the stochastic simulation distributions of the full and the simplified models are very similar. The result obtained from mathematical analysis of the simplified model can be expected to be applicable to the full model.

## 2.4.2 Deterministic Steady-State Analysis

In the previous section, we determined that the analytical distribution for the simplified model can suggest the steady-state distribution for the full model.

SPN simulation can further provide distributions for the full model at finite times. However, the direct mathematical analysis for the full model is very difficult. Often a deterministic analysis can indicate a pathway bifurcation that may also present for the stochastic system. Here we examine the use of deterministic analysis as an indicator for the bistable distribution of the full model.

First, for simplicity of notation, a variable  $z$  is defined as equal to the number of dimers:

$$z = [Dimer] \tag{2.19}$$

At the deterministic steady states, protein production and degradation in Fig. 2.1 are in balance:

$$\alpha_0[DNA0] + \alpha_1[DNA1] - \delta[SdiA] = 0 \tag{2.20}$$

In eq. 2.20 the third term is the degradation rate of protein. The summation of the first two terms is the production rate of protein, in eq. 2.7. Eq. 2.7 is used for the simplified model, but at the deterministic steady states it can also be used for the full model, because the simplification was based on the assumption that [mRNA] is at steady-state.

In Fig. 2.1, the transition of gene states of the full model is at equilibrium:

$$\frac{[DNA0]}{[DNA1]} = \frac{\beta}{z} \tag{2.21}$$

Furthermore, the gene states can be normalized to a single gene. Then we have

$$[DNA0] + [DNA1] = 1 \tag{2.22}$$

From eqs. 2.20, 2.21 and 2.22, the steady-state protein monomer number is obtained as

$$[SdiA] = \frac{\alpha_0}{\delta} \frac{\beta}{\beta + z} + \frac{\alpha_1}{\delta} \frac{z}{\beta + z} \quad (2.23)$$

At steady state, activation and dimerization in Fig. 2.1 are in equilibrium. So we obtain

$$\theta_a = \frac{[SdiA][AI2]}{[SdiAact]} \quad (2.24)$$

$$\theta_d = \frac{[SdiAact]^2}{[Dimer]} \quad (2.25)$$

Multiply eq. 2.25 by the square of eq. 2.24, and use eq. 2.19 to obtain:

$$[AI2] = (\theta z)^{1/2} [SdiA]^{-1} = (\theta z)^{1/2} \frac{\delta(\beta + z)}{\alpha_0\beta + \alpha_1 z} \quad (2.26)$$

where  $\theta = \theta_a^2 \theta_d$

From eq. 2.25, the number of activated protein complex is expressed as

$$[SdiAact] = (\theta_d z)^{1/2} \quad (2.27)$$

From eqs. 2.21 and 2.22, the number of DNA binding dimer is

$$[DNA1] = \frac{z}{\beta + z} \quad (2.28)$$

When we consider the total signal molecule number, we must count free signal molecules, signal molecules binding to the activated protein complex, and those binding to dimers (including the dimer binding DNA). Therefore the total number of signal molecules is:

$$\begin{aligned} [AI2total] &= [AI2] + [SdiAact] + 2[Dimer] + 2[DNA1] \\ &= (\theta z)^{1/2} \frac{\delta(\beta + z)}{\alpha_0\beta + \alpha_1 z} + (\theta_d z)^{1/2} + 2z + \frac{2z}{\beta + z} \end{aligned} \quad (2.29)$$

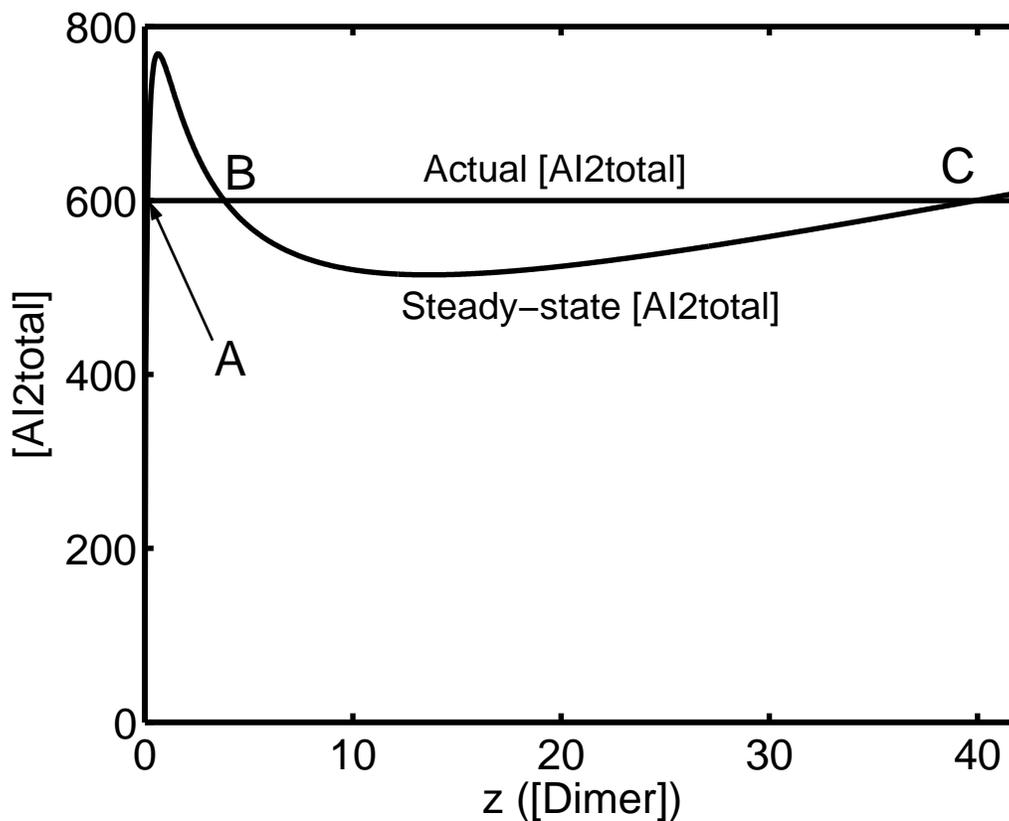


Figure 2.8: Total signal molecule number ( $[AI2total]$ ) vs.  $z([Dimer])$ .

Eq. 2.29 is used to obtain a plot of  $[AI2total]$  as a function of  $z$ , shown in Fig. 2.8. The full model assumes that the total number of signal molecules is a constant. It can be found from Fig. 2.8 that when this constant is within a certain range, the curve has more than one intersection point with the horizontal line, which means that the deterministic model has multiple steady states. In these cases, pathway bifurcation is possible and the stochastic distribution may also be bistable. Therefore the curve for  $[AI2total]$  vs.  $z$  can indicate the existence of bistability.

Let us suppose that the total AI-2 number is equal to 600, as it was in the stochastic simulation. Then the points A, B and C in Fig. 2.8 represent three steady

Steady states	A	B	C
$[Dimer]$	0.1394	3.8215	39.9287
$[SdiA]$	9.9422	55.0765	238.4189

Table 2.4: Steady states when  $[AI2_{total}] = 600$ .

states. The protein monomer number at each steady state is calculated from eq. 2.23 and shown in Table 2.4. Consider a point between A and B. In that case the steady-state total signal number is higher than the actual total signal number. This means there are not enough signal molecules, so the dimers and activated complexes will disassociate and their number will decrease. The system goes to point A. As for a state between B and C, the actual signal number is higher than the steady-state total signal number. This means there is a surplus of signal molecules, so more activated complexes and dimers will form and the state goes to point C. Therefore A and C are stable states for the deterministic model and B is a switch point. The protein monomer numbers for A and C are close to the peaks in Fig. 2.7.

## 2.5 Discussion

We present two gene regulation network models, both of which contain positive autoregulation. In the full model, shown in Fig. 2.1, the positive autoregulation is mediated through signal molecules. The gene product protein binds signal molecules first, then forms dimers which subsequently bind the gene and improve its expression. In the simplified model, shown in Fig. 2.2, no signal molecule is involved. The

gene product protein forms a dimer directly, then the dimers up-regulate the gene expression. The simplified model can be obtained from the full model with two simplifications.

Both models exhibit bistable distributions of protein numbers as shown in Figs. 2.3 and 2.7, which means that an initially uniform cell population can evolve into two subpopulations. Cells in one population have low protein numbers, while cells in the other have high protein numbers. The bistability of the two models is the result of the positive autoregulation mechanism and their stochastic nature. However, obtaining the bistable protein number distributions for the two models involves different levels of difficulty. For comparison purposes, on a standard single processor, our SPN simulation takes more than ten hours for the full model, when one thousand batches (single path simulations) are used to obtain the distribution. For the simplified model, the same type of simulation takes less than one hour. The calculation of the analytical steady state distribution for the simplified model is essentially instantaneous, but it requires the additional small-noise approximation, which requires large protein number, and the fast-transition approximation, which needs fast but finite transition between the gene states.

Let us further discuss the meaning and effect of the two simplifications we introduced. The first is that the transcription and translation steps are merged into one single step of gene expression, which assumes the mRNA number is always at steady state and can be calculated from eq. 2.6. This assumption has two effects. One is that the dynamic simulation is faster because the time it takes for mRNA number to reach steady state is ignored. This effect is not important though when

considering steady state or long times in simulation. The second effect is caused by ignoring the fluctuation of mRNA number, because with this assumption, when the full model is at steady state and the state of the gene does not change, the mRNA number is fixed. The mRNA fluctuation does not play an important role in the presence or not of bistability. In the simplified model, the fluctuation in the gene states can induce bistability while the fluctuation of monomer number can reduce bistability [12], suggesting that the fluctuation in gene states is the main source of bistability. The fact that the distributions in Fig. 2.3 and Fig. 2.7 are both bistable and have similar shapes, indicates that mRNA fluctuation is mostly unrelated to bistability and does not have a significant influence. The peaks in Fig. 2.3 are a little thinner and higher than in Fig. 2.7, indicating that ignoring the mRNA fluctuation only causes a slight reduction of the protein number fluctuation. Therefore, the second effect is also not important and the assumption behind this simplification is reasonable.

The second simplification is that the activation and dimerization of protein monomers are combined into one step. Here it is assumed that the number of unbound signal molecules is a constant. This assumption can influence molecule numbers at steady state. This is why the position of the peak for the higher protein number in Fig. 2.3 has an observable difference from the position of the corresponding peak in Fig. 2.7. To avoid unacceptable differences, the steady state free signal number for the full model cannot be too far away from the assigned constant value in the simplified model. For this simplification, there is an additional assumption that activation is much faster than dimerization so that it can be considered at

quasi-equilibrium. However, usually even dimerization is fast enough to be thought at equilibrium [12]. This assumption is expected to be reasonably satisfied.

The discussion above indicates that the number of signal molecules present can be a determining factor on the presence or not of stochastic bistability. The difficulty involved in the direct mathematical analysis of the full stochastic model led us to consider the use of the corresponding deterministic model. We saw that such an analysis can be a very useful indicator of stochastic bistability. Furthermore, we were able to directly focus on the effect of the number of signal molecules as shown in Fig. 2.8. This supports that further work is needed in connecting the signal mediated positive autoregulation model with uptake models for the signal molecules. We have such work under way for the AI-2 uptake in *E. coli*.

Finally, we note that the transient simulation results in Fig. 2.5 indicate that stochastic simulation is essential for a bistable system. For a monostable system, the mean for the stochastic simulation usually matches the deterministic path and the standard deviations are relatively small. Therefore deterministic simulation is usually sufficient and the time-consuming stochastic simulation may not be necessary. However, for a bistable system, the mean path for the stochastic simulation lies between the two deterministic stable steady states, while the deterministic simulation only goes to one of the two stable steady states depending on the initial condition. As for the standard deviations, they are very large as a result of the development of two distinct cell subpopulations.

## Chapter 3

# Cell Age Distribution in Relation to Stochastic Bifurcation of Key Protein Number

### 3.1 Introduction

Stochastic fluctuations in gene regulation networks may lead to bifurcation of protein number, which can cause an initially homogeneous population to partition into two subpopulations [8]. Often the bifurcation can be directly observed from experiment. For example, to emulate the two states of HIV-1, a vector can be constructed with GFP and selected HIV-1 genes. Then bright infected cells have high transcription level of those genes and thus are in active state. Dark cells have low transcription level and are in latent state [62].

The bifurcation of some proteins plays an important role in cell division. For example, the cyclin-dependent kinase is in low activity state while the cell is not dividing, and it has a high activity when the cell is dividing [63]. In *Xenopus* oocytes cells, the p42 mitogen-activated protein kinase (MAPK) and the cell-division cycle protein kinase Cdc2 form a “positive-feedback-based bistable memory module”, which can regulate cell division [64]. In *Escherichia coli*, SdiA protein, which may have a bistability is involved in the cell division. Overexpression of SdiA can activate  $P_{Q2}$ , a promoter upstream of *ftsQAZ* gene, and then initiate the expression

of ftsQAZ [38]. Protein FtsQAZ is necessary for cell division. FtsZ is used during the Z-ring formation and FtsA is required for the formation of cross wall [38]. Thus a high number of those proteins is required during cell division. In *E.coli*, there are about 5,000-20,000 FtsZ monomers in one cell during exponential growth [65, 66, 67]. These proteins above can be considered as “key proteins” for cell division. As a simplification, we assume that a cell divides once the key protein number inside reaches a threshold value.

The pattern of key protein number evolution can influence the age distribution of cell population. Because the cellular activities are often stochastic [5], key protein numbers in different cells would take different times to reach the threshold for cell division, although they have the same initial values. Then cells would divide at different times and form a specific age distribution. Age distribution is experimentally measurable [68, 69, 70, 71]. Therefore if we can calculate age distribution from the pattern of key protein number evolution, the calculated results can be compared with experimental results and the bifurcation of key protein number can be testified.

Mathematical analysis of cell age distribution has a long history. As early as in the 1930s, it was discovered that under identical condition, two sister cells have different generation time, i.e. the length of the duration from birth to division completion [72, 73]. Much work has been done to interpret the variability of the generation time and simulate the age distribution. Some work used deterministic models for biochemical processes inside cells [73, 74, 75, 76, 77]. In these models a particular variable, such as the age, size and mass of the cell, is assumed to be able to trigger cell division. This variable is deterministic. However its value when cell

division begins is not a fixed value but has a distribution, which therefore causes the variability of the generation time. Other work proposed stochastic models, in which biochemical processes are stochastic. At first it was proposed that cell division is composed of several successive steps and each step is a stochastic process [78, 79]. However this model does not reflect a real cell very well and was questioned by Koch and Schaechter [75]. Another type of model was later developed, in which consider the cell cycle is composed of several phases. Some of these phases are stochastic processes and others are deterministic [80, 81]. However all these models do not link molecular events inside cells to age distribution.

Recently, a stochastic model was suggested, which considers molecular events inside the cell [82]. In this model chromosome replication plays a key role in cell division and the formation of age distribution. However, with the development of molecular biology, current research focuses on particular genes rather than on the whole chromosome. These genes can form gene regulation networks, which may cause the bifurcation of key protein number. Then the amount of key component for cell division may have a bifurcation, which cannot happen while just considering the whole chromosomes. Therefore work needs to be done to relate age distribution with the possible bifurcation of a key protein number.

Here we present mathematical analysis to obtain age distribution of cell population, with the supposition that the bottleneck for cell division is the key protein number and the protein number can evolve into bistable distribution. The work has two parts. In the first part alive probability will be obtained from the Fokker-Planck equation and Stochastic Petri nets, and the relation between alive probability and

protein number distribution will be analyzed. In the second part age distribution will be obtained from alive probability function, based on balance population model. We mainly work on analytical results so that the effects of parameter values can be studied directly.

## 3.2 Methods for Obtaining the Alive Probability Function

### 3.2.1 Alive Probability Function

Alive probability function is important because it is an intermediate function for obtaining age distribution of cell population from key protein number distribution. In this section, we will study how protein number distribution relates to alive probability function. In the next section, we will discuss how to obtain age distribution from this function.

Suppose initially at time 0, a group of cells were just born. Then alive probability function  $F(t)$  can be defined to be:

$$F(t) = 1 - \frac{\text{no. of cells which have divided at time } t}{\text{initial cell no.}} \quad (3.1)$$

Hence the function is actually the probability that a cell, which is born at time 0, has not divided by the time  $t$ . The value for this function starts from 1 at  $t = 0$ , because the cell is of course alive when it is just born. Then the value decreases monotonically as time  $t$  increases. Finally, the value approaches 0 when  $t$  goes to infinity, which means the cell will eventually divide after long enough time.

Alive probability function is not a new concept. Function  $F_-(\tau)$  proposed

by Powell [74] and  $\alpha$ -plot by Smith and Martin [80] have the same meaning as the function  $F(t)$ . However it seems that this function does not have a commonly accepted name. Here we suppose that after cell division two daughter cells are born and the mother cell is “dead”. Therefore the cells which “have not divided” can be considered as “still alive” and the function  $F(t)$  can be called the “alive probability function”.

### 3.2.2 Computation from the Fokker-Planck Equation

Fokker-Planck equation is used to describe continuous Markov processes [10]. Recently it has been applied into cellular events [12, 5]. If the evolution of a protein number can be described with the Fokker-Planck equations [12]:

$$\partial_{(\delta t)}\rho(x, t) = -\partial_x A(x)\rho(x, t) + \frac{1}{2}\partial_x^2 B(x)\rho(x, t) \quad (3.2)$$

The steady state distribution of the protein number is [10, 12]:

$$\bar{\rho}(x) = \frac{\lambda}{B(x)} \exp\left(2 \int_0^x \frac{A(x')}{B(x')} dx'\right) \quad (3.3)$$

where  $x$  represents the protein number. The functions  $A(x)$  and  $B(x)$  depend on the system. Steady state distribution is important because it indicates the direction of protein number evolution.

The evolution of protein number follows Fokker-Planck equation (3.2) and cell divides once the protein number reaches a threshold. This problem is similar to the problem that a particle, whose position follows the Fokker-Planck equation, escapes from the region between zero and the threshold. To deal with the latter problem,

Gardiner define a function  $G(x, t)$ , which means the probability that the particle, initially at  $x$ , is still in the region at  $t$ . Alive probability function  $F(t)$  can be obtained from the function  $G(x, t)$ . So first we need to solve the function  $G(x, t)$  from a partial differential equation (PDE) [10]:

$$\partial_{(\delta t)} G(x, t) = A(x) \partial_x G(x, t) + \frac{1}{2} B(x) \partial_x^2 G(x, t) \quad (3.4)$$

The initial condition is:

$$\begin{aligned} G(x, 0) &= 1, x \in [0, x_T] \\ &= 0, \text{elsewhere} \end{aligned} \quad (3.5)$$

The boundary condition is:

$$\partial_x G(0, t) = 0 \quad (3.6)$$

$$G(x_T, t) = 0 \quad (3.7)$$

In the equations above,  $x_T$  represents the protein number threshold and  $A(x)$  and  $B(x)$  are the same as in (3.3). After obtaining  $G(x, t)$  from Eq. (3.4),  $F(t)$  is obtained by plugging initial value  $x_0$  into  $G(x, t)$ :

$$F(t) = G(x_0, t) \quad (3.8)$$

For systems, where  $A(x)$  and  $B(x)$  can be obtained, the numerical solution of Fokker-Planck equation can be efficient and convenient in obtaining the protein number distributions and the alive probability functions under different parameter values.

### 3.2.3 Computation from Stochastic Petri Net Simulation

Stochastic Petri network (SPN) is a good tool for simulating stochastic and discrete processes. In SPN, molecular events inside cells, such as transcription, translation, molecular association and disassociation, are considered as chemical reactions and the probability of taking place follows a negative exponential density function [15]. Mobius software can be used to simulate SPN [17].

The SPN designed to obtain alive probability function is shown in Fig. 3.1. In the SPN, the “Protein” place is also connected to other components which are not shown. Those components represent cellular activities which can influence protein number, i.e., the value for “Protein” place. The initial values for “Alive” and “Time” places are set to be 1 and 0. A threshold is set in the input gate of “DivisionIG”. Once simulation starts, i.e., the cell is born, the value for “Time” place is increased by one every minute. When the value for “Protein” place reaches the threshold for the first time, which means the cell divides and “dies” at that time, “Alive” place sets its value to zero and stops accumulating the value for “Time” place. So the value for “Time” place will record the life span of the cell. After simulating a sufficient large number of cells, we have the number of undivided cells at a particular time and let  $F(t)$  to be equal to undivided cell number divided by total cell number.

Compared with Fokker-Planck equation, Stochastic Petri net has two advantages. The first is that SPN is a discrete model, which is more accurate because molecule numbers inside cell are discrete. Second, SPN is convenient for modeling multi-species system, while Fokker-Planck equation with more than a few variables

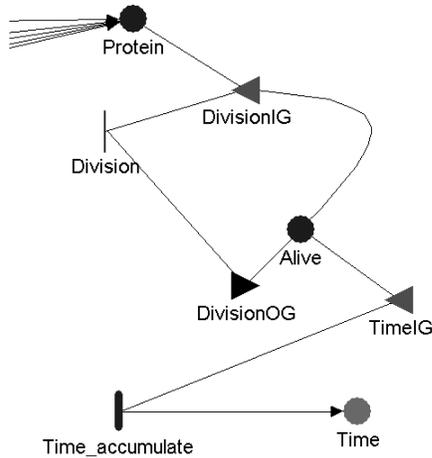


Figure 3.1: General stochastic Petri net to obtain alive probability function.

is often impossible to solve, even numerically [5]. If we use single-variable Fokker-Planck equation, we usually need to make some approximations, which may reduce the accuracy of the result. The disadvantage of SPN is that SPN simulation is often computationally very expensive. Due to the advantages and the disadvantage, we will use different method in different cases. When we want to analyze alive probability functions under several different parameter values, we will use Fokker-Planck equation. When we want to verify whether the conclusion is applicable to complex systems, we will use Stochastic Petri net.

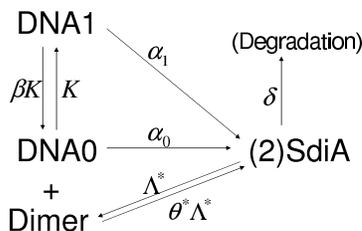


Figure 3.2: Self-promotion network.

### 3.3 Results for Alive Probability Function

#### 3.3.1 Results from Fokker-Planck Equation

Here we use the self-promotion network in Fig. 3.2 as an example. In this network SdiA protein can improve its own expression. We choose SdiA as an example because this protein is involved in cell division. For simplification, we assume that SdiA is a key protein so that cell divides immediately once SdiA protein number reaches a threshold. Although we use SdiA as an example here, this network is not too specific because this mechanism of positive feedback appears in several biological systems [57, 55] and previous study on the stochasticity in this network is general [12]. Therefore the conclusion from this SdiA network can also be used in other self-promotion networks.

In the self-promotion network in Fig. 3.2, if we suppose the reactions are affected by thermal fluctuations, with fast-transition and small-noise approximations, the evolution of protein number can be expressed with Fokker-Planck equations [12]. So the equations (3.3) and (3.4) can be applied here. For this network, the functions

$A(x)$  and  $B(x)$  are [12]:

$$A(x) = \frac{ba_0 + x^2}{b + x^2} - x - \frac{2xb(a_0 - 1)[((a_0 - 2) + x)x^2 + b(x - a_0)]}{\kappa(b + x^2)^4} \quad (3.9)$$

$$B(x) = \frac{1}{m_o} \left( \frac{b(a_0 + x) + x^2(1 + x)}{b + x^2} \right) + \frac{2bx^2(a_0 - 1)^2}{\kappa(b + x^2)^3} \quad (3.10)$$

So the results are dependent on the values for four parameters, which can be obtained from the parameters in Fig. 3.2:  $a_0 = \alpha_0/\alpha_1$ ,  $b = \beta\theta\delta^2/\alpha_1^2$ ,  $\kappa = K\alpha_1^2/(\theta\delta^3)$ ,  $m_o = \alpha_1/\delta$ . We should notice that  $x$  is dimensionless and equal to protein number  $m$  divided by  $m_o$ . In the same way, suppose the threshold is  $m_T$  and initial protein number is  $m_0$ . Then the dimensionless threshold and initial numbers are  $x_T = m_T/m_o$  and  $x_0 = m_0/m_o$ .

The explicit form of alive probability function is not easy to find because it requires an analytical solution for Eq. (3.4). However, by observing the numerical solution of (3.4) and (3.8), we propose a conjecture: if protein number distribution from (3.3) is bistable, then the alive probability function  $F(t)$  from (3.4) to (3.8) can be written in the form of double-exponential function:

$$F(t) = c \cdot \exp(-t/T_1) + (1 - c) \cdot \exp(-t/T_2), \quad T_1 > T_2 \quad (3.11)$$

All the numerical results in this chapter show that after fitting the parameters in (3.11), the double-exponential function matches the numerical solution of (3.4) to (3.8) very well. Next, several different cases will be discussed to show this match and how protein number distribution influences the alive probability function.

**Bifurcation with distant peaks.** When  $a_0 = 0.02$ ,  $b = 0.175$ ,  $\kappa = 64$ ,  $m_o = 400$ , the steady state distribution of protein number is bistable as shown in Fig.

3.3. The two peaks of  $\bar{p}(x)$  are separated and distant from each other, which means the barrier between the two peaks is significant and thus the transition between the two peaks is relatively difficult. We set  $x_0 = 0.25$  and  $x_T = 0.5$ , which indicates that initial protein number is 100 and the threshold is 200. The initial number is between the two peaks because bifurcation is liable to happen when start point is between the two peaks. And we usually assume initial number to be half of the threshold, because when the cell divides the protein number is at the threshold and then after division the initial protein number for the two daughter cells will be half of the threshold. The alive probability function is obtained and shown in Fig. 3.4. The squares are selected points of the numerical solution of (3.4) to (3.8). (We do not show all the points of numerical solution, otherwise they would cover the double-exponential curve.) The solid line is the double-exponential function with the parameters of  $c = 0.5464$ ,  $T_1 = 74.8774$  and  $T_2 = 0.6734$ , which are obtained from fitting. The fit is excellent. Therefore with appropriate parameter values, double-exponential function can be used to represent the alive probability function.

**Bifurcation with partially overlapped peaks.** When  $a_0 = 0.08$ ,  $b = 0.28$ ,  $\kappa = 500$ ,  $m_o = 500$ , the steady state distribution of protein number is also bistable as shown in Fig. 3.5 [12]. However in this distribution the two peaks are so close that the bottom parts of the two peaks even overlap. This means the barrier between the two peaks is weak and thus the transition between the two peaks is relatively easy. Suppose initial protein number is 150 and the threshold is 300, i.e.,  $x_0 = 0.3$ ,  $x_T = 0.6$ . Then alive probability function is obtained and shown in Fig. 3.6. The fitting result is  $c = 0.6139$ ,  $T_1 = 106.2460$  and  $T_2 = 17.1727$ . In this case  $T_1$  is

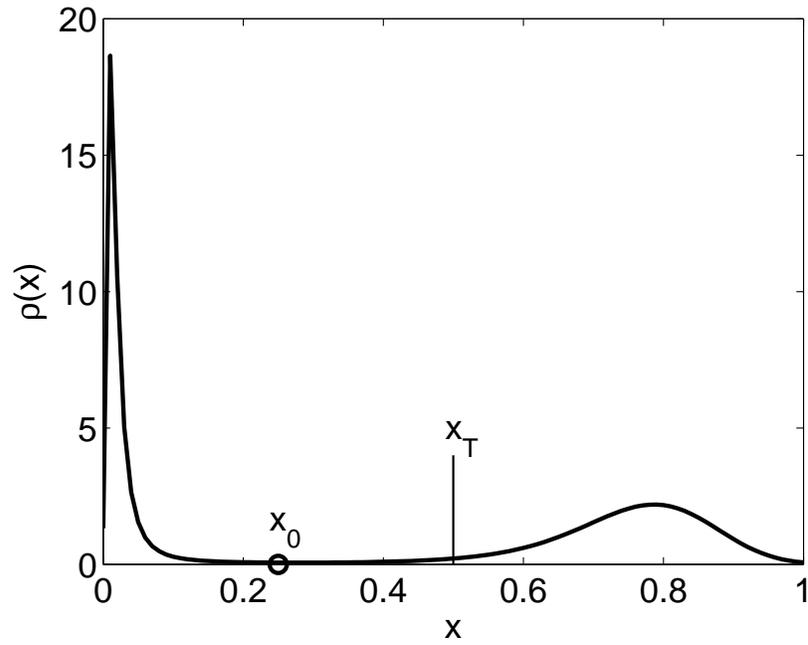


Figure 3.3: Protein number distribution for the case with bifurcation with distant peaks:  $a_0 = 0.02$ ,  $b = 0.175$ ,  $\kappa = 64$ ,  $m_o = 400$ ,  $\delta = 2.5$ , (Data from Chapter. 2) and  $x_0 = 0.25$ ,  $x_T = 0.5$ .

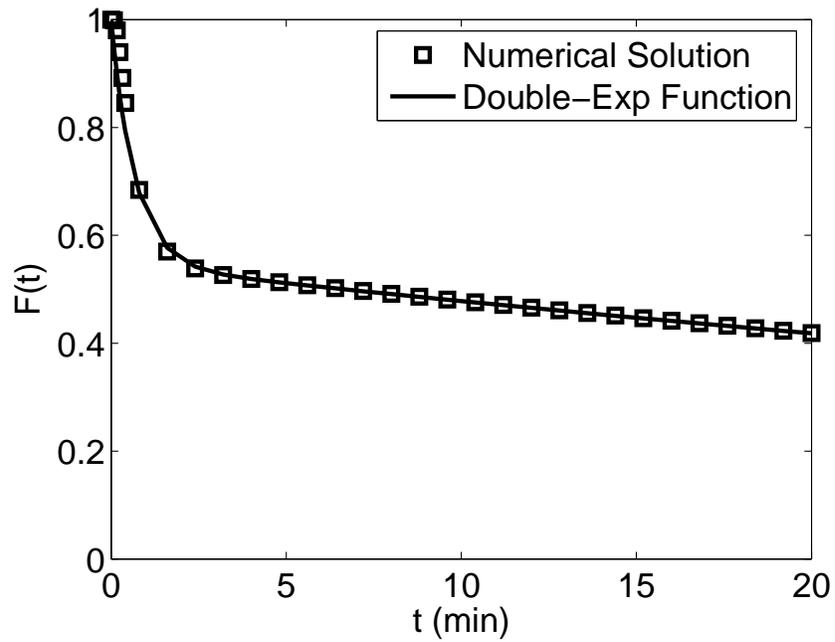


Figure 3.4: Alive probability function for the case with bifurcation with distant peaks. Fitting results:  $c = 0.5464$ ,  $T_1 = 74.8774$ ,  $T_2 = 0.6734$ .

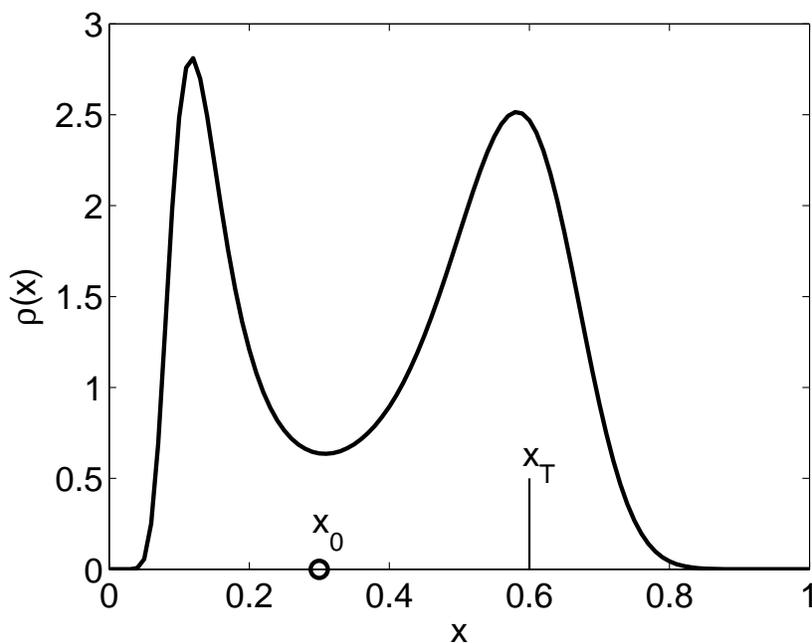


Figure 3.5: Protein number distribution for the case with bifurcation with partially overlapped peaks:  $a_0 = 0.08$ ,  $b = 0.28$ ,  $\kappa = 500$ ,  $m_o = 500$ ,  $\delta = 1$ , (Data from [12]) and  $x_0 = 0.3$ ,  $x_T = 0.6$ .

nearly 6 times  $T_2$ , while in the cases with significant bifurcation,  $T_1$  is more than one hundred times  $T_2$ . Therefore when the two peaks are not so far away from each other, i.e., the barrier between the peaks is not so strong, the ratio of  $T_1$  to  $T_2$  becomes relatively small. When the two peaks are so close that they almost totally overlap, that is, almost no barrier between the two peaks exists,  $T_1$  and  $T_2$  will be almost the same. In other words, the distribution becomes monostable and  $F(t)$  function takes a single-exponential form instead of double-exponential.

**No bifurcation and threshold point is in the peak area.** When  $a_0 = 0.02$ ,  $b = 0.1$ ,  $\kappa = 64$ ,  $m_o = 400$ , as shown in Fig. 3.3.1, the steady state protein number distribution has only one peak, which means at that time the network does not have bifurcation. Suppose  $x_0 = 0.4$  and  $x_T = 0.8$ , i.e., initial protein number is

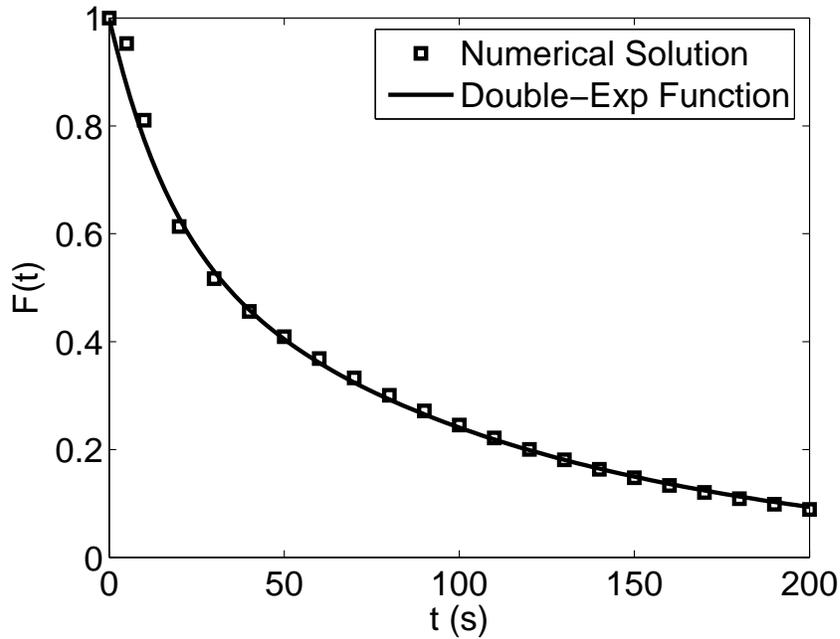


Figure 3.6: Alive probability function from Fokker-Planck equation for the case with bifurcation with partially overlapped peaks. Fitting results:  $c = 0.6139$ ,  $T_1 = 106.2460$ ,  $T_2 = 17.1727$ .

160 and the threshold is 320. Here  $x_T$  is in the peak and  $x_0$  is outside the peak area, so this case can be considered as one where the protein number is trying to reach the peak from outside. Alive probability function is shown in Fig. 3.8, in which  $F(t)$  quickly decreases to zero. The results of parameter fitting are:  $c = 0.0453$ ,  $T_1 = 16.5336$  and  $T_2 = 1.0782$ . Here  $c$  is so small that the  $T_1$  term can be ignored. Therefore in this case alive probability function  $F(t)$  can be considered as a single-exponential function.

**No bifurcation and initial point is in the peak area.** When  $a_0 = 0.02$ ,  $b = 0.25$ ,  $\kappa = 64$ ,  $m_o = 400$ , the distribution of protein number is also monostable as shown in Fig. 3.9. Set  $x_0 = 0.025$  and  $x_T = 0.1$ , which means the initial protein number is 10 and the threshold is 40. In this case  $x_0$  is in the peak area and  $x_T$  is

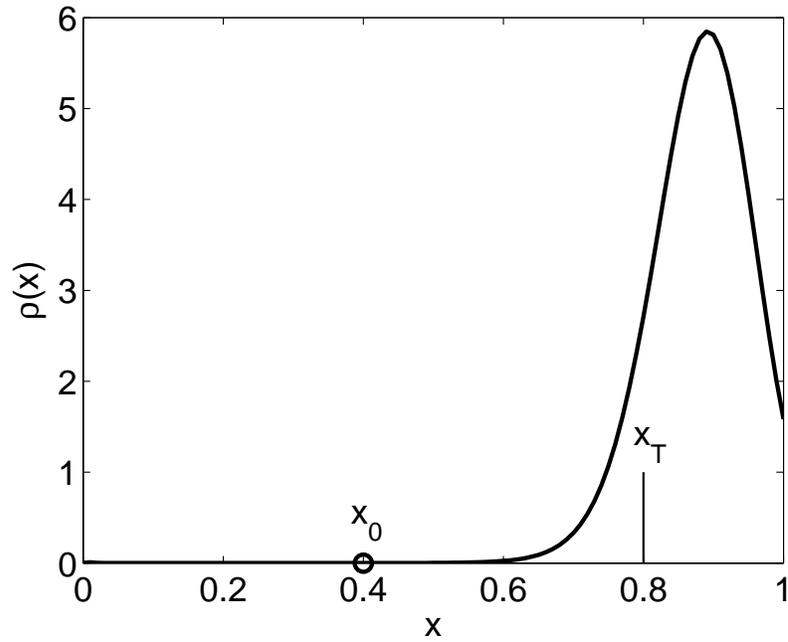


Figure 3.7: Protein number distribution for the case without bifurcation and threshold point is in the peak:  $a_0 = 0.02$ ,  $b = 0.1$ ,  $\kappa = 64$ ,  $m_o = 400$ ,  $\delta = 2.5$ , and  $x_0 = 0.4$ ,  $x_T = 0.8$ .

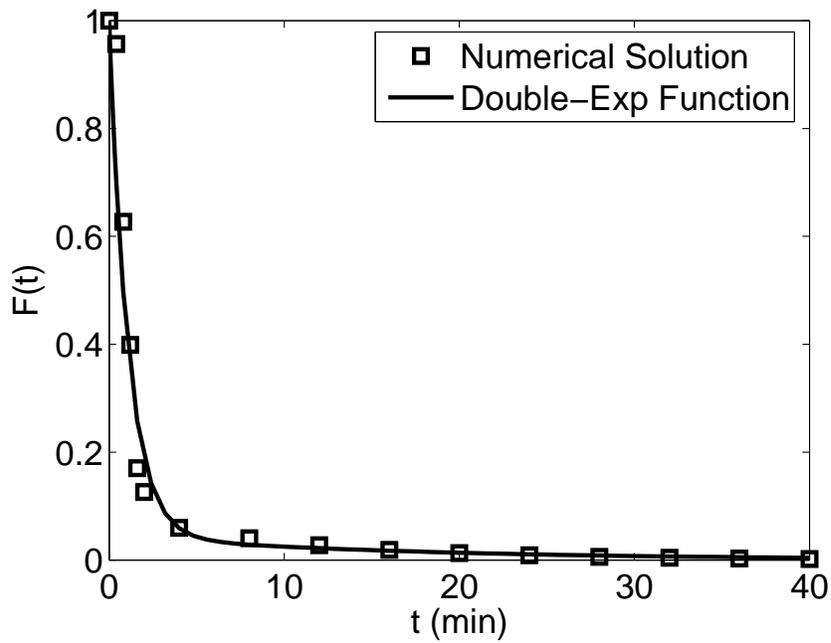


Figure 3.8: Alive probability function from Fokker-Planck equation for the case without bifurcation and threshold point is in the peak. Fitting results:  $c = 0.0453$ ,  $T_1 = 16.5336$ ,  $T_2 = 1.0782$ .

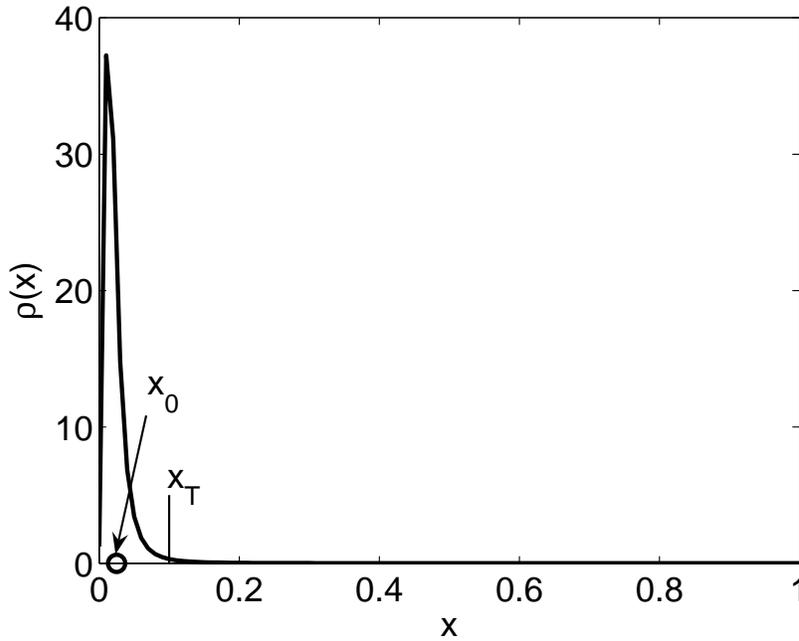


Figure 3.9: Protein number distribution for the case without bifurcation and start point is in the peak:  $a_0 = 0.02$ ,  $b = 0.25$ ,  $\kappa = 64$ ,  $m_o = 400$ ,  $\delta = 2.5$ , and  $x_0 = 0.025$ ,  $x_T = 0.1$ .

outside, so it can be considered that the protein number is trying to escape from the peak area. Then as shown in Fig. 3.10, alive probability function is obtained. Here  $F(t)$  decreases very slowly from one. The fitting results are:  $c = 1.0000$ ,  $T_1 = 19.2618$  and  $T_2 = 1.0037$ . Because  $c$  is equal to 1, the term of  $T_2$  is zero and thus can be cancelled. Therefore in this case  $F(t)$  is also a single-exponential function.

**No bifurcation but with double-exponential function.** Suppose the protein number distribution in Fig. 3.11 is the same as that in Fig. 3.9. But we change the initial protein number and the threshold so that they are 40 and 80 respectively. Therefore  $x_0 = 0.1$  and  $x_T = 0.2$ , both of which are to the right of the peak. In this case, the fitting results is:  $c = 0.8587$ ,  $T_1 = 138.9259$  and  $T_2 =$

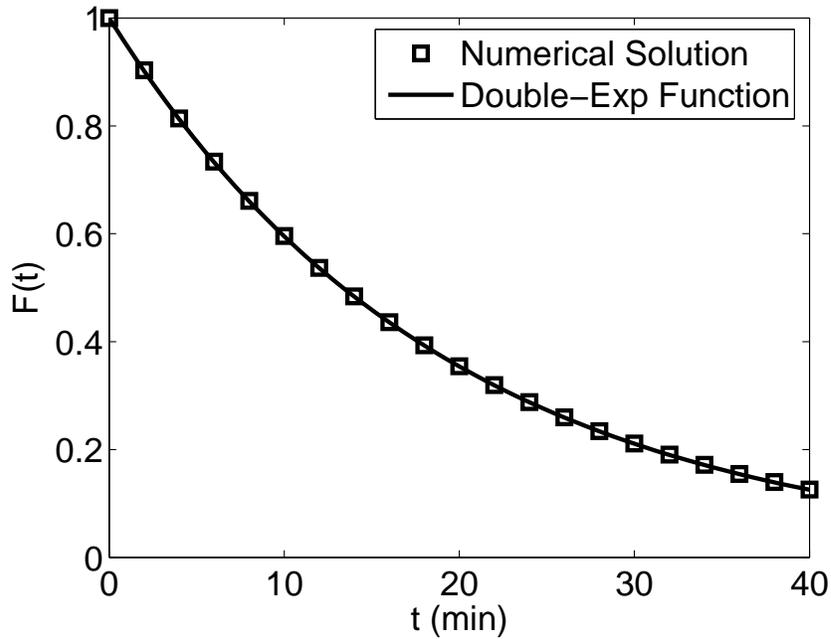


Figure 3.10: Alive probability function from Fokker-Planck equation for the case without bifurcation and start point is in the peak. Fitting results:  $c = 1.0000$ ,  $T_1 = 19.2618$ ,  $T_2 = 1.0037$ .

0.3907, then as shown in Fig. 3.12, alive probability function  $F(t)$  is also a double-exponential function. For this case  $F(t)$  is in the double-exponential form, although protein number distribution is monostable. Let us increase the threshold. Then after fitting, for this monostable case,  $T_1$  increases significantly. But for bistable cases, when threshold increases,  $T_1$  almost remains the same value as long as the threshold is inside the right peak area or between the two peaks. This characteristic distinguishes this case from the real bifurcation cases.

### 3.3.2 Results from Stochastic Petri Nets

Based on the network in Fig. 3.2, an SPN is set up as shown in Fig. 3.13. In this SPN “SdiA” place represents the key protein and the places introduced before

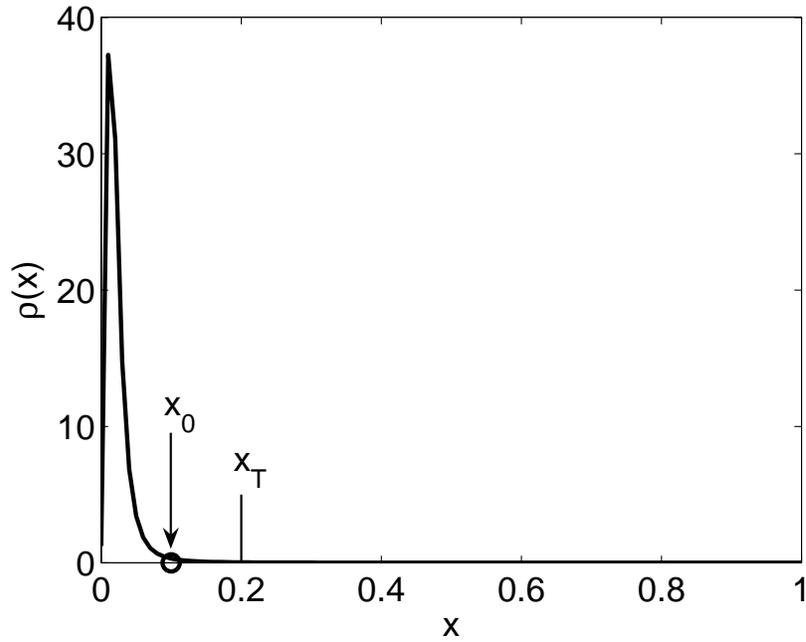


Figure 3.11: Protein number distribution for the case without bifurcation but with double-exponential function:  $a_0 = 0.02$ ,  $b = 0.25$ ,  $\kappa = 64$ ,  $m_o = 400$ ,  $\delta = 2.5$ , and  $x_0 = 0.1$ ,  $x_T = 0.2$ .

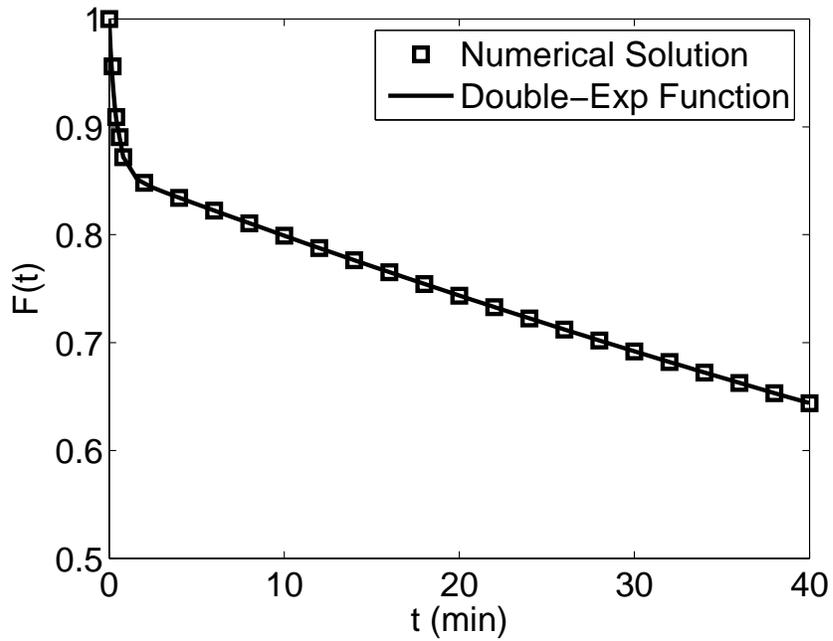


Figure 3.12: Alive probability function from Fokker-Planck equation for the case without bifurcation but with double-exponential function. Fitting results:  $c = 0.8587$ ,  $T_1 = 138.9259$ ,  $T_2 = 0.3907$ .

can be seen. The parameter values in the networks in Fig. 3.13 are the same as the SPN simulated in previous work. The initial protein number and the threshold are set to be 100 and 200 respectively. With the method by Kepler and Elston [12], these parameter values can be found to correspond to the parameters in the case distant-peak bifurcation (shown in Figs. 3.3 and 3.4). Figure 3.14 shows the result of protein number distribution 50 minutes after cells are born, which is obtained previously. It can be found that after 50 minutes (close to the mean generation time of *E.coli*), the distribution is close to steady state (solid curve obtained from the result in Fig. 3.3) and has a bistability. The alive probability function obtained from SPN simulation are shown in Fig. 3.15. (Only selected points from SPN simulation are used because all the points will cover the continuous curves.) We notice that the function  $F(t)$  from SPN can also be written as a double-exponential function. Therefore the conclusion that the bifurcation of key protein number results in double-exponential form of  $F(t)$  applies not only to the result of Fokker-Planck equation, but also to that of SPN simulation. The values for the parameter  $c$  from these two methods are very close, which indicates that with the two methods the results unrelated to time scale can be very close. However the values for the parameters  $T_1$  and  $T_2$  are different. And for SPN simulation, one more parameter, the time delay  $\tau$ , is added to obtain a good fitting. So the double exponential function is:

$$\begin{aligned}
 F(t) &= 1, \quad t < \tau \\
 F(t) &= c \cdot \exp\left(-\frac{t-\tau}{T_1}\right) + (1-c) \cdot \exp\left(-\frac{t-\tau}{T_2}\right), \quad T_1 > T_2, \quad t \geq \tau \quad (3.12)
 \end{aligned}$$

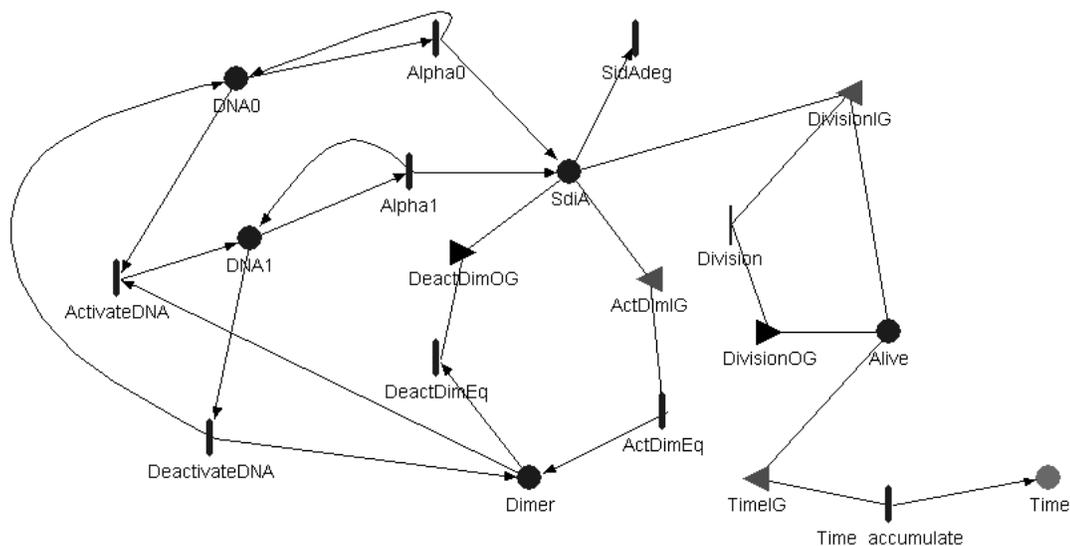


Figure 3.13: SPN to obtain alive probability function of self-promotion network.

$F(t)$  from Fokker-Planck equation drops faster than that from SPN. This is because of the approximations made to derive Fokker-Planck equation. For example, the fast-transition approximation ignores the time used to reach the equilibrium of reversible reactions [12].

The second case for SPN is the signal-mediated self-promotion network as shown in Fig. 3.16. Here we are still discussing a general case although we also use SdiA as an example. In this network there is still a mechanism of positive feedback, which makes bifurcation possible. However this mechanism takes effect through signal molecules (AI-2 molecule in Fig. 3.16). The protein needs to bind the signal molecule first, and then form a dimer and activate its own expression. Moreover, this network is more detailed than the self-promotion network we discussed earlier. In this network, the steps of transcription and translation are separate, while the

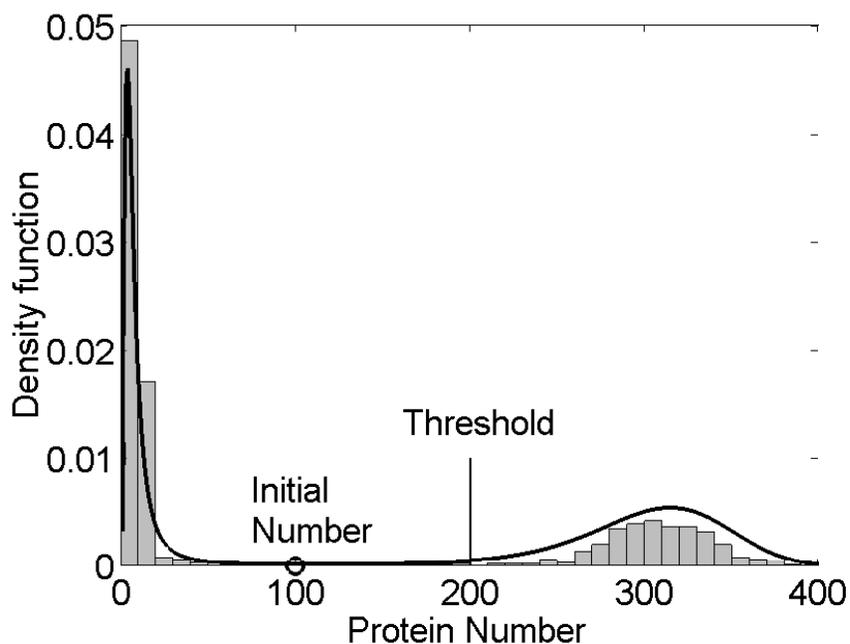


Figure 3.14: Protein number distribution of self-promotion network from Stochastic Petri nets (SPN). The solid curve is the steady state distribution. The bargraph is the distribution of cells at the age of 50 minutes, which is from SPN simulation.

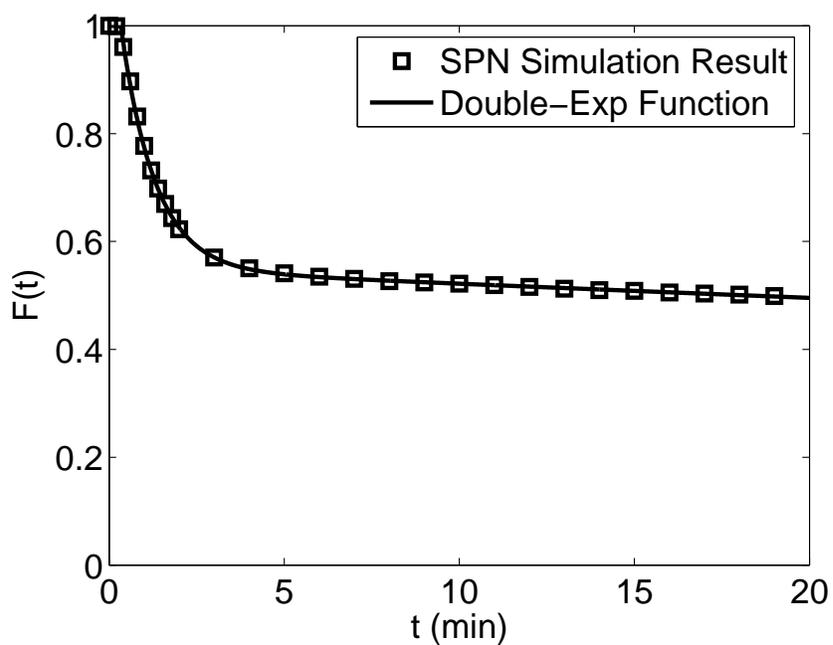


Figure 3.15: Alive probability function of self-promotion network from Stochastic Petri nets (SPN). Fitting results:  $c = 0.5484$ ,  $T_1 = 193.5935$ ,  $T_2 = 0.9782$ ,  $\tau = 0.3323$ .

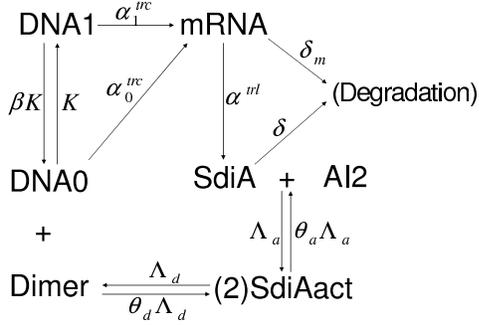


Figure 3.16: Signal-mediated self-promotion network.

steps merge into one expression step in the self-promotion network. Therefore, the behavior of the signal-mediate self-promotion network may be different from the self-promotion network.

The function  $F(t)$  for this network cannot be easily obtained with Fokker-Planck equation. So we can only use SPN to obtain its alive probability function. The SPN for signal-mediated self-promotion network is shown in Fig. 3.17. In this SPN, “SdiA” place also represents key protein and the same places are used to obtain  $F(t)$  as in the SPN in Fig. 3.13. The parameter values used in this SPN are the same as that in previous work.

The results of SPN simulation are shown in Fig. 3.18 and Fig. 3.19. Figure 3.18 shows the protein number distribution after 50 minutes. We notice that the protein number distribution is also bistable. In Fig. 3.19 we see the alive probability function. The function  $F(t)$  here takes a double-exponential form again. But in this double-exponential function, the values for the time-related parameters  $T_1$ ,  $T_2$  and  $\tau$  are larger than both Fokker-Planck equation and SPN simulation results for the self-promotion network.  $F(t)$  here drops slower than for the self-promotion network

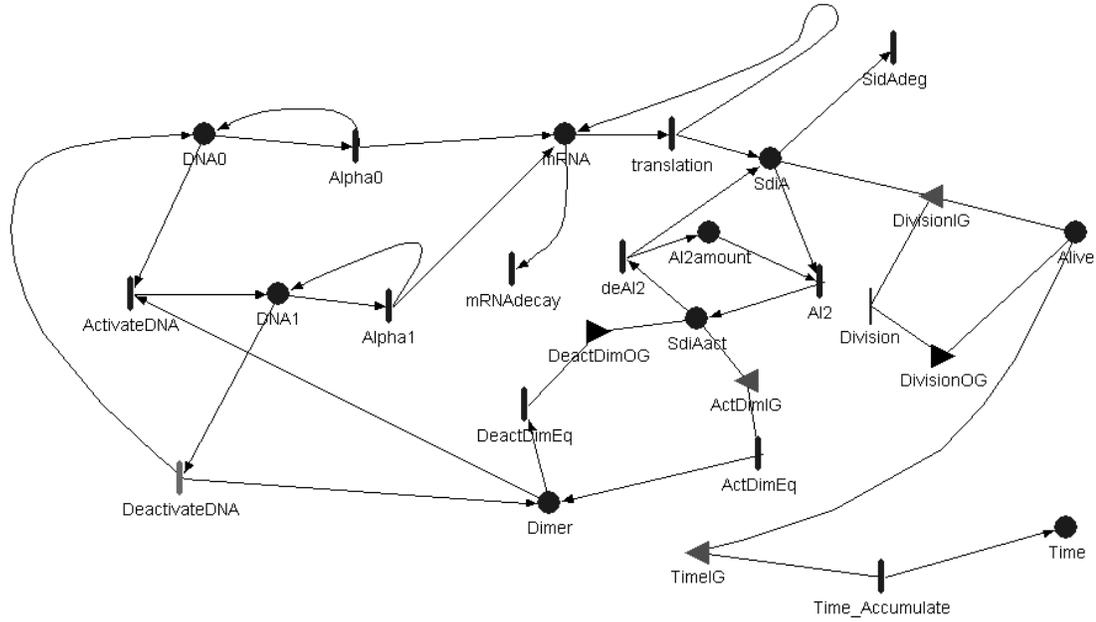


Figure 3.17: SPN to obtain alive probability function of signal-mediated self-promotion network.

because of the assumptions made for the simplification of the network in Fig. 3.16 into the self-promotion network. For example, one assumption ignores the time needed for the mRNA number to reach steady state.

The alive probability functions for both networks above can be written in the double-exponential form, which suggests that this form can be of wide applicability. Let us summarize some common properties of the two networks, which contribute to  $F(t)$  being a double-exponential function. First, a key protein regulates cell division through a threshold mechanism. Second, the processes in the network are considered to be stochastic Markov processes, which can lead to the loss of synchronization. Third, the key protein number has a bifurcation, which comes from the positive feedback in each of the two networks.

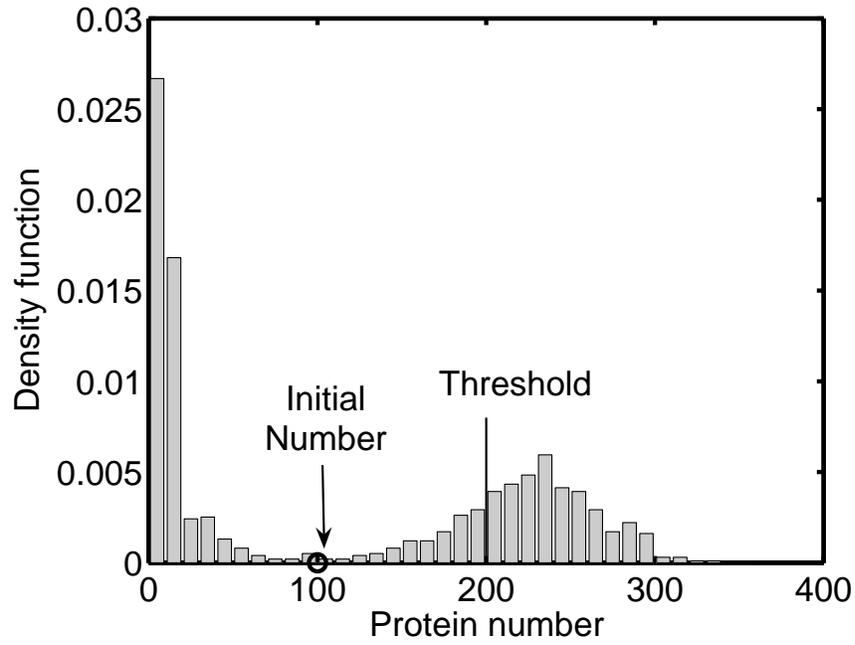


Figure 3.18: Protein number distribution of signal-mediated self-promotion network from Stochastic Petri nets (SPN).

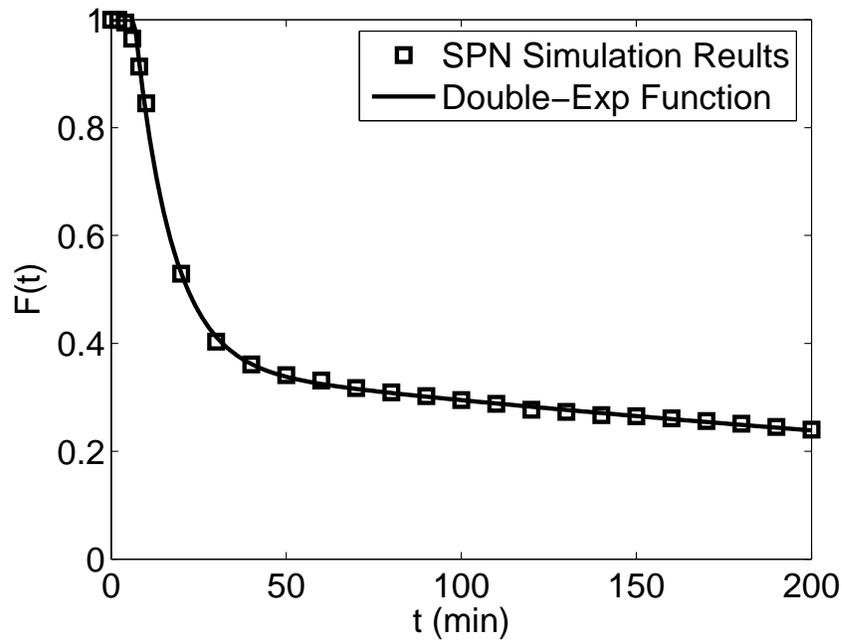


Figure 3.19: Alive probability function of signal-mediated self-promotion network from Stochastic Petri nets (SPN). Fitting results:  $c = 0.3588$ ,  $T_1 = 474.0490$ ,  $T_2 = 10.5924$ ,  $\tau = 6.6477$ .

### 3.3.3 Explanation of the Double-exponential Form in Relation to Bifurcation of Key Protein Number

Based on the results for the two networks we studied, we may conclude that when steady state distribution of a key protein for division is bistable, then the alive probability function  $F(t)$  is a double-exponential function; while when the protein distribution is monostable,  $F(t)$  is a single-exponential function. It is required that in the bistable cases,  $x_0$  should be between the two peaks and  $x_T$  should be either inside the right peak area or between two peaks. However, a special case of monostable distribution also shows double-exponential  $F(t)$ . We also discussed that the double-exponential form for alive probability function is not limited to a specific method or to a specific network.

Although we have used the double-exponential form in several different cases, we have not yet discussed a theoretical reason for the form. Here we propose that the bifurcation of key protein numbers of the cell population in a homogenous culture can explain the double-exponential form of  $F(t)$ . The two peaks in the steady state distribution of protein number actually correspond to two directions that the cells may take. If the initial protein number is between the two peaks, there will be a bifurcation during the evolution of the protein number. In some cells, protein number evolves to the right peak and therefore reaches the threshold quickly. So these cells will divide quickly. In other cells, protein number evolves to the left peak, and after this happens the protein number is liable to stay there, therefore making it hard to reach the threshold. So it would take very long time for cells in

$x_0$	$c$	$T_1$	$T_2$
0	1.0000	78.1322	–
0.05	1.0000	74.0013	–
0.1	0.9272	74.0011	1.0049
0.15	0.8157	74.9209	0.9889
0.2	0.6900	74.0140	0.7948
0.25	0.5464	74.8774	0.6734
0.3	0.4052	74.7216	0.5477
0.35	0.2737	74.4318	0.4350
0.4	0.1619	72.2416	0.3319
0.45	0.0679	74.2357	0.2487
0.5	0	-	0

Table 3.1: Parameter Values of Double-exponential Function.

this group to divide. In this way, the bifurcation of protein number causes  $F(t)$  to take a double-exponential shape.

Based on the explanation above, we can discuss the physical meanings of the parameters. Consider the bifurcation case with distant peaks (Figs. 3.3 and 3.4) and hold the threshold  $x_T = 0.5$ . Then let us change the initial point  $x_0$  and fit the parameter values of the double-exponential function. The results are shown in Table 3.1.

**Physical meaning of  $c$ .**  $c$  denotes the fraction of cells which go to the left peak. From Fig. 3.20, we could see that as  $x_0$  increases,  $c$  decreases. This means

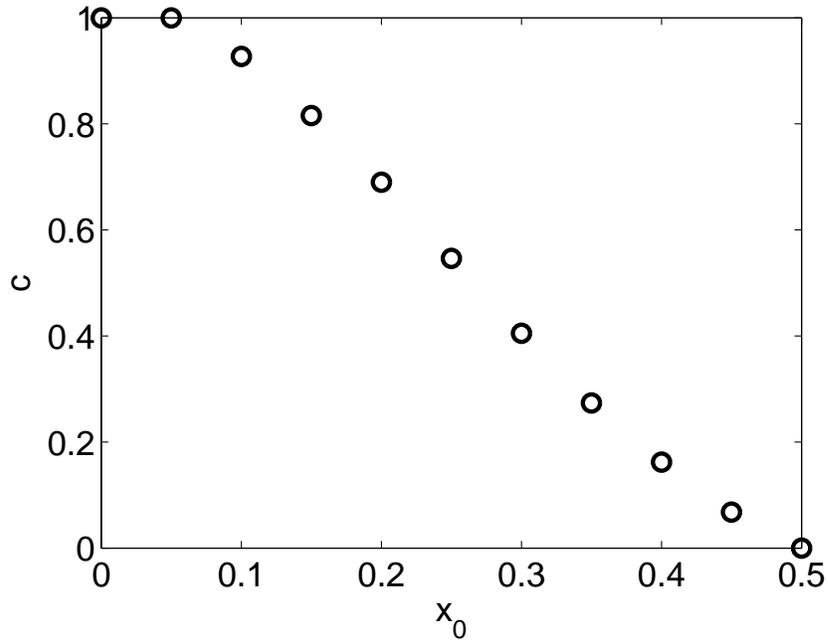


Figure 3.20:  $c$  vs  $x_0$  curve.

that as  $x_0$  goes up, SdiA number is more likely to directly go to the right peak. When  $x_0$  is around the left peak,  $c$  is equal to 1. When  $x_0$  is equal to  $x_T$ ,  $c$  is equal to 0, which means in all the cells, the protein number goes across the threshold quickly, so all the cells divide at once. It could also be noted that if  $x_0$  increases from 0.1 to 0.4, the parameter  $c$  goes down almost linearly (Fig. 3.20).

**Physical meaning of  $T_1$ .**  $T_1$  is essentially the mean time it takes the protein number to reach the threshold for cells which go to the left peak. For these cells, protein number actually goes to left peak very fast so that the values for  $T_1$  should nearly be equal to the mean first passage time (MFPT), i.e., the average time length for protein number to escape from the left-peak value. So this MFPT is independent of the initial value  $x_0$  and the  $T_1$  values in Table 3.1 are almost constant. MFPT

$T(x)$  can be obtained by solving the following second-order ODE [10]:

$$A(x)\frac{dT}{dx} + \frac{1}{2}B(x)\frac{d^2T}{dx^2} = -1 \quad (3.13)$$

with the boundary conditions:

$$\frac{dT}{dx}(0) = 0 \quad (3.14)$$

$$T(x_T) = 0 \quad (3.15)$$

Then  $T(x_0)$  is the MFPT from  $x_0$  to the threshold  $x_T$ . Use the summit value of left peak  $x_0 = 0.005$  and then MFPT  $T(x_0) = 79.1$ . Use  $x_0 = 0.075$ , which corresponds to a point at the edge of the left peak and then MFPT  $T(x_0) = 74.4$ . Both of the values are close to the  $T_1$  in Table 3.1. Therefore the physical meaning of  $T_1$  we propose is reasonable.

**Physical meaning of  $T_2$ .**  $T_2$  is the mean time for protein number to reach threshold for the cells which go to the right peak. Therefore, as  $x_0$  goes up, the initial protein number becomes closer to the threshold and  $T_2$  becomes the smaller. The ratio of  $T_1$  to  $T_2$  can be used to describe the intensity of the barrier between the two peaks. That  $T_1/T_2$  is close to 1 means that the barrier is weak and cells can easily transit from one peak to the other.

## 3.4 Methods for Obtaining Age Distribution

### 3.4.1 Generation Time Distribution

Next we shall have to calculate the age distribution from the alive probability function. Generation time distribution  $f(t)$  is necessary in this step. Its definition

is that, in a very small time interval  $[t, t + \Delta t]$ , the probability that the cell divides is  $f(t)\Delta t$ . Generation time distribution has been proposed and used to obtain age distribution since very early [72, 74].  $f(t)$  has also been called “life-length distribution” and “interdivision time distribution” [75].

Alive probability function  $F(t)$  and generation time distribution  $f(t)$  have the following relations:

$$f(t) = -\frac{dF(t)}{dt} \quad (3.16)$$

$$F(t) = 1 - \int_0^t f(t')dt' \quad (3.17)$$

From equations (3.16) and (3.17), generation time distribution and alive probability function can be obtained from each other and then used to calculate the age distribution.

### 3.4.2 Total-Based Method to Obtain Age Distribution

Total-based method considers all the cells as a whole. With generation time distribution, Powell [74] derived age distribution for balanced growth (steady state). For dynamic growth, a population balance model can be used to obtain age distribution. Population balance model is a mathematical model which considers the segregation of bacteria. This model is widely used and can also provide mass distribution [83]. However it is very difficult to solve so that its solution is usually numerical [84]. An analytical solution is possible but it is very complex to obtain [77]. Here we propose a new method for an analytical solution, utilizing the Laplace Transform.

The population balance model is a partial differential equation (PDE) [77]:

$$\frac{\partial n(t, a)}{\partial t} + \frac{\partial n(t, a)}{\partial a} = -\Gamma(a)n(t, a) \quad (3.18)$$

where  $t$  is time,  $a$  is age.  $n(t, a)$  is unnormalized age distribution, which means at the time  $t$  the number of cells with age between  $[a, a + \Delta a]$  is  $n(t, a)\Delta a$ .  $\Gamma(a)$  is the division rate density function and equal to:

$$\Gamma(a) = \frac{f(a)}{F(a)} \quad (3.19)$$

The boundary condition is:

$$n(t, 0) = 2 \int_0^\infty \Gamma(a)n(t, a)da \quad (3.20)$$

The initial condition is:

$$n(0, a) = n_0(a) \quad (3.21)$$

Equation (3.18) is difficult to solve, not only because it is a PDE but also because its boundary condition has an integral. To simplify it, we define a production function  $P(t)$ , which means that during a time interval after time zero,  $[t, t + \Delta t]$ , the number of newly born cells is  $P(t)\Delta t$ . Therefore by definition we have:

$$n(t, a) = P(t - a)F(a) \quad (3.22)$$

Use (3.22) into (3.18), left side is automatically equal to right side. Use (3.22) into boundary condition (3.20), we obtain an equation containing  $f(t)$  and  $P(t)$ :

$$P(t) = 2 \int_0^\infty P(t - a)f(a)da \quad (3.23)$$

The equation (3.23) was also obtained by Ramkrishna [85] through a very complicated mathematical derivation. In our approach, the functions  $f(t)$  and  $P(t)$  have clear biological meaning making the method easier to understand.

Compared with the PDE (3.18), equation (3.23) seems to be relatively easier to solve because it only contains single-variable functions. However the integral in (3.23) still makes the equation hard to solve. Here we apply the Laplace Transform to obtain  $P(t)$  from  $f(t)$ .

Before the Laplace Transform is used, we need to assume that at the very beginning ( $t = 0$ ), the cells of the first generation are just born, that is to say,  $n_0(a) = N_0\delta(0)$ . So the production function at  $t = 0$  is:

$$P(0) = N_0\delta(0) \tag{3.24}$$

where  $N_0$  is the original number of first-generation cells.

Use (3.24) into (3.22), then the age distribution becomes:

$$\begin{aligned} n(t, a) &= P(t - a)F(a), \quad t > a \\ n(t, a) &= N_0F(a)\delta(a - t), \quad t = a \end{aligned} \tag{3.25}$$

Use (3.24) into (3.23) to set:

$$P(t) = 2N_0f(t) + 2 \int_0^{t-} P(t - a)f(a)da, \quad t > 0 \tag{3.26}$$

Then let us use Laplace Transform. First we need to get the Laplace Transform of  $f(t)$ :

$$f^L(s) = \mathcal{L}[f(t)] \tag{3.27}$$

Then get the transform of  $P(t)$ :

$$P^L(s) = \frac{2N_0 f^L(s)}{1 - 2f^L(s)} \quad (3.28)$$

Equation (3.28) is obtained by taking Laplace Transform on both sides of (3.26). With this formula,  $P(t)$  can be obtained:

$$P(t) = \mathcal{L}^{-1}[P^L(s)] \quad (3.29)$$

In summary, we have developed a new way to obtain an analytical form of the age distribution  $n(t, a)$  from  $f(t)$ . The steps are: from  $F(t)$  to  $f(t)$  with (3.16); from  $f(t)$  to  $P(t)$  with (3.27), (3.28) and (3.29); from  $P(t)$  to  $n(t, a)$  with (3.25).

Actually in our work, we are also interested in the fraction of the first-generation cells in a cell population rather than just the age distribution of the population. The total cell number of a population is the integration of age distribution:

$$N(t) = \int_0^t n(t, a) da \quad (3.30)$$

Use (3.25) in (3.30) to obtain:

$$N(t) = N_0 F(t) + \int_0^{t-} P(t-a) F(a) da \quad (3.31)$$

To avoid integration, we can also use the Laplace Transform:

$$N^L(s) = N_0 F^L(s) + P^L(s) F^L(s) = F^L(s)(N_0 + P^L(s)) \quad (3.32)$$

With the equation above, similarly to the way for obtaining  $P(t)$ , we can obtain  $P^L(s)$ , and then  $N^L(s)$ , and then  $N(t)$ . The number of first-generation cells is  $N_0 F(t)$ . So the fraction of first-generation cells in a cell population,  $R(t)$ , is:

$$R(t) = \frac{N_0 F(t)}{N(t)} \quad (3.33)$$

We can summarize the steps to obtain  $R(t)$ : from  $F(t)$  to  $f(t)$  with (3.16); from  $f(t)$  to  $P(t)$  with (3.27), (3.28) and (3.29); from  $P(t)$  to  $N(t)$  with (3.31); from  $N(t)$  to  $R(t)$  with (3.33).  $R(t)$  may be possible to measure by experiments.

### 3.4.3 Generation-Based Method to Obtain Age Distribution

Generation-based method obtains the cell age distribution for each generation and then integrates the results for each generation into an overall result. Liou *et al.* [77] proposed successive generations approach to obtain the analytical age distribution of each generation. However the solutions by this approach contain repeated integrations and are very complex. Also in this approach age distributions are obtained generation by generation, and thus, to obtain the age distribution of one generation, age distributions of all previous generations need to be calculated. So we will attempt to reduce the intensive computation required by this approach.

The population balance equations for age distribution of each generation are as follows. First we have the PDE for first generation:

$$\frac{\partial n_1(t, a)}{\partial t} + \frac{\partial n_1(t, a)}{\partial a} = -\Gamma(a)n_1(t, a) \quad (3.34)$$

with boundary condition and initial condition:

$$n_1(t, 0) = 0 \quad (3.35)$$

$$n_1(0, a) = n_0(a) \quad (3.36)$$

And then for the other generations:

$$\frac{\partial n_k(t, a)}{\partial t} + \frac{\partial n_k(t, a)}{\partial a} = -\Gamma(a)n_k(t, a), \quad k \geq 2 \quad (3.37)$$

where boundary and initial conditions:

$$n_k(t, 0) = 2 \int_0^\infty \Gamma(a)n_{k-1}(t, a)da \quad (3.38)$$

$$n(0, a) = 0 \quad (3.39)$$

To deal with these PDEs, similarly to the approach in the previous section, we define the production for each generation,  $P_k(t)$ . Then by definition we also have:

$$n_k(t, a) = P_k(t - a)F(a) \quad (3.40)$$

Use (3.40) into (3.36), and we have:

$$\begin{aligned} P_1(t) &= n_0(-t)/F(-t), \quad t \leq 0 \\ P_1(t) &= 0, \quad t > 0 \end{aligned} \quad (3.41)$$

Use (3.40) into (3.38), and we have:

$$P_k(t) = 2 \int_0^\infty P_{k-1}(t - a)f(a)da, \quad k \geq 2 \quad (3.42)$$

To use the Laplace Transform, we need to assume all the first-generation cells are born at the time  $t = 0$ . Then  $n_0(a) = N_0\delta(0)$ . So the production function of the first generation is Dirac delta function:

$$P_1(t) = N_0\delta(0) \quad (3.43)$$

In this way, no cell exists before  $t = 0$ . Then in (3.42), the upper limit of the integration can be changed. Then the equation becomes:

$$P_k(t) = 2 \int_0^t P_{k-1}(t - a)f(a)da, \quad k \geq 2 \quad (3.44)$$

Therefore the Laplace Transforms of the production functions  $P_k(t)$  are:

$$P_1^L(s) = N_0 \quad (3.45)$$

$$P_k^L(s) = 2P_{k-1}^L(s)f^L(s), \quad k \geq 2 \quad (3.46)$$

With (3.45) and (3.46) the Laplace Transform of production function can be written as:

$$P_k^L(s) = N_0[2f^L(s)]^{k-1} \quad (3.47)$$

In this way, from  $F(t)$  and  $f(t)$ , we can directly obtain  $P_k(t)$  with (3.47), and then obtain age distribution for the  $k$ th generation with (3.40). We do not need to calculate the age distributions generation by generation.

We are also interested in the fraction of first-generation cells. We begin with the cell number of each generation:

$$N_k(t) = \int_0^\infty n_k(t, a)da = \int_0^t P_k(t-a)F(a)da \quad (3.48)$$

The upper limit is  $t$  because no cell exists before  $t = 0$  for the  $P_1(t)$  we assume in (3.43). Then the Laplace Transform is:

$$N_k^L(s) = P_k^L(s)F^L(s) \quad (3.49)$$

From (3.17) we know the relation between the Laplace Transform of  $f(t)$  and  $F(t)$ :

$$F^L(s) = \frac{1}{s} - \frac{f^L(s)}{s} \quad (3.50)$$

Plug equations (3.47) and (3.50) into (3.49), and then obtain the Laplace Transform of  $N_k(t)$ :

$$N_k^L(s) = N_0[2f^L(s)]^{k-1}\left(\frac{1}{s} - \frac{f^L(s)}{s}\right) = \frac{N_0}{s}[(2f^L(s))^{k-1} - \frac{1}{2}(2f^L(s))^k], \quad k < K \quad (3.51)$$

where  $K$  denotes the latest generation, which is excluded. For this generation, the cells have not divided. So instead of (3.50), we can assume that for those cells  $F(t) = 1$  and thus  $F^L(s) = 1/s$ . Then from (3.47) and (3.49), the cell number can be represented by:

$$N_K^L(s) = \frac{P_k^L(s)}{s} = \frac{N_0}{s} [2f^L(s)]^{K-1} \quad (3.52)$$

Then from (3.51) and (3.52), the Laplace Transform of total cell number is:

$$N^L(s) = \sum_{k=1}^K N_k^L(s) = \frac{N_0}{s} + \frac{N_0}{2s} \sum_{k=2}^K [2f^L(s)]^{k-1} \quad (3.53)$$

First-generation cell number is a special case for (3.51):

$$N_1^L(s) = \frac{N_0}{s} [1 - f^L(s)] = N_0 F^L(s) \quad (3.54)$$

So the number of first generation is:

$$N_1(t) = N_0 F(t) \quad (3.55)$$

Finally, from  $f(t)$ , we can obtain  $N_1(t)$  with (3.55), and  $N(t)$  with (3.53). The fraction of first-generation cells is:

$$R(t) = \frac{N_1(t)}{N(t)} = \frac{N_0 F(t)}{N(t)} \quad (3.56)$$

## 3.5 Results for Age Distribution

### 3.5.1 Case without Time delay

In the previous section, we concluded that if key protein number in cells has bifurcation, alive probability function  $F(t)$  can be written in the double-exponential

form. Here we discuss the age distribution obtained with such an  $F(t)$ . First we discuss a simple case, in which the time delay  $\tau$  is ignored. This case is relatively easier than the case with time delay. So we prefer using the conclusions in this case if the time delay is very small. We choose the parameter values for signal-mediated self-promotion network (Fig. 3.16). In this case the alive probability function is:

$$F^{(0)}(t) = 0.3588\exp(-t/474.0490) + (1 - 0.3588)\exp(-t/10.5924) \quad (3.57)$$

The superscript (0) here means no time delay. Then the Laplace Transform of  $F(t)$  is:

$$F^{(0)L}(s) = \frac{0.6412}{s + 0.09441} + \frac{0.3588}{s + 0.002110} \quad (3.58)$$

So the other Laplace Transforms are:

$$f^{(0)L}(s) = 1 - sF^{(0)L}(s) = 1 - \frac{0.6412s}{s + 0.09441} - \frac{0.3588s}{s + 0.002110} \quad (3.59)$$

$$P^{(0)L}(s) = \frac{2N_0 f^{(0)L}(s)}{1 - 2f^{(0)L}(s)} = \frac{0.01208s + 0.0003984}{s^2 - 0.02428s - 0.0001992} N_0 \quad (3.60)$$

$$N^{(0)L}(s) = F^{(0)L}(s)[P^{(0)L}(s) + N_0] = \frac{s + 0.03522}{s^2 - 0.02428s - 0.0001992} N_0 \quad (3.61)$$

Hence the total cell number is:

$$N^{(0)}(t) = [1.7561\exp(0.03224t) - 0.7561\exp(-0.006177t)]N_0 \quad (3.62)$$

Choose time  $t = 50$ . Then  $N^{(0)}(t) = 8.2478N_0$ ,  $F^{(0)}(t) = 0.3286$ . So the first-generation cell fraction is:

$$R^{(0)}(t) = 4\% \quad (3.63)$$

This number is not large. But experimental techniques can observe less than one percent abnormal cells in a population [73]. So this number may be measurable.

The results above are the results under specific parameter values. Next we will analyze the effect of different parameter values. Double-exponential function is too complex to obtain the functions above related to age distribution. Because the value for  $T_1$  is usually very large, we will assumed as an approximation that  $T_1$  is infinity and the  $T_1$  term in (3.11) is constant. And also let  $\mu = 1/T_2$ . Then the double-exponential function becomes a simplified form of exponential plus constant:

$$F(t) = (1 - c)\exp(-\mu t) + c \quad (3.64)$$

This analytical form of alive probability function can be handled easily. The relation between alive probability function and age distribution is described based on the function (3.64). Plug (3.64) into (3.16), (3.27), (3.28) and (3.32), we have:

$$N^{(0)L}(s) = F^{(0)L}(s)[P^{(0)L}(s) + N_0] = \frac{1 - c}{1 - 2c} \cdot \frac{N_0}{s - \mu(1 - 2c)} - \frac{c}{1 - 2c} \cdot \frac{N_0}{s}, \quad c \neq \frac{1}{2} \quad (3.65)$$

$$N^{(0)L}(s) = F^{(0)L}(s)[P^{(0)L}(s) + N_0] = \frac{N_0}{s} + \frac{\mu N_0}{2s^2}, \quad c = \frac{1}{2} \quad (3.66)$$

So the total cell number is:

$$N^{(0)}(t) = \frac{1 - c}{1 - 2c} N_0 \exp[\mu(1 - 2c)t] - N_0 \frac{c}{1 - 2c}, \quad c \neq \frac{1}{2} \quad (3.67)$$

$$N^{(0)}(t) = N_0 + \frac{\mu N_0 t}{2}, \quad c = \frac{1}{2} \quad (3.68)$$

Figure 3.21 shows the total cell number when  $F(t)$  is in two different forms, with the parameter values in (3.57). We notice that before 60 minutes the difference between the two total number functions are small. So the function (3.64) is a reasonable approximation.

With this alive probability function, the total cell number has three kinds of shapes shown in Figs. 3.22 to 3.24. It is very interesting that under different parameter values the cells show different growth curves. Based on this, the function  $R(t)$  can be discussed. In the situation of  $c > 1/2$ ,  $N(t)$  has an upper limit, so  $R(t) = N_0 F(t)/N(t)$  has a lower limit  $2c - 1$ . That means the fraction of original cells can not go to zero. When  $c$  is not too close to  $1/2$ , the fraction of original cells in the population is always significant:

$$R^{(0)}(t) > 2c - 1, \quad c > \frac{1}{2} \quad (3.69)$$

For the other two situations, we could select time  $t_s$  which is long enough but is still much smaller than  $T_1$ . Then in the time period before  $t_s$ ,  $R(t)$  can not be lower than  $R(t_s)$ :

$$R^{(0)}(t) > R^{(0)}(t_s), \quad c \leq \frac{1}{2} \quad (3.70)$$

Therefore, if the key protein number has a bifurcation, after a relatively long period, the first-generation cells will still occupy a significant fraction in the cell population.

### 3.5.2 Case with Time Delay

In the alive probability function (3.57), the time delay is ignored. However the time delay often has significant effect on the results. This time delay may be related to the time needed by protein number to reach the threshold, or due to time needed by other cellular activities between reaching threshold and cell division. If the time delay is not ignored, the alive probability function for signal-mediated

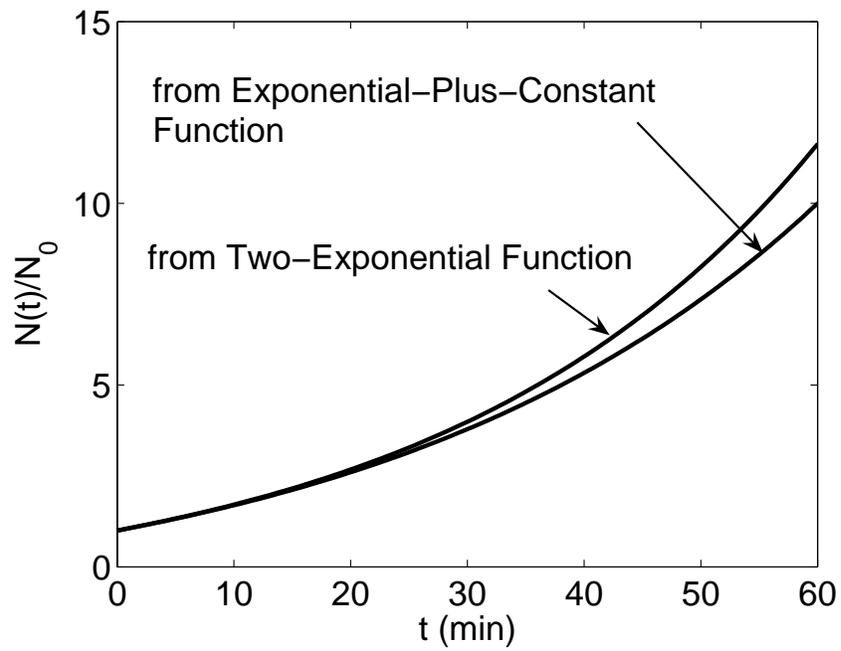


Figure 3.21: Total cell number when alive probability functions  $F(t)$  are in two different forms.

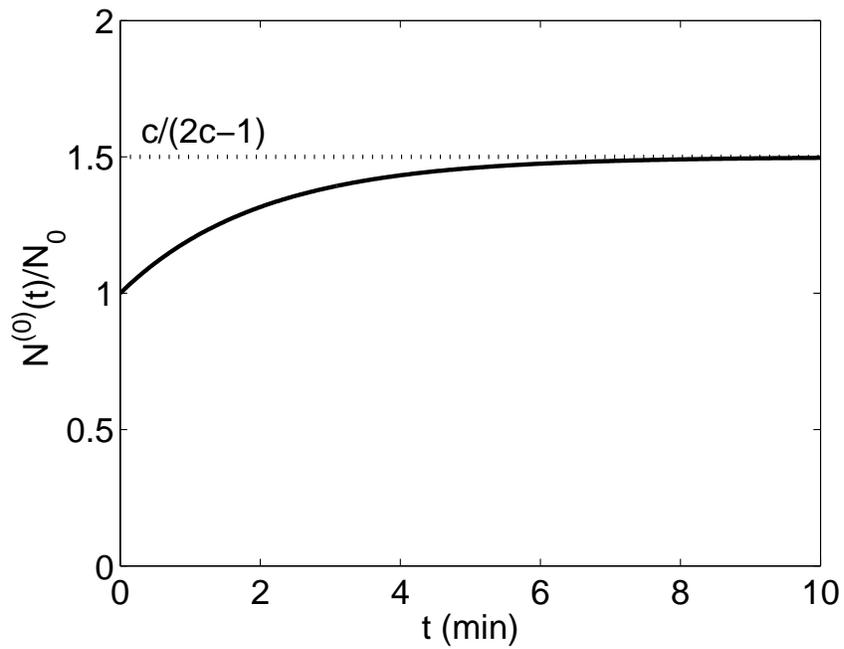


Figure 3.22: The shape of total cell number for the case without time delay when  $\mu = 1$ ,  $c = \frac{3}{4}$ .

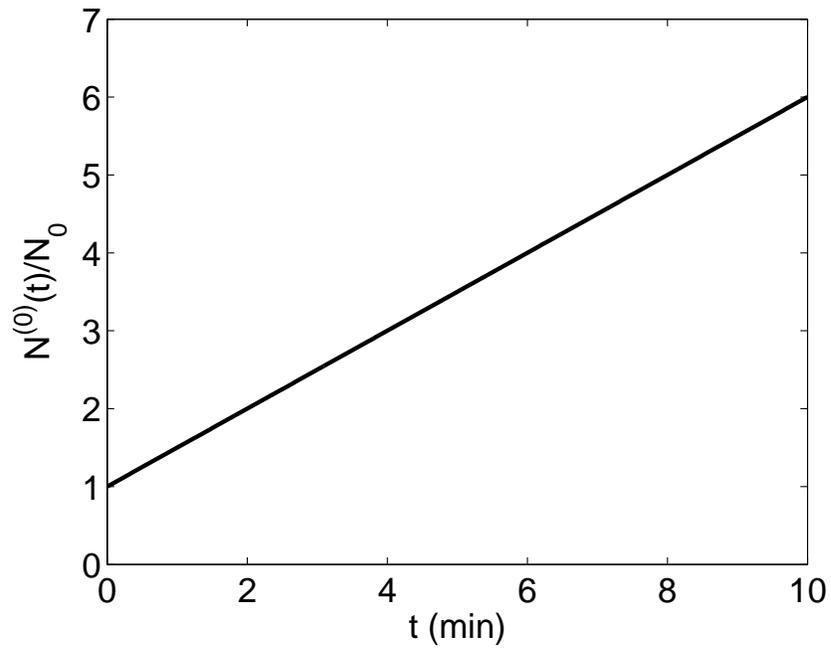


Figure 3.23: The shape of total cell number for the case without time delay when  $\mu = 1$ ,  $c = \frac{1}{2}$ .

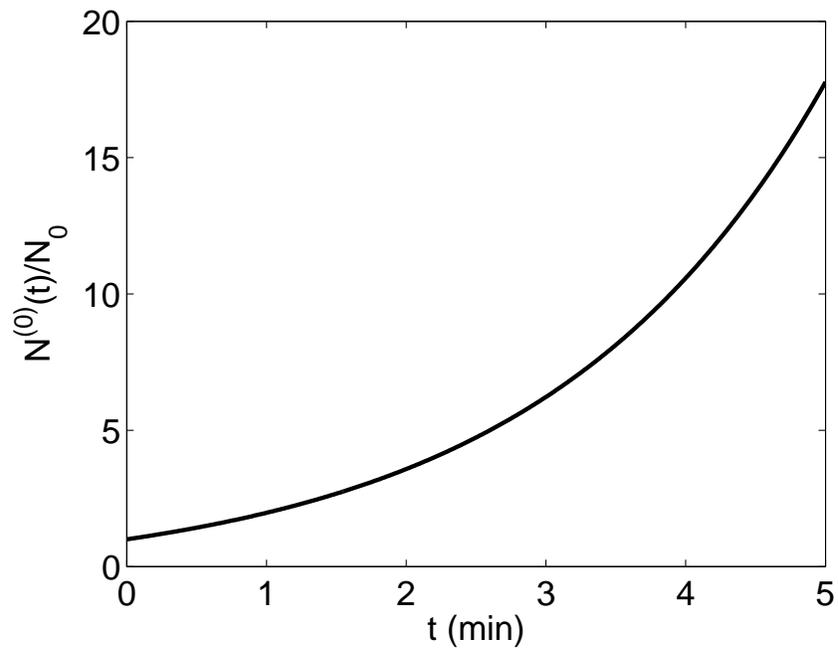


Figure 3.24: The shape of total cell number for the case without time delay when  $\mu = 1$ ,  $c = \frac{1}{4}$ .

self-promotion network is:

$$\begin{aligned}
F^{(\tau)}(t) &= 1, \quad t < 6.6477 \\
&= (1 - 0.3588)\exp[-(t - 6.6477)/10.5924] \\
&\quad + 0.3588\exp[-(t - 6.6477)/474.0490], \quad t \geq 6.6477 \quad (3.71)
\end{aligned}$$

The superscript  $(\tau)$  here means time delay exists. The generation time distribution is:

$$\begin{aligned}
f^{(\tau)}(t) &= 0, \quad t < 6.6477 \\
&= 0.06053\exp[-0.09441(t - 6.6477)] \\
&\quad + 0.0007569\exp[-0.002110(t - 6.6477)], \quad t \geq 6.6477 \quad (3.72)
\end{aligned}$$

When there is a time delay, in order to obtain the age distribution, we have to consider the cells of each generation. We choose the time of 50 minutes. Then the number of generations we need to consider is:

$$K = \text{fix}\left(\frac{50}{6.6477}\right) + 1 = 8 \quad (3.73)$$

where we  $\text{fix}(x)$  returns the integer portion of  $x$ .

The Laplace Transform of  $f(t)$  is:

$$f^{(\tau)L}(s) = \exp(6.6477s) \frac{0.06129s + 0.0001992}{(s + 0.09441)(s + 0.002110)} \quad (3.74)$$

Plug this equation into (3.53), to set the Laplace Transform of total cell number function:

$$N^{(\tau)L}(s) = \frac{N_0}{s} + \frac{N_0}{2} \sum_{k=2}^K \frac{2^{k-1}(0.06129s + 0.0001992)^{k-1}}{s(s + 0.09441)^{k-1}(s + 0.002110)^{k-1}} \exp[(k - 1)6.6477s] \quad (3.75)$$

To obtain the inverse Laplace Transform, we define a function  $Q_k(t)$ , whose Laplace Transform is:

$$Q_k^L(s) = \frac{2^{k-1}(0.06129s + 0.0001992)^{k-1}}{s(s + 0.09441)^{k-1}(s + 0.002110)^{k-1}} \quad (3.76)$$

Then the function  $N^{(\tau)L}(s)$  is:

$$N^{(\tau)L}(s) = N_0 Q_1^L(s) + \frac{N_0}{2} \sum_{k=2}^K Q_k^L(s) \exp[(k-1)6.6477s] \quad (3.77)$$

We can obtain  $Q_k(t)$ , the inverse Laplace Transform of (3.76). But the function  $Q_k(t)$  contains many terms. So here we just plot curves of the functions from  $Q_1(t)$  to  $Q_8(t)$  as shown in Figs. 3.25 to 3.32. The total cell number  $N^{(\tau)}(t)$  is:

$$N^{(\tau)}(t) = N_0 Q_1(t) + \frac{N_0}{2} \sum_{k=2}^8 Q_k[t - (k-1)6.6477] \quad (3.78)$$

The result is shown in Fig 3.33.

At the time  $t = 50\text{min}$ ,  $F^{(\tau)}(t) = 0.3381$  and  $N^{(\tau)}(t) = 3.3439$ . So the fraction of first-generation cells is:

$$R^{(\tau)}(t) = N_0 F^{(\tau)}(t) / N^{(\tau)}(t) = 10.1\% \quad (3.79)$$

Therefore the fraction here is still significant.

### 3.6 Discussion

It is known that self-promotion gene regulation networks may exhibit stochastic bistability. In this work we examine what effect this might have on cell age distribution, including the fraction of first-generation cells in a population, in the case where a protein considered as a key protein for cell division is involved in this

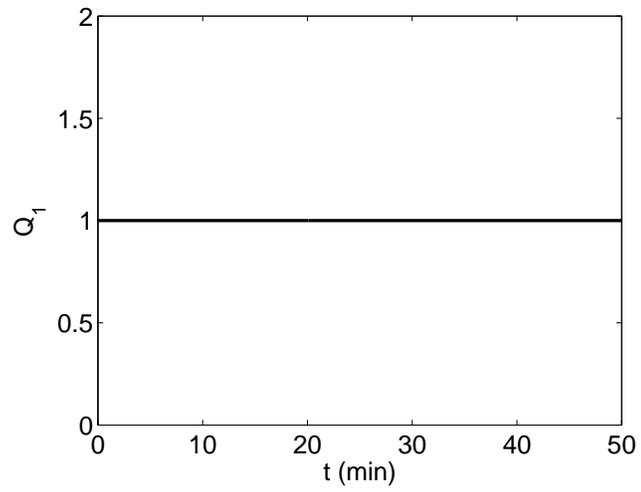


Figure 3.25: The functions  $Q_1(t)$  for the case with time delay.

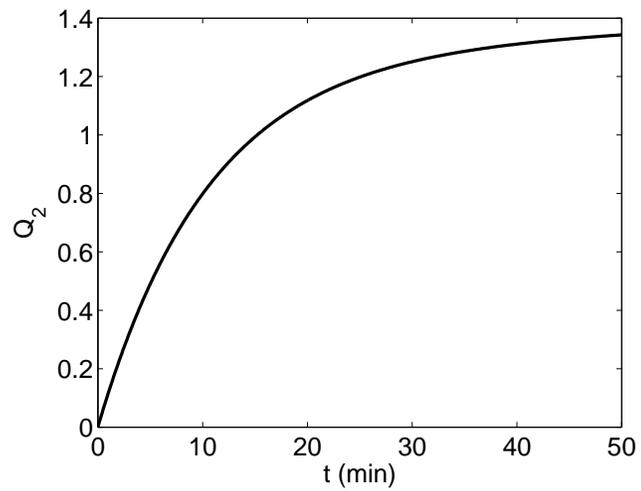


Figure 3.26: The functions  $Q_2(t)$  for the case with time delay.

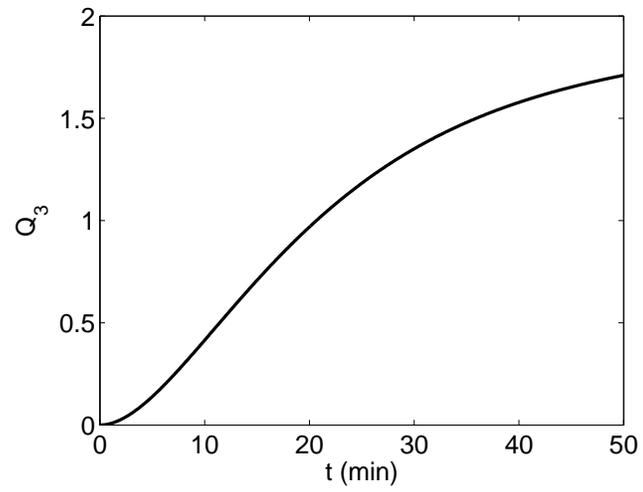


Figure 3.27: The functions  $Q_3(t)$  for the case with time delay.

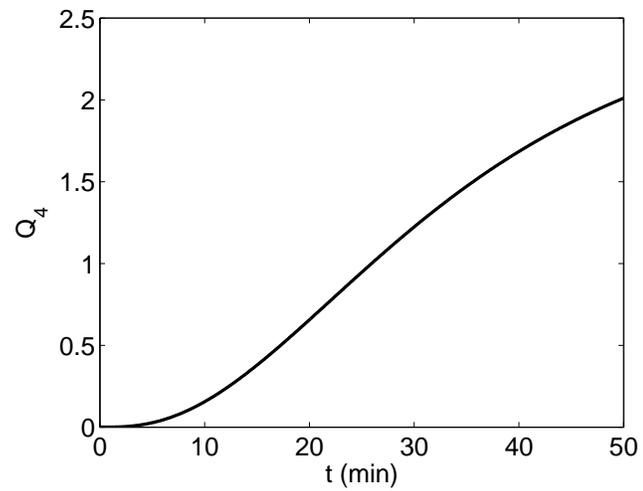


Figure 3.28: The functions  $Q_4(t)$  for the case with time delay.

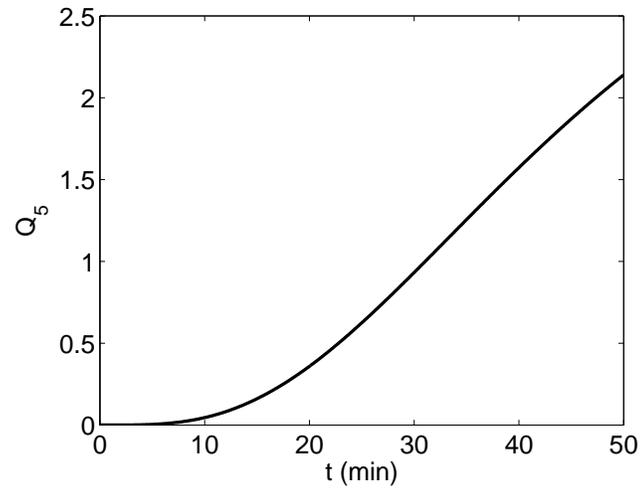


Figure 3.29: The functions  $Q_5(t)$  for the case with time delay.

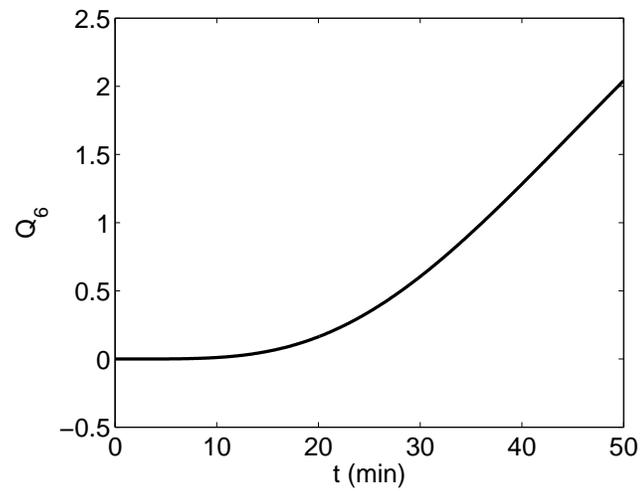


Figure 3.30: The functions  $Q_6(t)$  for the case with time delay.

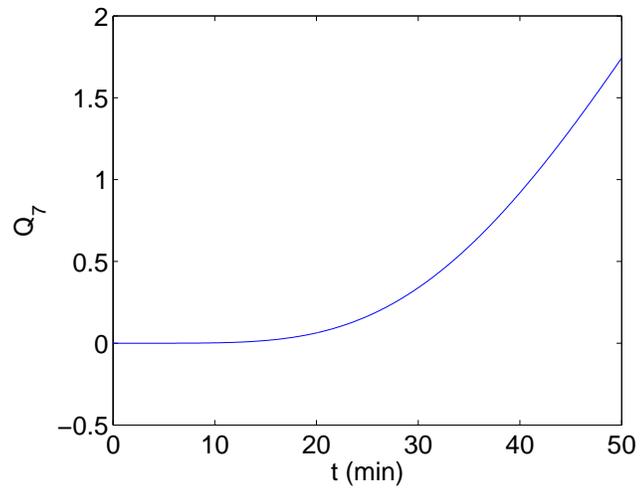


Figure 3.31: The functions  $Q_7(t)$  for the case with time delay.

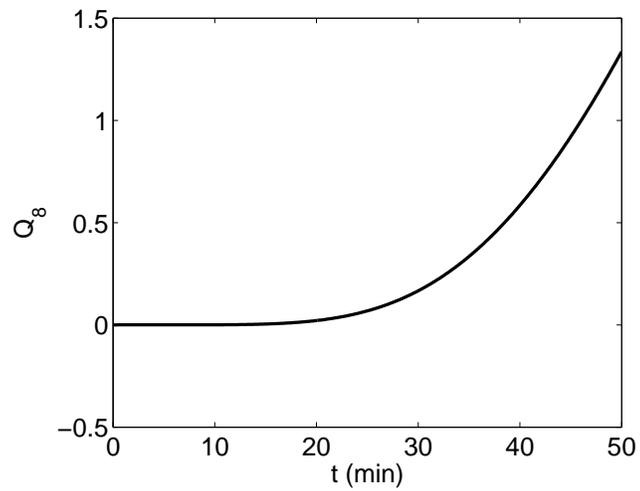


Figure 3.32: The functions  $Q_8(t)$  for the case with time delay.

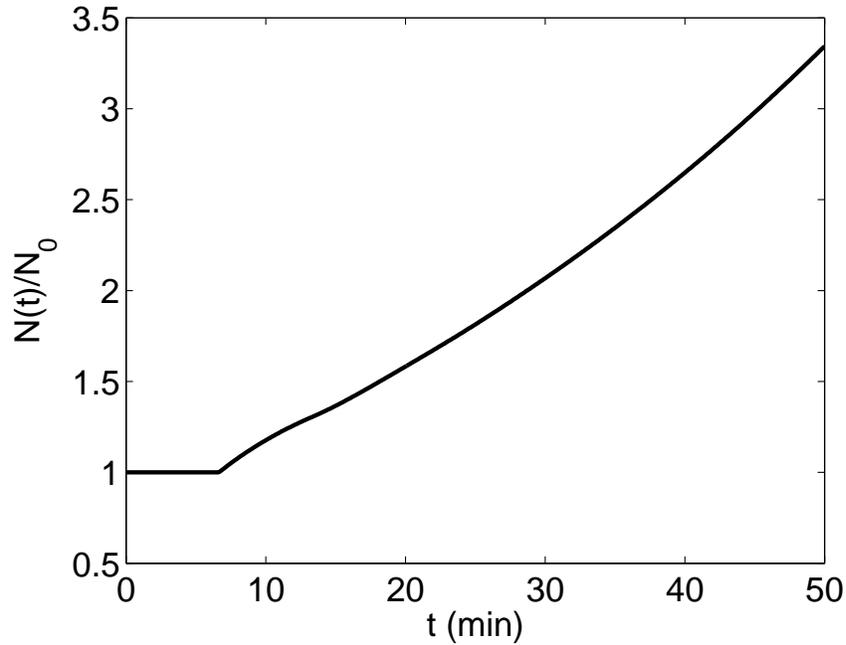


Figure 3.33: Total cell number for the case with time delay.

stochastic behavior. We assume that (1) the evolution of gene regulation networks satisfy Markov property; (2) the molecular number of key protein has bifurcation and therefore steady state distribution of key protein numbers has two peaks; (3) cell division happens once the key protein number reaches a threshold; (4) the initial number of key protein is between the two peaks in the distribution and the threshold is between the two peaks or around the peak of high protein number. Under these assumptions, it is found that the alive probability function can be written in the form of a double-exponential function. The more difficult the transition between the two peaks is, the more apparent the double-exponential form is. Based on this alive probability function, the age distribution of cell population can be obtained. It is found that after several generation times, the proportion of first generation cells is

still significant enough to be detectable.

We suggest the alive probability function can be written in the form of double-exponential function when the key protein number distribution is bistable. Usually the solution directly from Fokker-Planck equations or Stochastic Petri nets is numerical and does not contain analytical form, while the suggested double-exponential function has a simple analytical form and thus can be plugged into population balance model to calculate the age distribution. Although the double-exponential form has not been proven, the alive probability functions obtained from both Fokker-Planck equations and Stochastic Petri nets match this form very well. Laplace transformation is also used to solve the population balance models. Population balance models are often solved numerically [76, 84]. An analytical solution is very complex [77]. Laplace transformation can help us obtain a relatively simple analytical solution quickly. Therefore both double-exponential function and Laplace transformation contribute to obtaining an analytical form of age distribution, especially an analytical form of first generation cell fraction.

The significance of the first generation cell fraction is not difficult to understand. The stochasticity of gene regulation networks causes the bifurcation of the first-generation cells, which are initially homogeneous. Then one group of cells has higher number of key division proteins and therefore divides soon and keeps growing. The other group of cells has lower number of key proteins and therefore holds for a long time without division. Because of the existence of the latter group, the fraction of first generation cells can be significant after a relatively long time. It was indeed detected that a certain fraction of the oldest cells can exist in a cell population [73].

It should be noted that the bifurcation of replication activity does not only happen in bacteria. Some species of viruses also have this kind of bifurcation. When these viruses infect cells, viruses in some cells are in “active” state and replicate quickly and then kill these cells and release their gene; while viruses in other cells are in “latent” state and replicate slowly [62]. Therefore the methods and conclusions in this work may be of much wider use.

## Chapter 4

### Deterministic and Stochastic Modeling of Autoinducer-2 Uptake

#### Regulation Network in *Escherichia coli*

##### 4.1 Introduction

“Quorum sensing” is the biological process through which bacteria communicate with each other by using chemical signal molecules [20]. Autoinducer-2 (AI-2) is an important signal molecule and much work has been done on its synthesis, uptake and functioning [30, 86, 46]. A very important aspect of such work is the study on AI-2 uptake to reveal how bacteria transport AI-2 molecules from the surrounding environment into their own cells. Recent research indicates that AI-2 uptake is regulated by similar mechanisms in some bacteria [32, 36, 86, 87]. The regulation networks of AI-2 uptake in *Salmonella typhimurium* and *Escherichia coli* are shown in Figs. 4.1 and 4.2. In both networks, the AI-2 uptake is regulated by a set of *lsr* genes. AI-2 is first transported into cells by the transporter apparatus encoded by *lsrACDB*. Then AI-2 is phosphorylated by the kinase LsrK. LsrR can repress the transcription of *lsr* genes but the LsrR repression can be derepressed by the phospho-AI-2. Finally, the protein LsrF and LsrG involve the degradation of AI-2 [36, 88].

The development of a detailed differential equation model for a cellular bio-

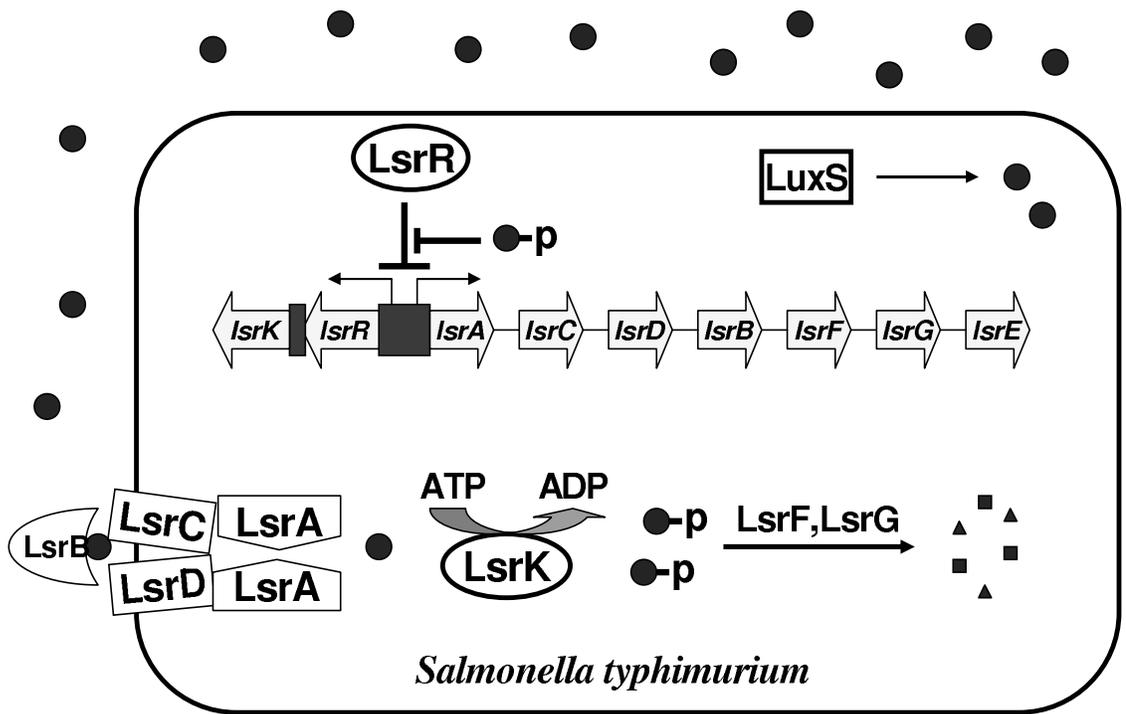


Figure 4.1: The regulation networks of AI-2 uptake in *Salmonella typhimurium* (Adapted from [88]).

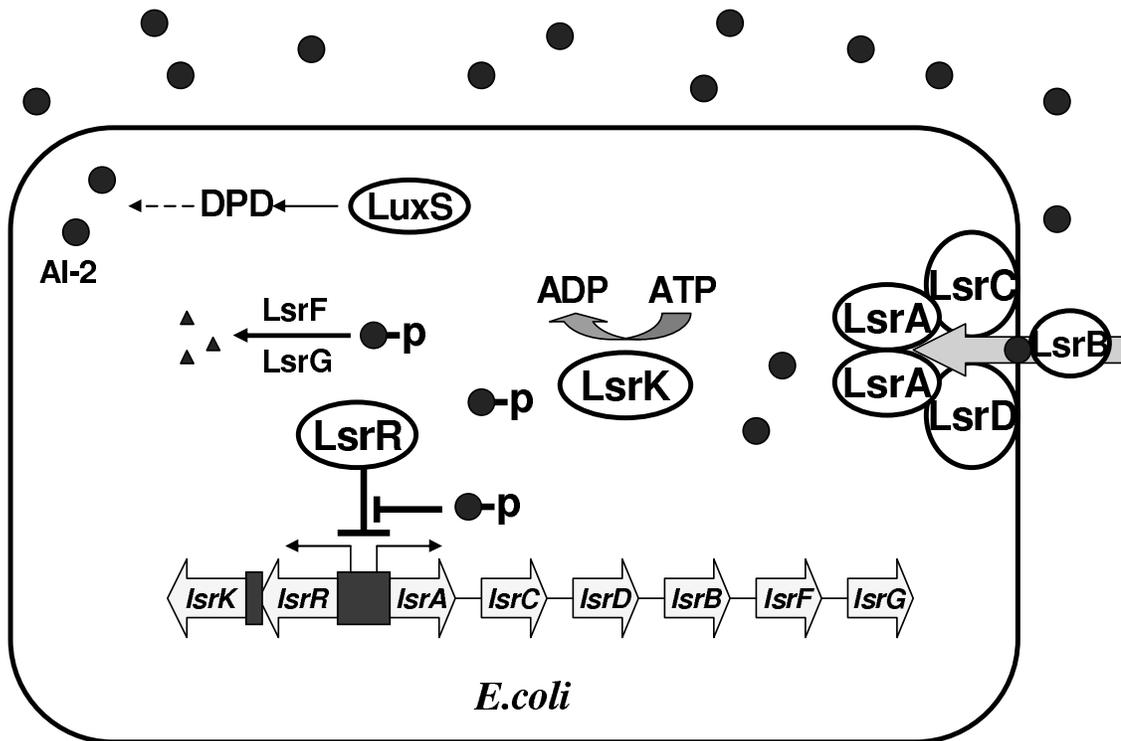


Figure 4.2: The regulation networks of AI-2 uptake in *Escherichia coli* (Adapted from [36]).

chemical pathway often requires the estimation of many parameters resulting in computationally intensive problems [89]. In some cases, parameter values are estimated by a trial-and-error fit to the experimental data [90, 91], something that requires the repetition of simulation, which is inefficient, and cannot obtain the best parameter values. An alternative is to calculate parameter values through mathematical optimization techniques, and several applications on biological systems with different optimization methods have been reported [91, 92, 93]. However, for complex gene networks, the simultaneous estimation of all parameters through optimization may turn out to be inefficient due to limited experimental data, measurement errors and other factors. For the AI-2 uptake network in *E.coli*, experimental data have been obtained from different mutations of bacteria [32, 36]. Each mutation corresponds to a simplified regulation network and thus can help us estimate some parameters. Therefore we suggest that parameter values can be estimated separately from experimental data for different mutations.

A deterministic ordinary differential equation (ODE) model assumes molecule numbers are deterministic and continuous. However, when molecule numbers are small, they need to be considered discrete and stochastic fluctuations become important [4, 5]. Stochastic models have been simulated for some biological systems [3, 8], including the AI-2 quorum sensing system in *E.coli* [51]. However this AI-2 model focuses more on AI-2 synthesis and the uptake of AI-2 is simplified into a single step. Moreover, this model did not compare the stochastic model with the deterministic one. It is usually not necessary to use stochastic models if the means of stochastic simulations follow closely the result of deterministic simulations. Some

stochastic simulations predict bifurcations, in which stochastic means show significant and very important differences from deterministic results [8, 12]. However, this leaves the open question of whether stochastic models are important if there is no stochastic bistability in the system.

In this work, a deterministic model for the AI-2 uptake regulation network in *E.coli* is first constructed. The parameters in this model are estimated step by step with experimental data for wild-type *E.coli* and their mutants. Stochastic simulation is also performed and a comparison between stochastic and deterministic simulations is made.

## 4.2 Models

### 4.2.1 Kinetic Network

Based on the regulation network in Fig. 4.2, a kinetic network is constructed and shown in Fig. 4.3. The pathway from external AI-2 (AI2ex) to decomposed AI-2 (AI2d) indicates the evolution of AI-2 molecules. First, external AI-2 molecules are transported into cells. Experiments suggested that there is an *lsr*-mediated uptake mechanism and an alternative mechanism for AI-2 uptake [32]. The alternative mechanism is represented with rate constant  $k_{tpa}$  in Fig. 4.3. For the *lsr*-mediated uptake mechanism, it was reported that protein LsrB is the AI-2 transporter in *S.typhimurium* [88]. This may also be the case in *E.coli* and in the kinetic network, LsrB is considered as a catalyst. It was also reported that proteins LsrC and LsrD are used to construct the transport channel [88]. Hence sufficient amounts of the two

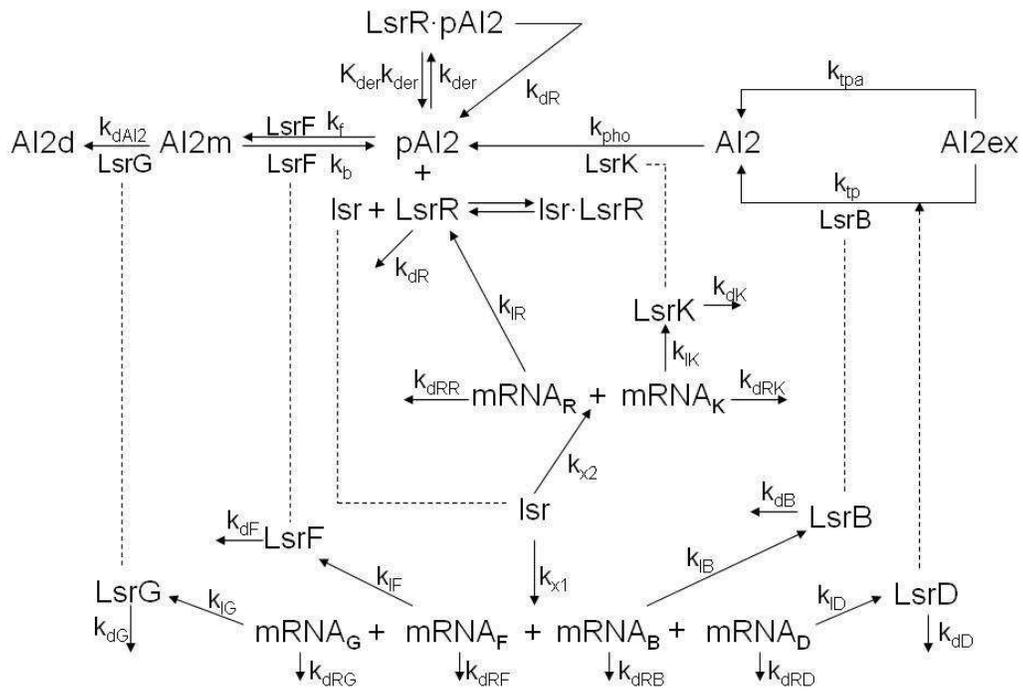


Figure 4.3: AI-2 uptake network of wild type *E. coli*.

proteins are necessary for uptake. Here we arbitrarily choose LsrD as the bottleneck and assume that the LsrD number must reach a threshold for *lsr*-mediated uptake to take place. After entering the cells, free AI-2 molecules (AI2) are phosphorylated into phospho-AI-2 (pAI2) by the kinase LsrK [32]. Phospho-AI-2 molecules involve the regulation of *lsr* genes and finally they are decomposed by proteins LsrF and LsrG. Understanding of the phospho-AI-2 decomposition is still vague and two possible explanations have been proposed. Here we choose one of them, which assumes that the decomposition consists of two steps through an intermediate state (AI2m) and that the first step is reversible. Each of the two steps is catalyzed by one of the proteins LsrF and LsrG [87]. In the kinetic network, it is arbitrarily assumed that LsrF is responsible for the first step and LsrG for the second.

The Lsr proteins play important roles in all the steps of the AI-2 uptake process and they are encoded from two operons: *lsrRK* operon and *lsrACDBFG* operon [32]. Transcripts of the two operons are from the same promoter (*lsr*). LsrR can bind the operator and let the promoter be repressed (*lsr*·LsrR). However the repression can be weakened by phospho-AI-2 [32]. So it can be suggested that phospho-AI-2 forms another complex (LsrR·pAI2) with LsrR so that the number of LsrR is reduced.

## 4.2.2 Deterministic Model

The deterministic ODE model that describes the kinetic network is listed in Table 4.1.  $X$  is  $OD_{600}$  and represents cell amount, so the first ODE is for cell growth.

---

$\frac{dX}{dt}$	$= \mu X$
$\frac{d[AI2]}{dt}$	$= k_{tp}[LsrB][AI2ex] + k_{tpa}[AI2ex] - k_{pho}[LsrK][AI2] - \mu[AI2]$
$\frac{d[pAI2]}{dt}$	$= k_{pho}[LsrK][AI2] + k_{dR}[LsrR \cdot pAI2] - k_{der}[pAI2][LsrR] + K_{der}k_{der}[LsrR \cdot pAI2]$ $- k_f[LsrF][pAI2] + k_b[LsrF][AI2m] - \mu[pAI2]$
$\frac{d[AI2m]}{dt}$	$= k_f[LsrF][pAI2] - k_b[LsrF][AI2m] - k_{dAI2}[LsrG][AI2m] - \mu[AI2m]$
$\frac{d[AI2d]}{dt}$	$= k_{dAI2}[LsrG][AI2m] - \mu[AI2d]$
$\frac{d[mRNA_B]}{dt}$	$= k_{x1}[lsr] - k_{dRB}[mRNA_B] - \mu[mRNA_B]$
$\frac{d[mRNA_D]}{dt}$	$= k_{x1}[lsr] - k_{dRD}[mRNA_D] - \mu[mRNA_D]$
$\frac{d[mRNA_F]}{dt}$	$= k_{x1}[lsr] - k_{dRF}[mRNA_F] - \mu[mRNA_F]$
$\frac{d[mRNA_G]}{dt}$	$= k_{x1}[lsr] - k_{dRG}[mRNA_G] - \mu[mRNA_G]$
$\frac{d[mRNA_K]}{dt}$	$= k_{x2}[lsr] - k_{dRK}[mRNA_K] - \mu[mRNA_K]$
$\frac{d[mRNA_R]}{dt}$	$= k_{x2}[lsr] - k_{dRR}[mRNA_R] - \mu[mRNA_R]$
$\frac{d[LsrB]}{dt}$	$= k_{lB}[mRNA_B] - k_{dB}[LsrB] - \mu[LsrB]$
$\frac{d[LsrD]}{dt}$	$= k_{lD}[mRNA_D] - k_{dD}[LsrD] - \mu[LsrD]$
$\frac{d[LsrF]}{dt}$	$= k_{lF}[mRNA_F] - k_{dF}[LsrF] - \mu[LsrF]$
$\frac{d[LsrG]}{dt}$	$= k_{lG}[mRNA_G] - k_{dG}[LsrG] - \mu[LsrG]$
$\frac{d[LsrK]}{dt}$	$= k_{lK}[mRNA_K] - k_{dK}[LsrK] - \mu[LsrK]$
$\frac{d[LsrR]}{dt}$	$= k_{lR}[mRNA_R] - k_{dR}[LsrR] - k_{rep}[lsr][LsrR] + K_{rep}k_{rep}[lsr \cdot LsrR] - k_{der}[pAI2][LsrR]$ $+ K_{der}k_{der}[LsrR \cdot pAI2] - \mu[LsrR]$
$\frac{d[lsr \cdot LsrR]}{dt}$	$= k_{rep}[lsr][LsrR] - K_{rep}k_{rep}[lsr \cdot LsrR] - \mu[lsr \cdot LsrR]$
$\frac{d[lsr]}{dt}$	$= -k_{rep}[lsr][LsrR] + K_{rep}k_{rep}[lsr \cdot LsrR] + \mu[lsr \cdot LsrR]$
$\frac{d[LsrR \cdot pAI2]}{dt}$	$= k_{der}[pAI2][LsrR] - K_{der}k_{der}[LsrR \cdot pAI2] - k_{dR}[LsrR \cdot pAI2] - \mu[LsrR \cdot pAI2]$
$\mu$	$= \frac{\mu_{max}(X_m - X)}{K'_s + (X_m - X)}$
$[AI2ex]$	$= [AI2total] - X([AI2] + [pAI2] + [LsrR \cdot pAI2] + [AI2m] + [AI2d])/C_a$

---

Table 4.1: ODEs of AI-2 uptake network in *E. coli*. [\*] means molecule number of \* in one cell.

Other ODEs are kinetics, in which the square brackets  $[\cdot]$  represent the number of corresponding molecules in one cell. In those ODEs, the last term that includes  $\mu$  comes from cell division [94]. In the kinetics ODE for  $lsr$ , the last term has a positive sign because genes replicate as cells divide and thus we need to let the sum of  $d[lsr]/dt$  and  $d[lsr \cdot LsrR]/dt$  be zero. There are two more equations at the end of Table 4.1. The first one is used to calculate specific growth rate  $\mu$ . The second is for external AI-2 activity  $[AI2ex]$ .  $[AI2total]$  represents total amount of released AI-2. In this equation the coefficient,  $C_a$ , used for last several terms comes from the assumption that one unit of normalized AI-2 activity (AI-2 activity divided by  $OD_{600}$ ) represents a certain number of AI-2 molecules [51].

Some parameter values are based on the literature as shown in Table 4.2. Those parameters are based on previous research. The values for translation rate, degradation rate of Lsr proteins and phosphorylation rate  $k_{pho}$  are adjusted based on Li *et al.* [51] within the same order of magnitude. The values for mRNA degradation rate and protein binding rate,  $k_{rep}$  and  $k_{der}$ , are adjusted based on the work in Chapter. 2. The remaining parameters need to be obtained by fitting experimental data.

### 4.2.3 Optimization Model for Parameter Estimation

Model parameters not listed in Table 4.2, will be estimated by solving an appropriate optimization problem:

$$\min_{\mathbf{p}} J = \sum_i (y_{i,sim} - y_{i,exp})^2 \quad (4.1)$$

Parameter	value ( $min^{-1}$ )	Parameter	value ( $min^{-1}$ )
$k_{dRB}$	0.3	$k_{dRD}$	0.4
$k_{dRF}$	0.1	$k_{dRG}$	0.1
$k_{dRK}$	0.5	$k_{dRR}$	0.4
$k_{lB}$	0.5	$k_{lD}$	0.8
$k_{lF}$	0.3	$k_{lG}$	0.3
$k_{lK}$	0.4	$k_{lR}$	0.6
$k_{dB}$	0.05	$k_{dD}$	0.02
$k_{dF}$	0.03	$k_{dG}$	0.03
$k_{dK}$	0.04	$k_{dR}$	0.06
$k_{rep}$	10	$k_{der}$	10
$k_{pho}$	0.5		

Table 4.2: Parameter values adjusted from literature.

subject to

$$\mathbf{f}\left(\frac{d\mathbf{v}}{dt}, \mathbf{v}, \mathbf{p}, \mathbf{p}_a, t\right) = 0 \quad (4.2)$$

$$\mathbf{v}(t_0) = \mathbf{v}_0 \quad (4.3)$$

$$y_{i,sim} = g(\mathbf{v}, t_i) \quad (4.4)$$

The objective function (4.1) is the sum of squares of the differences between experimental data and their corresponding model simulation data. The first set of constraints (4.2) are the ODEs describing the kinetics. The constraints can be the full ODE model in Table 4.1, but also can be part of the ODEs to describe the kinetics of mutants.  $\mathbf{v}$ ,  $\mathbf{p}_a$  and  $\mathbf{p}$  represent the ODE variables, assigned parameters in Table 4.2 and parameters to be fitted. The second set of constraints (4.3) assign initial conditions. Here the initial values for  $X$  and  $lsr$  are 0.025 and 1 respectively and the other initial values are zero. The third set of constraints (4.4) indicate how to obtain simulation values which correspond to experiment values.  $t_i$  are the time points of experimental data. In this work we will use two sets of experimental data: external AI-2 activity and  $\beta$ -galactosidase units. When external AI-2 activity is used, the simulation data are:

$$y_{i,sim} = AI2ex(t_i) \quad (4.5)$$

When  $\beta$ -galactosidase units are used, the units can be obtained by multiplying transcription rate with a constant [51]:

$$y_{i,sim} = C_b k_x [lsr] \quad (4.6)$$



In Fig. 4.4 shows the SPN for AI-2 uptake network in *E.coli*. The circle places in the SPN indicate cellular species, including gene, mRNA, proteins, signal molecules and their complexes. The values for those places indicate the molecule numbers of the species. One special feature is that the “reach” place at the bottom right of the SPN figure indicates whether LsrD protein number reaches the threshold. The bold vertical lines indicate activities including transcription, translation, protein binding and signal processing. Those activities are stochastic and the time of their happening follows a negative exponential distribution [15]. The thin vertical line is an instant activity, which changes the value for “reach” place as soon as LsrD number reaches the threshold. There are also some triangles which are input gates and indicate the preconditions of related activities [17].

For most species, their amounts in both deterministic and stochastic models are expressed in the form of molecule number. So most parameter values in stochastic model are the same as those in deterministic model. However in the deterministic model, external AI-2 amount is expressed in the form of activity. So in order to obtain external AI-2 number per cell for stochastic model, a transformation needs to be made for activity and the uptake rate constants  $k_{tp}$  and  $k_{tpa}$  also need to be converted. The details will be discussed in the next section.

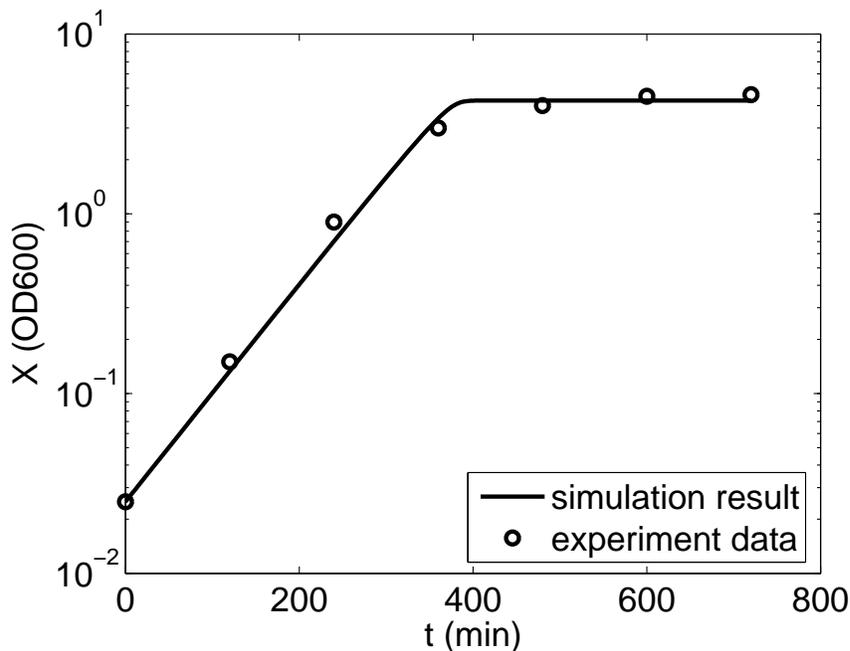


Figure 4.5: Fitting results for cell growth data. Estimated parameter values:  $\mu_{max} = 0.015$ ,  $K'_s = 0.3216$ ,  $X_m = 4.2637$ . Experiment data are from Wang *et al.* [32]

## 4.3 Results

### 4.3.1 Parameter Estimation

First, the cell growth parameters are estimated by fitting the  $OD_{600}$  data. The result shown in Fig. 4.5 indicates that the simulation with obtained parameter values matches the experimental data well. It is also can be noticed that the switch time point between exponential stage and stationary stage is just a little earlier than 400 minutes.

Next, the kinetic networks of different mutants are used to fit experiment data. Those networks are simplified from the wild type network in Fig. 4.3 and are

summarized in Figs. 4.6 to 4.10. The simplifications come from two causes: Some of the *lsr* genes are mutated; and only one of the two sets of experimental data (external AI-2 activity and  $\beta$ -galactosidase units) is considered. Both causes allow us to ignore part the wild type network. The kinetic networks of mutants have different complexities. For networks fitting external AI-2 activity, no AI-2 uptake happens in  $\Delta lsrK$  mutant so no kinetic network for this case is shown. The network of  $\Delta lsrACDBFG$  mutant only contains a single step which is the alternative uptake (Fig. 4.7). The network for  $\Delta lsrR$  mutant is more complex than the previous two because both *lsr*-mediated and alternative AI-2 uptake happen in this case (Fig. 4.8).

For networks fitting  $\beta$ -galactosidase units, the network of  $\Delta lsrR$  is the simplest because there is no repression here and full transcription happens (Fig. 4.6). For the  $\Delta lsrK$  mutant there is a repression on transcription (Fig. 4.9). For the  $\Delta lsrACDBFG$  mutant there are, not only repression, but also de-repression and therefore the network is the most complex (Fig. 4.10). Because of the difference in complexity, we will estimate the parameter values step by step from simple network to complex network.

Fitting results for mutant networks are summarized in Table 4.3. With the estimated parameters, deterministic simulations for each step are shown in Figs. 4.11 to 4.11. The fitting steps are ordered from simple networks to more complex ones, so that the latter steps can use the parameter values from former steps. In the first step of Table 4.3, total amounts of synthesized AI-2 are interpolated. In the second, third and fourth step, the parameters have different values before and

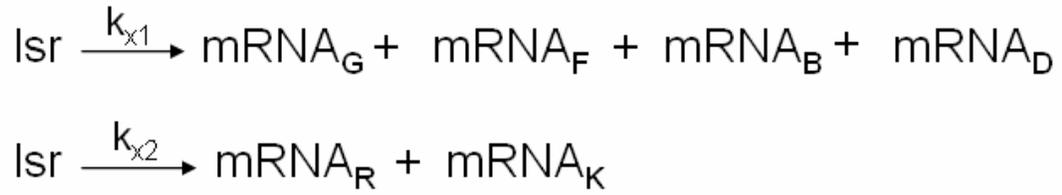


Figure 4.6: Kinetic network of mutant  $\Delta\text{lsrR}$  to fit  $\beta$ -galactosidase units (full transcription).



Figure 4.7: Kinetic network of mutant  $\Delta\text{lsrACDBFG}$  to fit external AI-2 activity (alternative uptake).

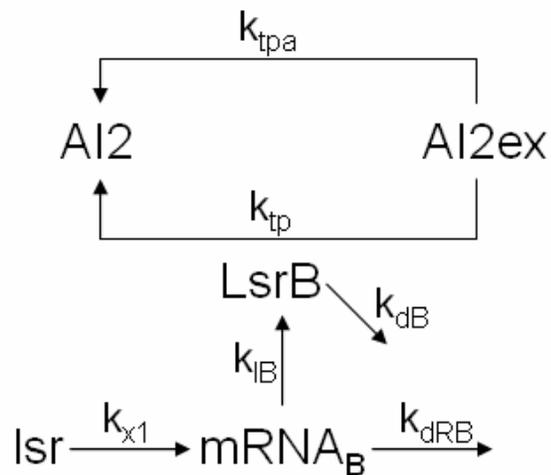


Figure 4.8: Kinetic network of mutant  $\Delta\text{lsrR}$  to fit external AI-2 activity (both uptake).

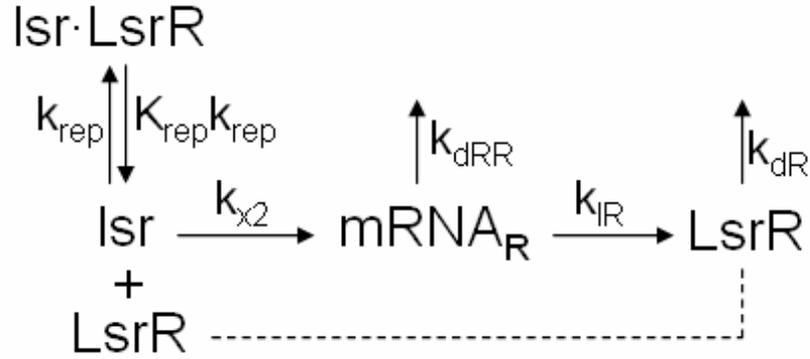


Figure 4.9: Kinetic network of mutant  $\Delta lsrK$  to fit  $\beta$ -galactosidase units (repression).

after 5 hours. Experiment data indicate that before around 5 hours almost no transcription from *lsr* promoter and no AI-2 uptake happen [32, 36]. Although the transcription level of *lsrRK* operon at 4 hours is not zero (shown in Fig. 4.13, from [36]), at that time there is no uptake and thus the transcription is meaningless. So we can set the transcription rates and uptake rate to be equal to zero before 5 hours. We also need to note that  $\beta$ -galactosidase units fitted in the fourth step are the units for *lsrRK* operon, while in other steps the units are for *lsrACDBFG* operon. In the fifth step, the value for uptake rate constant  $k_{tp}$  depends on whether LsrD number reaches the threshold  $T_D$ . We discussed earlier that *lsr*-mediated AI-2 uptake requires a sufficient amount of LsrD. Here in order to simply the problem, it is assumed that the threshold is reached when we reach 5 hours. This assumption is further discussed below.

The threshold of LsrD number should be in a reasonable region. In mutant  $\Delta lsrK$ , the transcription level is significantly larger than zero after 5 hours, which means some LsrD proteins are produced. However AI-2 uptake still does not happen

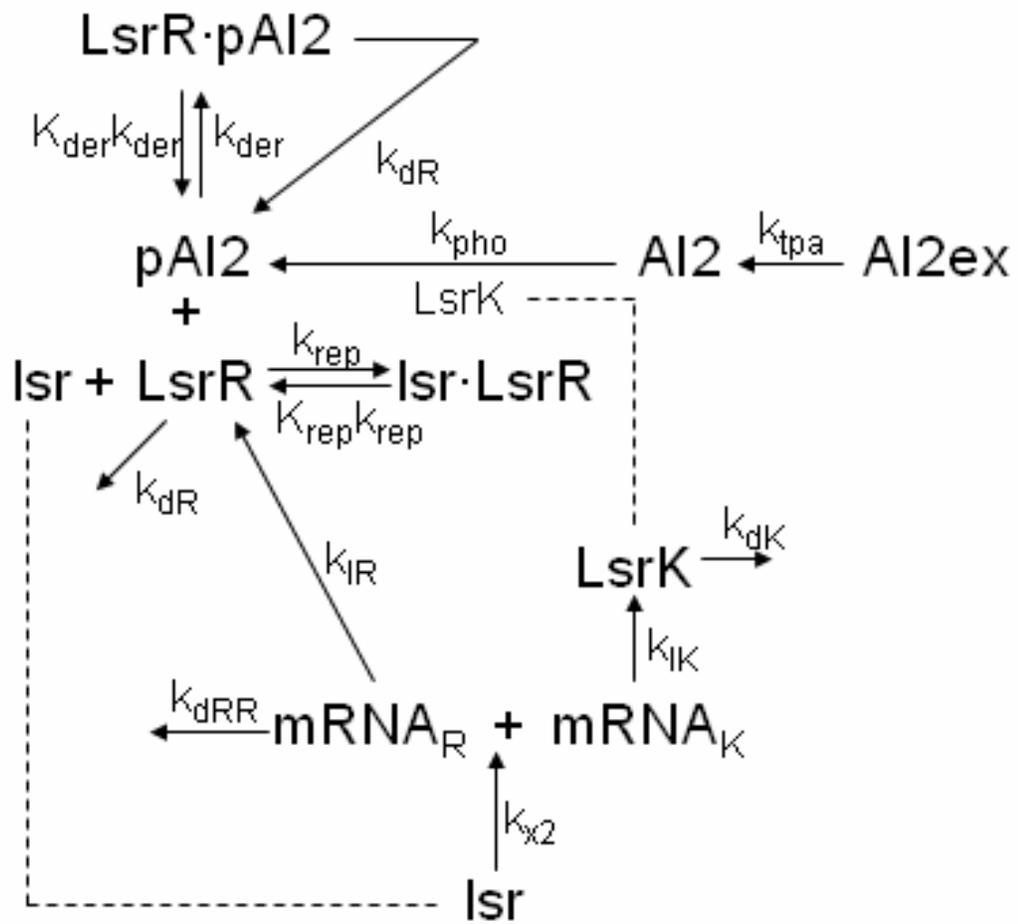


Figure 4.10: Kinetic network of mutant  $\Delta lsrB$  to fit  $\beta$ -galactosidase units (depression).

Step	Mutant	Experiment data fitted	Remark	Fitting result
1	$\Delta lsrK$	Ext. AI-2 activity	No uptake	[AI2total] shown in Fig. 4.11
2	$\Delta lsrR$	$\beta$ -gal. units	Full transcription1	$k_{x1} = \begin{cases} 0 & t < 5h \\ 7.8875 & t \geq 5h \end{cases}$
3	$\Delta lsrR$	$\beta$ -gal. units of $lsrRK$	Full transcription2	$k_{x2} = \begin{cases} 0 & t < 5h \\ 21.7764 & t \geq 5h \end{cases}$
4	$\Delta lsrACDBFG$	Ext. AI-2 activity	Alternative uptake	$k_{tpa} = \begin{cases} 0 & t < 5h \\ 0.1443 & t \geq 5h \end{cases}$
5	$\Delta lsrR$	Ext. AI-2 activity	Both uptake	$k_{tp} = \begin{cases} 0 & LsrD < T_D \\ 0.3837 & LsrD \geq T_D \end{cases}$
6	$\Delta lsrK$	$\beta$ -gal. units	Repression	$K_{rep} = 7.2171$
7	$\Delta lsrACDBFG$	$\beta$ -gal. units	De-repression	$K_{der} = 8.5987$

Table 4.3: Fitting results for mutant kinetic networks.

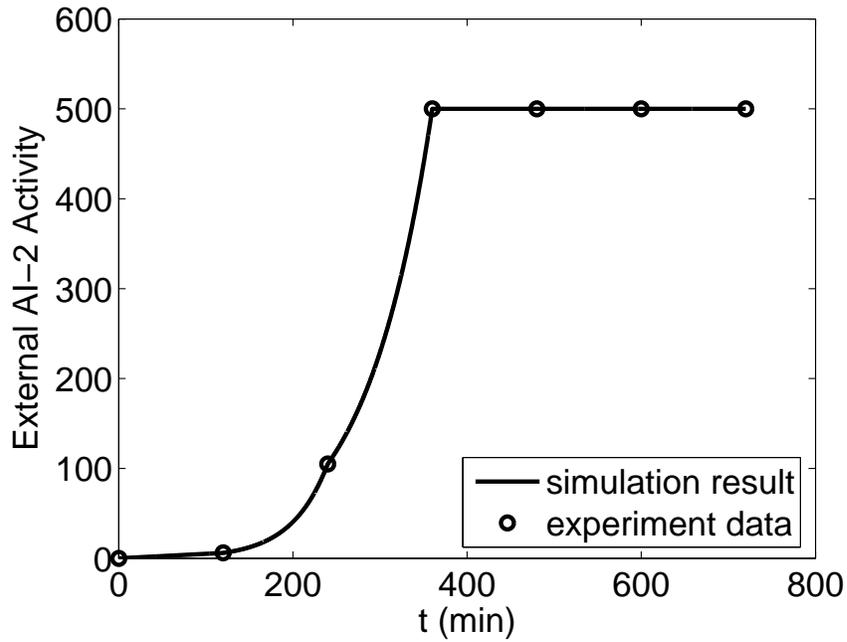


Figure 4.11: Deterministic simulation of mutant network corresponding to the step in Table 4.3 for “no uptake”. All experiment data are from [32] and [36]

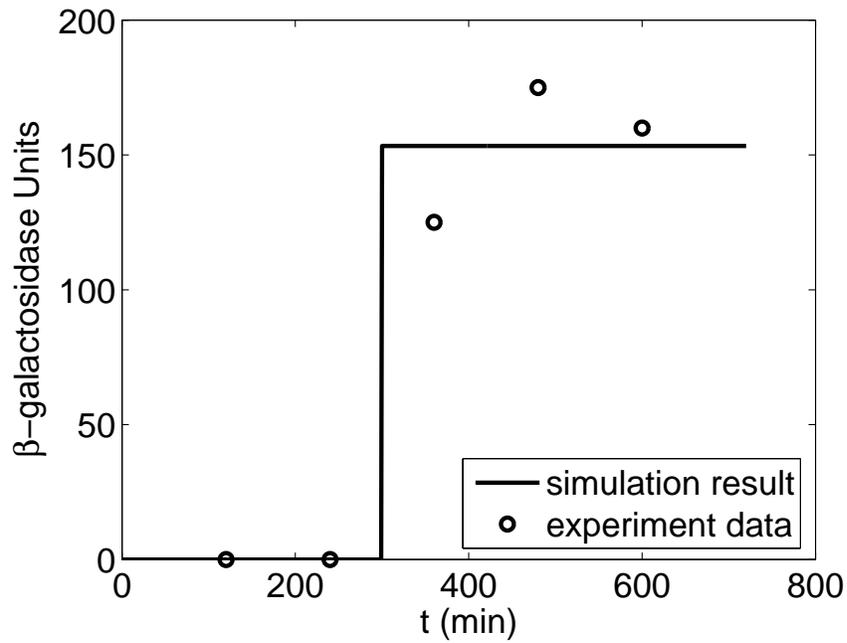


Figure 4.12: Deterministic simulation of mutant network corresponding to the step in Table 4.3 for “full transcription1”. All experiment data are from [32] and [36]

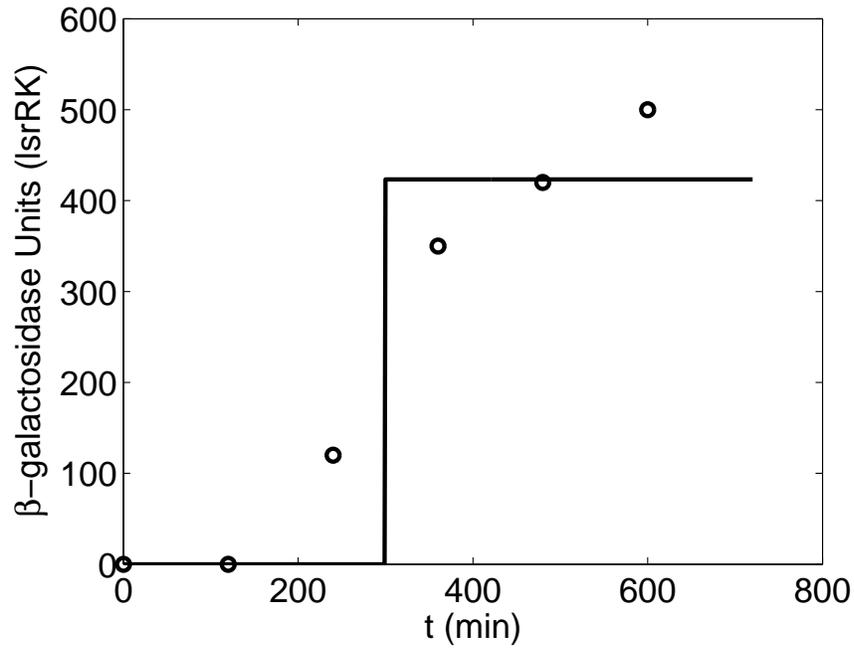


Figure 4.13: Deterministic simulation of mutant network corresponding to the step in Table 4.3 for “full transcription2”. All experiment data are from [32] and [36]

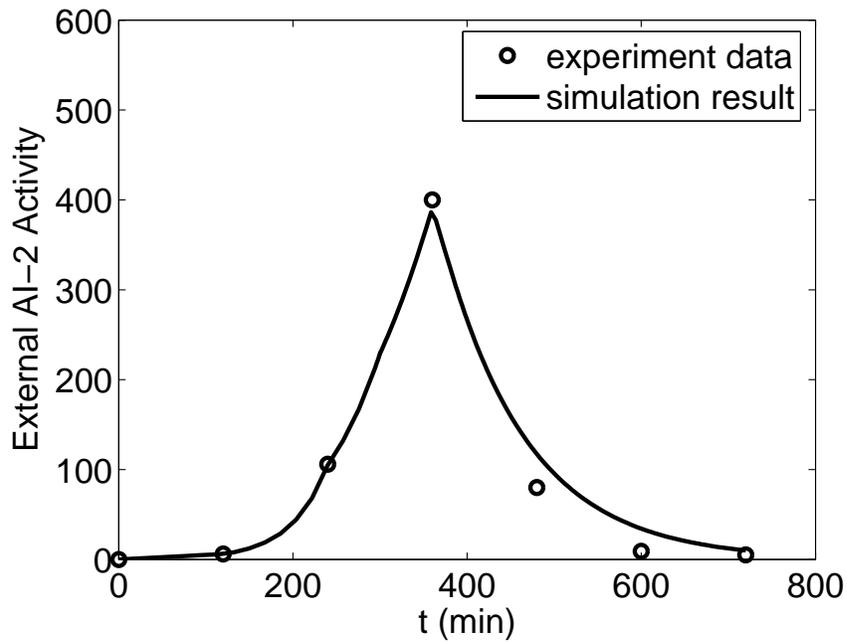


Figure 4.14: Deterministic simulation of mutant network corresponding to the step in Table 4.3 for “alternative uptake”. All experiment data are from [32] and [36]

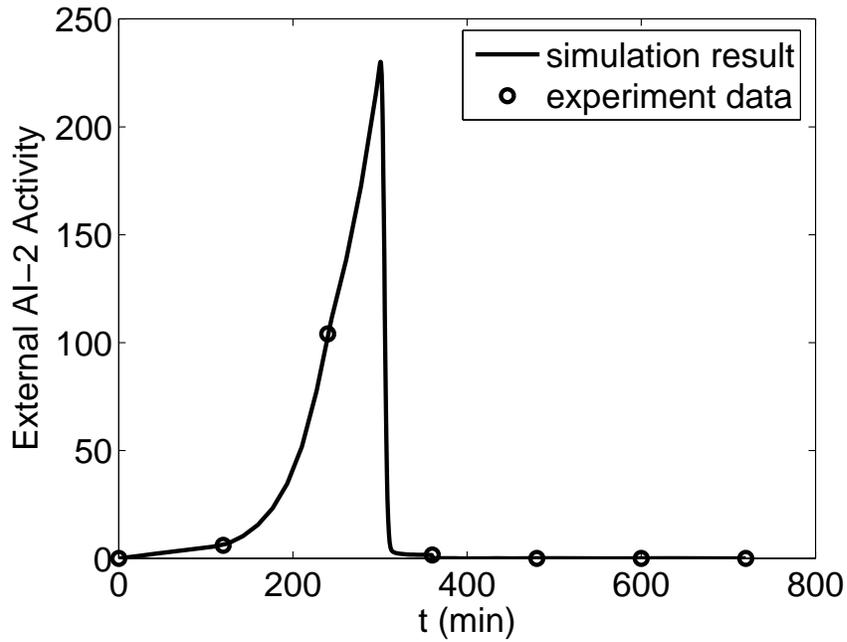


Figure 4.15: Deterministic simulation of mutant network corresponding to the step in Table 4.3 for “both uptake”. All experiment data are from [32] and [36]

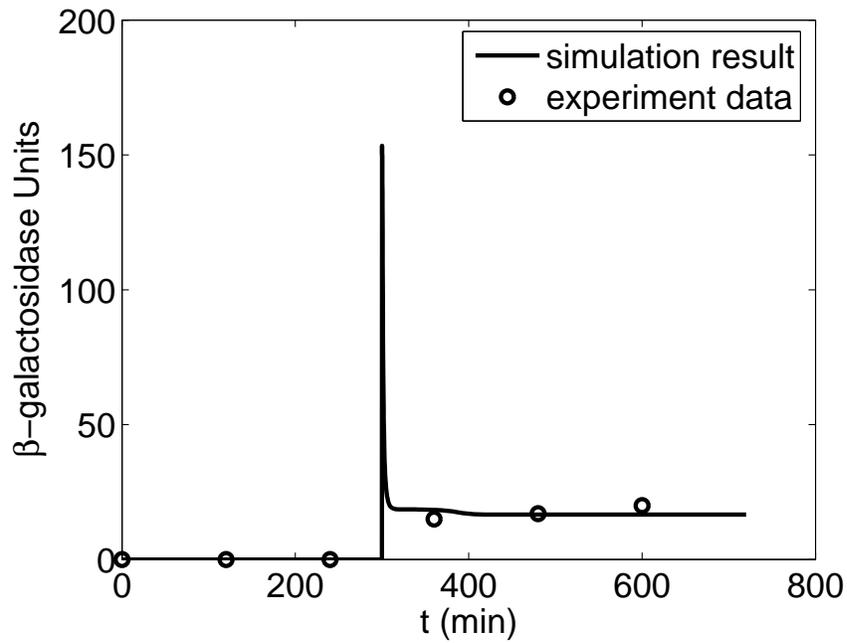


Figure 4.16: Deterministic simulation of mutant network corresponding to the step in Table 4.3 for “repression”. All experiment data are from [32] and [36]

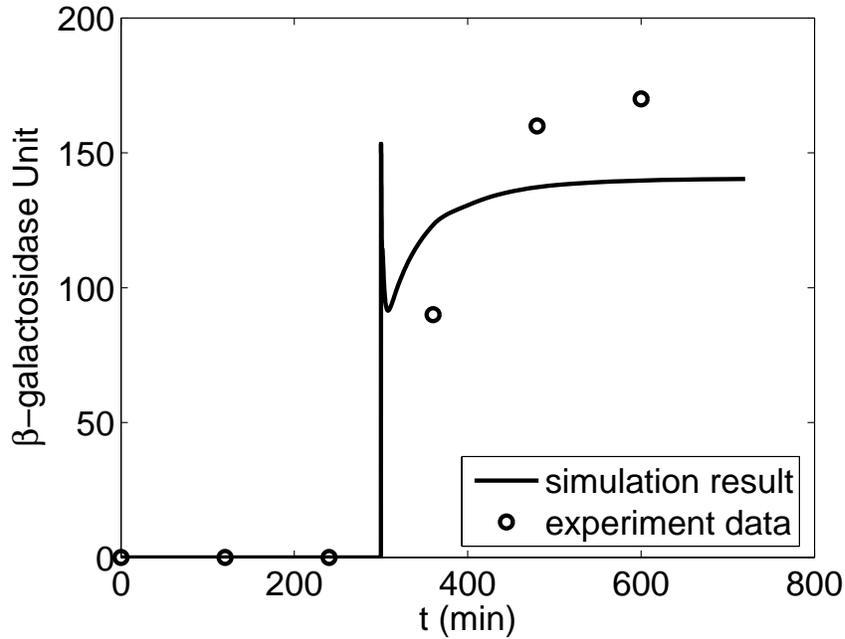


Figure 4.17: Deterministic simulation of mutant network corresponding to the step in Table 4.3 for “de-repression”. All experiment data are from [32] and [36]

at that time. So the threshold should be larger than LsrD numbers for this case. From Fig. 4.18, it is suggested that a threshold higher than 90 is acceptable. For the estimation  $k_{tp}$ , it is assumed that in mutant  $\Delta lsrR$ , LsrD number reaches the threshold soon after 5 hours. The assumption is reasonable if LsrD threshold is not too high. In Fig. 4.19 the threshold is 390, and simulation result still matches experiment data well. Therefore the threshold for LsrD number should be between 90 and 390.

The remaining undetermined parameters are  $T_D$ ,  $k_f$ ,  $k_b$  and  $k_{dAI2}$ . Experimental data for wild type *E.coli* will be used to obtain these parameter values. However, the wild type kinetic network as shown in Fig. 4.3 is so large that computational problems may be caused. We use effective *lsr* transcription level and AI-2 uptake rate to simplify the network. Effective transcription level is used to calculate  $T_D$ .

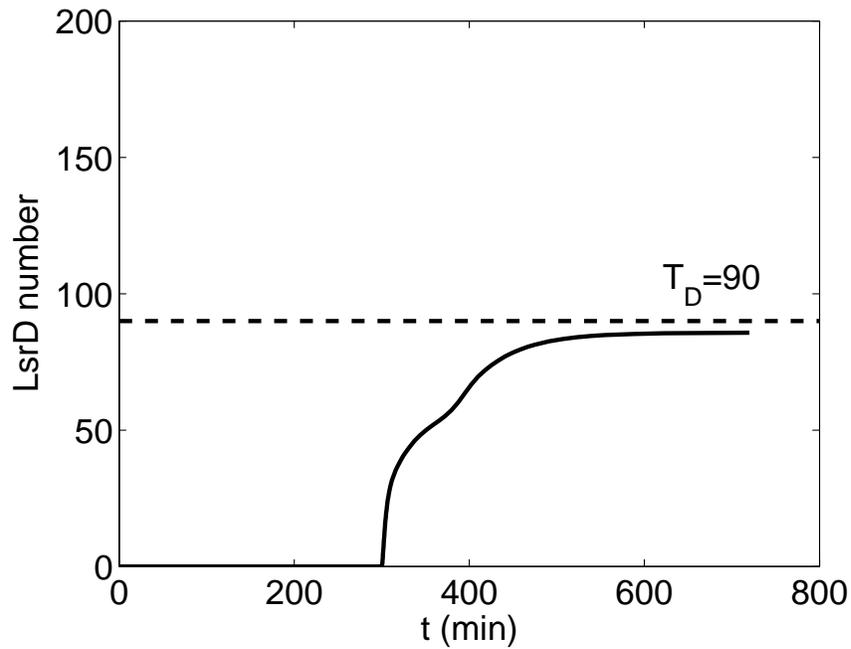


Figure 4.18: Analysis of LsrD threshold. LsrD number of mutant  $\Delta lsrK$  suggests lower bound of LsrD threshold.

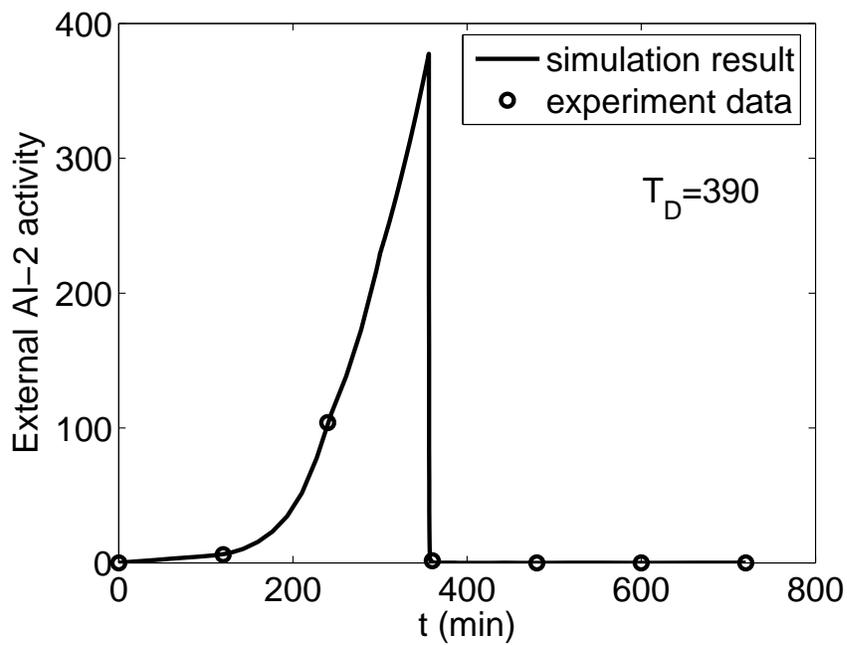


Figure 4.19: Analysis of LsrD threshold. External AI-2 activity of mutant  $\Delta lsrR$  when the threshold is 390, which suggests upper bound of LsrD threshold.

The network in this case only includes the expression of LsrD (Fig. 4.20). The effective level is interpolated from experimental data points of wild type  $\beta$ -galactosidase units (Fig. 4.21, from [32]). An artificial point is added at 5 hours because previous computation indicates that transcription level increases quickly after 5 hours. The value for this point is the average of the three subsequent data points. From experimental data of external AI-2 activity, wild type *E.coli* has the same activity as  $\Delta lsrACDBFG$  mutant at 6 hours, and has an activity of zero at 8 hours. It is suggested that LsrD number reaches  $T_D$  and *lsr*-mediated uptake happens between 6 and 8 hours. We set the time at 7 hours arbitrarily resulting in  $T_D$  equal to 297, which is the LsrD number at 7 hours as shown in Fig. 4.22. Effective AI-2 uptake rate is used to obtain the other parameter values. Fig. 4.24 shows the uptaken AI-2 amount, which is equal to activity of  $\Delta lsrK$  mutant minus activity of wild type. Then the effective uptake rate in Fig. 4.25 is the slope value for the uptaken amount. The kinetic network in this case is shown in Fig. 4.23. The AI-2 uptake and related regulations are simplified into a single step. The phosphorylation is also ignored because previous computation shows the step to be very quick. By using this network to fit the experimental  $\beta$ -galactosidase units, the parameters are estimated as follows:  $k_f = 1.070 \times 10^{-3}$ ,  $k_b = 5.954 \times 10^{-5}$  and  $k_{dAI2} = 1.077 \times 10^{-5}$ . In summary, in estimating  $T_D$ ,  $\beta$ -galactosidase units are used to calculate effective transcription level and external AI-2 activity is fitted; while in estimating  $k_f$ ,  $k_b$  and  $k_{dAI2}$ , external AI-2 activity is used to calculate effective AI-2 uptake rate and  $\beta$ -galactosidase units are fitted.

With all the parameter values available, the behavior of the full wild type

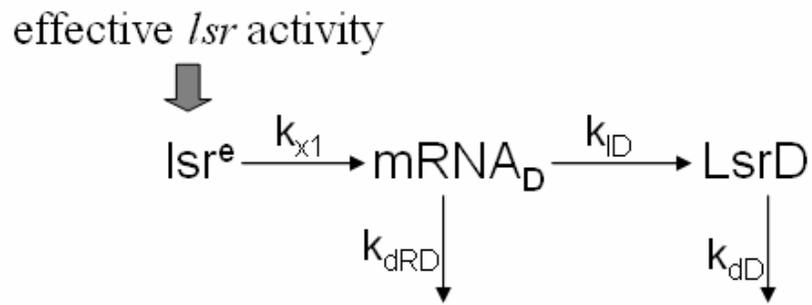


Figure 4.20: Kinetic network with effective transcription level.

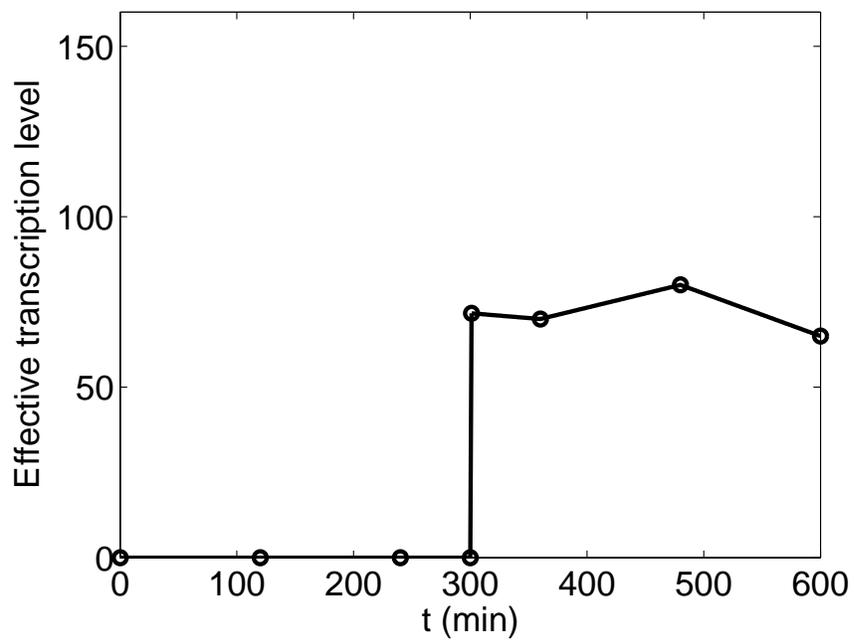


Figure 4.21: Effective transcription level for obtaining  $T_D$ .

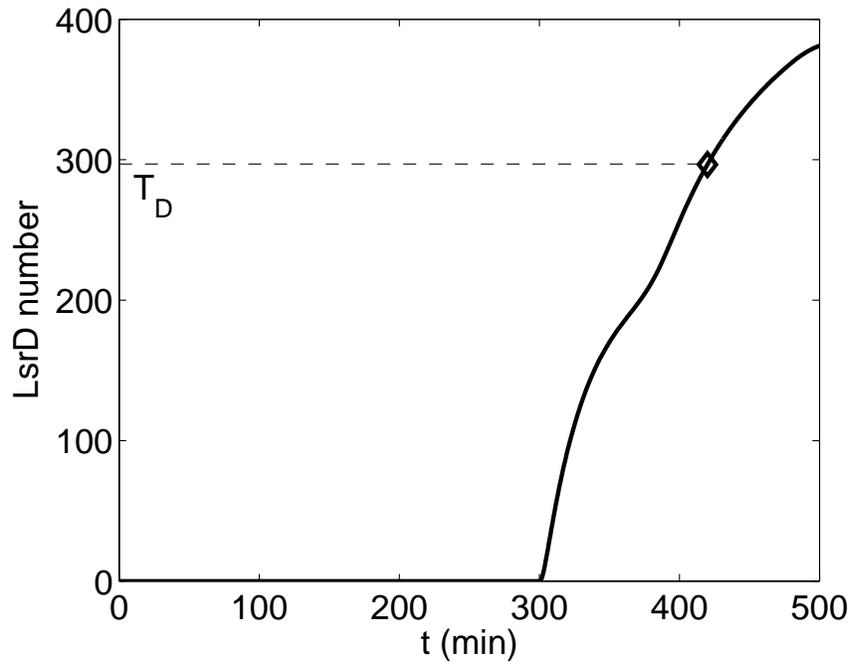


Figure 4.22: Fitting result for obtaining  $T_D$  from the network with effective transcription level:  $T_D = 297$ .

network can be simulated by using all the the ODEs in Table 4.1. The simulation results follow the experimental data well, as shown in Figs. 4.27 and 4.28. In Fig. 4.27, external AI-2 activity has a sharp drop between 6 and 8 hours because at that point the number of LsrD protein reaches the threshold  $T_D$  and remaining AI-2 molecules outside enter the cells quickly through *lsr*-mediated uptake. In Fig. 4.28,  $\beta$ -galactosidase units have two spikes. At the time of 5 hours, transcription level is high because there is no LsrR protein at first. Soon after that, LsrR proteins are produced and repress the *lsr* promoter. So the transcription level drops down and forms a spike. At time around 7 hours, *lsr*-mediated AI-2 uptake happens. AI-2 molecules enter cells and reduce the repression by LsrR. Thus transcription jumps up again. Then with high transcription level, more LsrF proteins are produced and

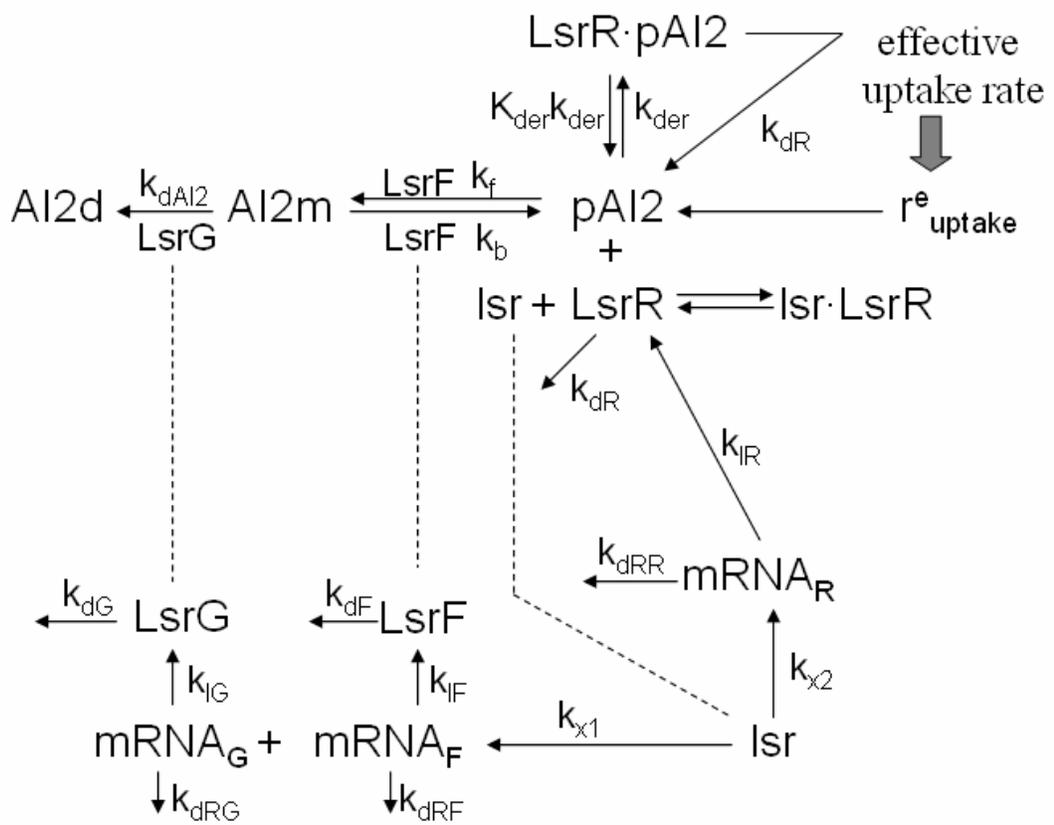


Figure 4.23: Kinetic network with effective AI-2 uptake rate.

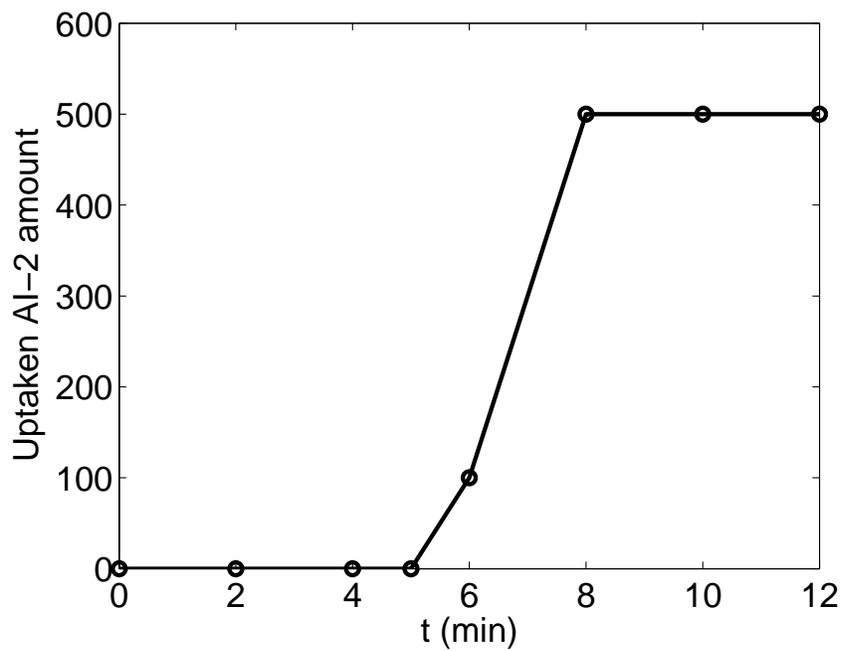


Figure 4.24: Effective AI-2 uptaken amount for obtaining  $k_f$ ,  $k_b$  and  $k_{dAI2}$ .

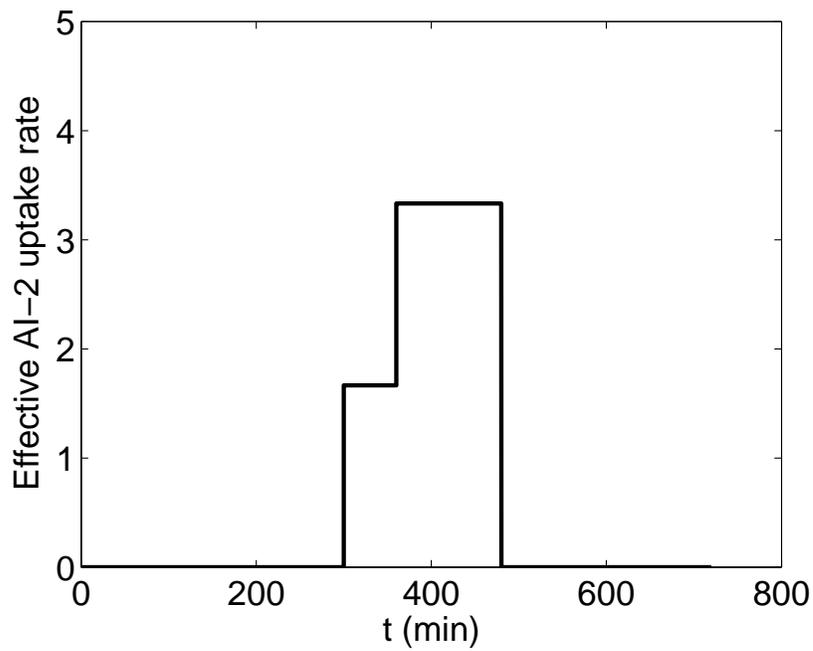


Figure 4.25: Effective AI-2 uptake rate for obtaining  $k_f$ ,  $k_b$  and  $k_{dAI2}$ .

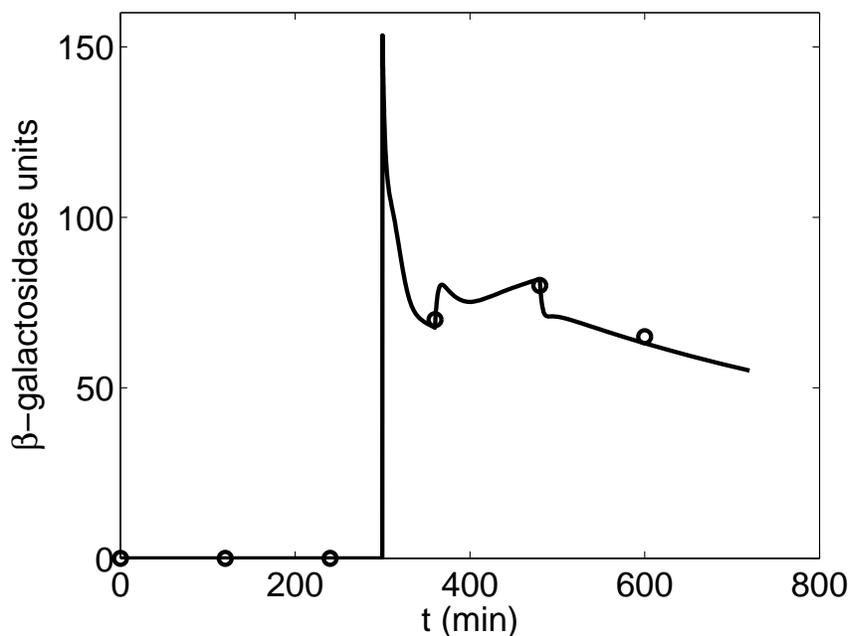


Figure 4.26: Fitting result for obtaining  $k_f$ ,  $k_b$  and  $k_{dAI2}$  from the network with effective AI-2 uptake rate:  $k_f = 1.070 \times 10^{-3}$ ,  $k_b = 5.954 \times 10^{-5}$  and  $k_{dAI2} = 1.077 \times 10^{-5}$ .

help phospho-AI-2 (pAI2) transform into the intermediate (AI2m). So the effect of de-repression is reduced and transcription level drops down again. Then another spike is formed. In the next section, we see how stochastic simulation explains that such spikes will be as sharp when a cell population is observed as a whole.

### 4.3.2 Results of Stochastic Simulation

Stochastic simulation is performed from a time of 390 minutes to 490 minutes. Deterministic simulation suggests that after 390 minutes, cell mass  $X$  almost reaches its maximum  $X_m$  and cell population enters stationary stage. Therefore the effect of cell growth need not be considered in stochastic simulation. At this stage cells do not divide so quickly, so SPN model in Fig. 4.4 can be simulated for 100 minutes,

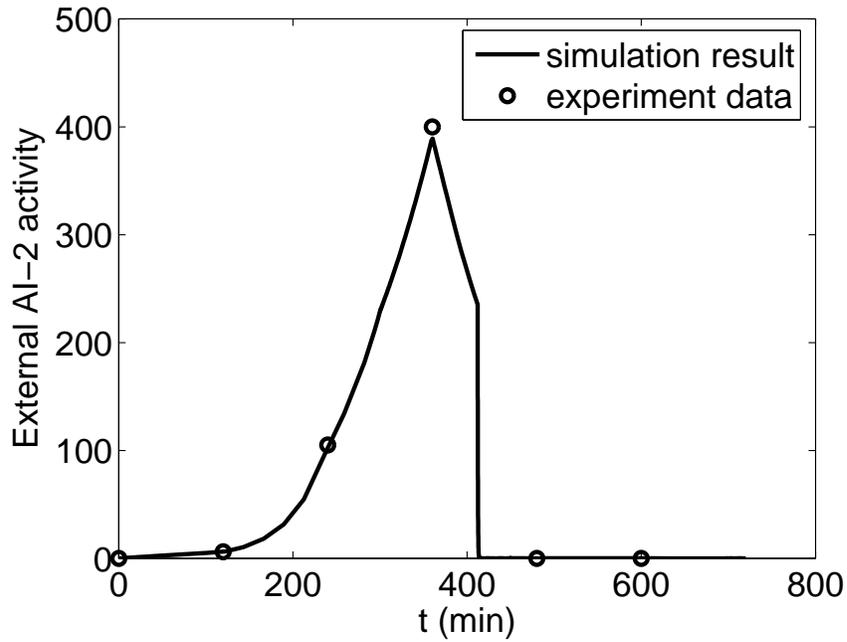


Figure 4.27: Deterministic simulation results of external AI-2 activity of wild type *E.coli*. Experiment data are from Wang *et al.* [32]

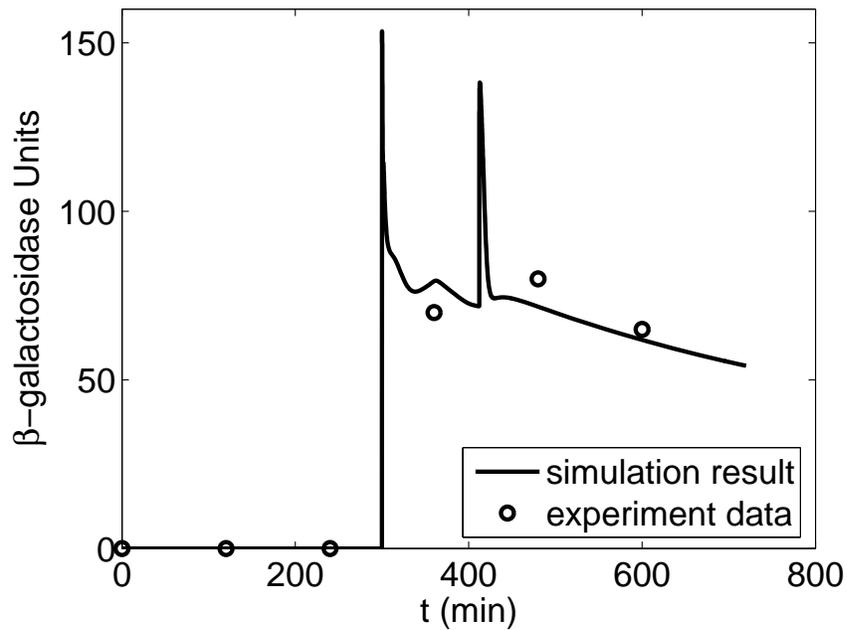


Figure 4.28: Deterministic simulation results of  $\beta$ -galactosidase units of wild type *E.coli*. Experiment data are from Wang *et al.* [32]

which is longer than the doubling time during exponential period.

The results of stochastic simulation are shown in Figs. 4.29 to 4.32. Fig. 4.29 and 4.30 are results which can provide a view of the whole time frame, while the results in Fig. 4.31 and 4.32 limit the time to between 390 and 490 minutes, which shows more detail. For external AI-2 activity as shown in Fig. 4.29 and 4.31, during the first several minutes the stochastic mean matches the deterministic result very well. After that, a difference between the stochastic mean and deterministic result appears. External AI-2 activity from the deterministic ODE simulation continues to decrease slowly and drops sharply to zero at around 415 minutes, while the mean activity from stochastic simulation does not have the sharp change but goes down gradually to zero. After 470 minutes, both stochastic and deterministic activities are equal to zero so they match again. For  $\beta$ -galactosidase units in Fig. 4.30 and 4.32, the deterministic result has a spike at around the time of 415 minutes, while the peak in stochastic mean is shorter and wider. And after 450 minutes, stochastic mean almost follows deterministic result again.

The difference between stochastic and deterministic results happens because LsrD numbers for deterministic and stochastic simulations reach their threshold in different ways. For deterministic simulation, LsrD number reaches the threshold at a fixed time, at which point a sharp drop in external AI-2 activity and a spike in  $\beta$ -galactosidase units appear. For the stochastic simulation, LsrD numbers in different cells reach the threshold at different times. The standard deviation of external AI-2 activity from stochastic simulation as shown in Fig. 4.31 can help us understand what happens. During the first several minutes, the standard deviation

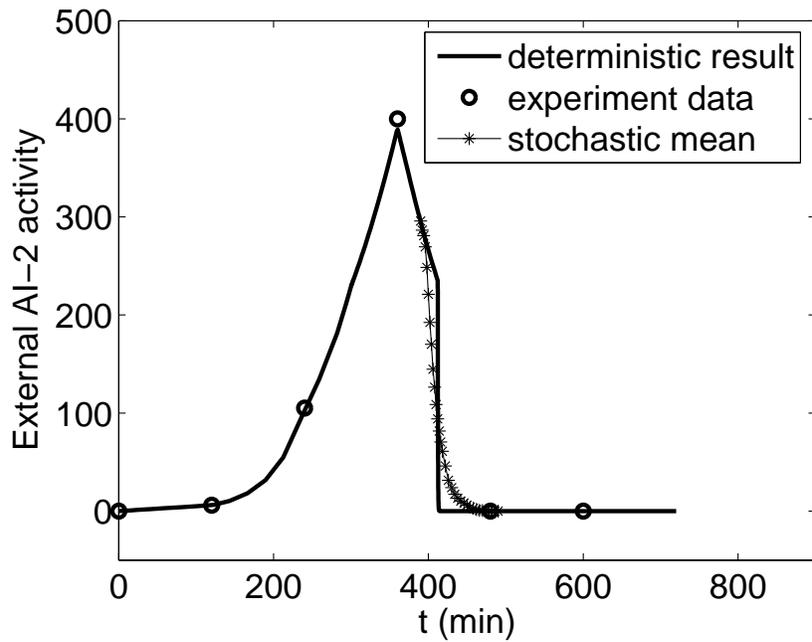


Figure 4.29: Results of external AI-2 activity in the whole time range from deterministic simulation and stochastic simulation.

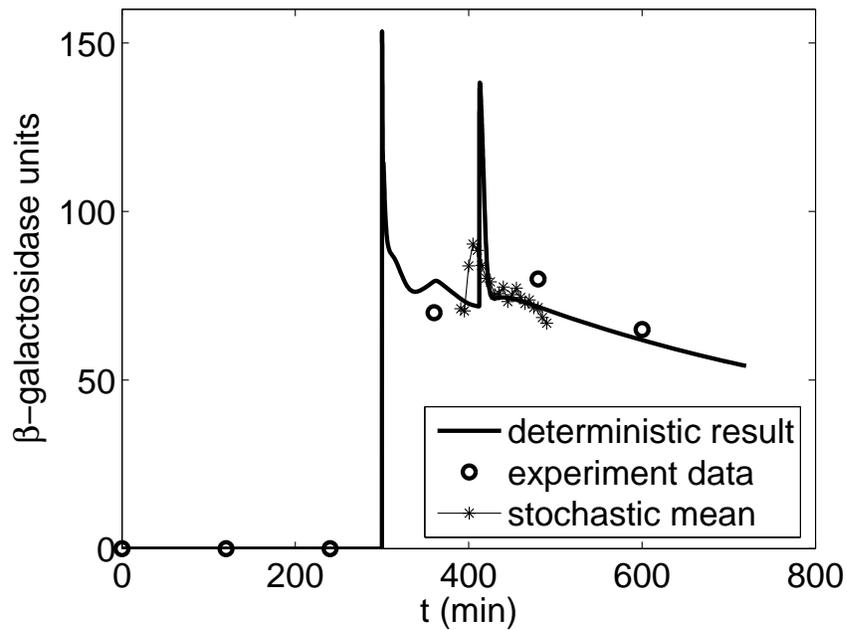


Figure 4.30: Results of  $\beta$ -galactosidase units in the whole time range from deterministic simulation and stochastic simulation.

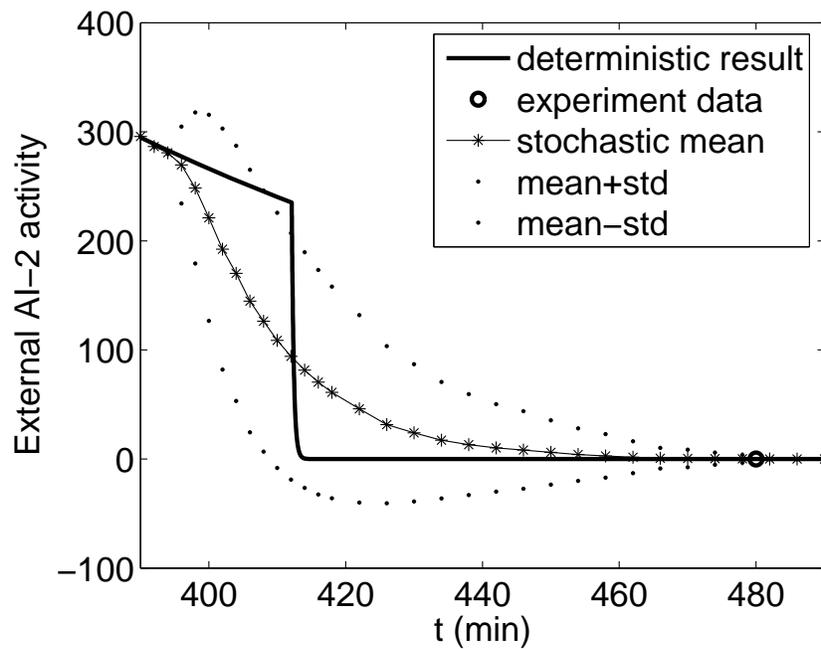


Figure 4.31: Results of external AI-2 activity in time range between 390 and 490 minutes from deterministic simulation and stochastic simulation.

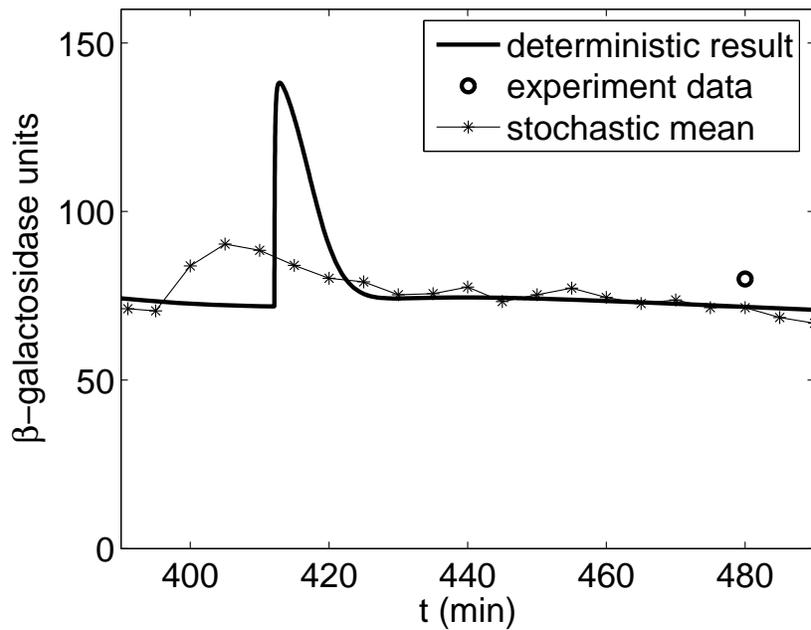


Figure 4.32: Results of  $\beta$ -galactosidase units in time range between 390 and 490 minutes from deterministic simulation and stochastic simulation.

is very small which means in all the cells LsrD numbers are lower than the threshold and thus only alternative AI-2 uptake happens. After that, in some cells LsrD numbers reach the threshold and activity becomes zero, so the standard deviation for the cell population becomes large. The more cells in which LsrD numbers reach the threshold, the larger the standard deviation. The standard deviation becomes the largest when both the number of cells with LsrD more the threshold and that with LsrD number less than the threshold. Then the standard deviation decreases as more cells have a LsrD number higher than the threshold. Finally after 470 minutes the standard deviation is very small again because in almost all the cells LsrD number reaches the threshold and AI-2 activity becomes zero.

#### 4.4 Discussion

We demonstrated the successful estimation of parameter values in the regulation network of AI-2 uptake in *E.coli*. As shown in Fig. 4.3, the wild type kinetic network of AI-2 uptake is very complex, and two main sets of experimental data, external AI-2 activity and *lsr* transcription level, are available. Rather than attempting to define a very large optimization problem, we chose an approach that takes advantage of experimental data available for different mutant cell types. In this way the kinetic networks are simplified as shown in Figs. 4.6 to 4.10. As a result, several parameter values can be obtained in a series of steps. In each step one type of mutant cells are considered and one set of experimental data are fitted. This allows the careful examination and understanding of the quality of the fit, which is

very important when one considers the rather limited number of experimental data. The final remaining unknown parameter values are obtained from experimental data of wild type cells. We introduced effective *lsr* transcription level and effective AI-2 uptake rate, so that the kinetic network can also be simplified and the two sets of experimental data can be considered separately. With all the parameters estimated, the simulation result of the whole network matches experimental data very well. The work here provides new thoughts for parameter estimation, where the optimization task is improved by utilizing biological rather than mathematical concepts, through the consideration of mutant cells.

Stochastic simulation for this network indicates that stochastic models can be valuable even when the system does not exhibit stochastic bistability. Although for monostable systems, stochastic means usually follow closely the deterministic paths. The results for the AI-2 uptake network have a significant difference. The presence of a threshold for the LsrD protein number needed to allow *lsr*-mediated uptake is supported by literature [88]. In deterministic simulation, at the time the threshold is reached, external AI-2 activity has a steep drop and *lsr* transcription level has a spike. In stochastic simulation, each cell reaches the threshold at different time, so for the cell population, the drop of external AI-2 activity is mild and there is no spike for *lsr* transcription level. Therefore if there is a threshold mechanism in a monostable system, the results from deterministic simulation and stochastic simulation can be significantly different and stochastic model is valuable as it better represents the type of experimental measurements we can expect from a cell population.

## Chapter 5

### Conclusion

It has been discovered in recent years that the stochasticity plays an important role in gene regulation networks. Such networks are composed of many molecular activities, which are basically stochastic due to thermal fluctuations [5]. Molecule numbers for many species are very small and reactions are relatively slow. Hence, the effects of stochasticity become significant and cannot be ignored [4]. The simulation of stochastic models can predict behavior distinct results from that of deterministic models. To further understand the impact of stochasticity on gene regulation networks, we have selected the quorum sensing system in *E.coli* as an example and examined stochastic effects in it.

#### 5.1 Protein Number Distribution

Stochasticity in bistable systems is important, because in this case cells with the same initial condition can evolve into two different groups. We used the regulation network of the *sdiA* gene expression as an example. This network has a positive-autoregulation which is mediated by signal molecules and thus may be a bistable system.

First the stochastic simulation of the network was performed. The result indicated that, from the same initial condition and after around fifty minutes, there

is a bifurcation among cell population: some cells have high numbers of protein molecules and others have low protein numbers.

Second, it was discovered that the signal-mediated positive-autoregulation network (full network) can be simplified into another network. The simplified network has relatively fast stochastic simulation and can be conveniently described with Fokker-Planck equation [12]. Therefore the results for the simplified network can be obtained quickly and confirmed with two different methods. We discussed the simplifications and the assumptions behind them, and found that the simplifications do not influence the bistable property and do not affect the shape of protein number distribution much. Therefore the simplified network can be used to predict the behavior of the full network.

Third, we have made a comparison between stochastic and deterministic simulation. From one initial condition, deterministic simulation follows a fixed path, while stochastic simulation has a bifurcation, and thus results in means different from the deterministic path and in large variances. These results show that deterministic simulation is not appropriate for bistable system and stochastic simulation is necessary.

Fourth, we found that the stochastic bistability of the full network can be influenced by signal molecules. When the signal molecule number is in a special range, there will be three steady states, two of which are stable. So the system is bistable. When the signal molecule number is outside this range, there will be only one steady state and the system becomes monostable.

## 5.2 Age Distribution

If the protein happens to be a key protein for cell division, this bifurcation may have an experimentally measurable effect on cell age distribution. SdiA protein can regulate the expression of *ftsQAZ*, and FtsQAZ protein is necessary for cell division. As a result, the SdiA protein may be considered as a key protein for cell division.

To relate the protein distribution to cell age distribution, we first calculate a function  $F(t)$ , which represents the probability that a cell has not divided at time  $t$ , based on the evolution of key protein number. The results from both Fokker-Planck equation and stochastic Petri nets indicate that the function  $F(t)$  can be written in a double-exponential form, when the key protein number has a bifurcation. This is the first attempt to construct a mathematical relation between the bifurcation of protein number and the function  $F(t)$ .

Based on  $F(t)$  function in this analytical form, we took the second step of calculating the age distribution of the cell population. The results indicate that if there is a bifurcation for key protein number, then there would be a significant fraction of very old cells in the cell population. This conclusion is not difficult to understand. With the bifurcation, the cells which evolve into high key protein numbers can divide quickly, while the cells which evolve into low key protein numbers would remain undivided for long times. The existence of small fraction of very old cells has been reported in the literature. But we first proposed here that the bifurcation of key protein number can be the explanation of this experimental phenomenon.

### 5.3 AI-2 uptake

The study of AI-2 mediated SdiA positive autoregulation networks showed that the signal molecule AI-2 can be a determining factor on the presence of bistability. We next considered the regulation network for the AI-2 uptake in *E.coli*.

First we obtained kinetic parameter values for this network by fitting a deterministic model to experimental data in the literature. We took an approach where data for mutant cells and effective data for wild type cells, corresponding to simplified network structures, were used to build the more complex wild type network step-by-step. Of course, finally the model matches the wild type cell measurements. Usually, improving the efficiency of parameter estimation is accomplished by improved optimization techniques. As optimization techniques have been thoroughly studied, the potential of this mathematical approach has become limited. Our approach is based on a biological concept which can open a new way to improve the efficiency of parameter estimation.

This network has a repression-derepression mechanism and is a monostable system. However there is still difference between the results of stochastic and deterministic simulation. The results of deterministic simulation show a step change in external AI-2 activity and a spike in transcription level, while in the results of stochastic simulation, the change is mild and the spike disappears. This difference comes from a threshold mechanism. The *lsr*-mediated AI-2 uptake requires sufficient LsrD protein, so we can propose this uptake cannot take place until the LsrD protein number reaches a threshold. In deterministic simulation, LsrD protein

number reaches the threshold at a fixed time. But in stochastic simulation, protein numbers can reach the threshold at different times in different cells, and thus the average changes are not as sharp as in deterministic simulation. The difference of these results indicates that stochastic models can also be necessary for a monostable system if it has a threshold.

## 5.4 Suggestions for Future work

We have used the quorum sensing system in *E.coli* as an example to study the stochasticity in gene regulation networks, and found that stochasticity can have significant impact on gene regulation networks for both bistable and monostable systems. But there are still open questions of interest for the future work.

In Chap. 2, we studied protein number distributions for the full network and the simplified network. We discussed the influence of signal molecule number on the bistability of the distributions. However, we did not discuss the influence of the kinetic parameters on the distributions. For the simplified network, the phase diagram by Kepler and Elston [12] provided a complete view of the influence of parameters on the protein number distribution. But no such work has been done for the full network. The SPN simulation we used is time-consuming and therefore trying many sets of different parameter values in the full network would be computationally very expensive. Erban *et al.* [95] proposed an equation-free approach which can perform a “bifurcation analysis”. With this approach we may be able to obtain a phase diagram of the full network, and distinguish the parameter values in which

the network has a stochastic bistability.

In Chap. 3, we developed a double exponential form for the alive probability function. This form was obtained by fitting to the simulation data and we provided a biological explanation. However we did not prove this form in a mathematical way. Further work may be done to provide a mathematical basis for the double exponential form. For simplified network, the alive probability function can be obtained from equation (3.4), which is from the Fokker-Planck equation. Equation (3.4) is a linear PDE and should be further studied for the simplified network. Finally we may be able to obtain an analytical solution for equation (3.4), which may provide justification for the double exponential form.

In Chap. 4, we stated that the regulation network of AI-2 uptake is a monostable system. Although simulation results support this statement, it would be useful to do a rigorous analysis for the network. The AI-2 uptake regulation network is very complex and its simulation is time-consuming. However, the equation-free approach by Erban *et al.* [95] may also be used here for the “bifurcation analysis” of this network. Hence this approach may confirm whether the network is monostable or not.

## Bibliography

- [1] U. Alon, *An introduction to systems biology: design principles of biological circuits* (Chapman & Hall/CRC, Boca Raton, FL, 2006), Chaps. 2.
- [2] C.V. Rao and A.P. Arkin, “Control motifs for intracellular regulatory networks,” *Annu. Rev. Biomed. Eng.* **3**, 391 (2001).
- [3] H.H. McAdams and A. Arkin, “Simulation of prokaryotic genetic circuits,” *Annu. Rev. Biophys. Biomol. Struct.* **27**, 199 (1998).
- [4] H.H. McAdams and A. Arkin, “Its a noisy business! Genetic regulation at the nanomolar scale,” *Trends in Genetics* **15**, 65 (1999).
- [5] C.V. Rao, D.M. Wolf and A.P. Arkin, “Control, exploitation and tolerance of intracellular noise,” *Nature* **420**, 231 (2002).
- [6] M.S. Ko, H. Nakauchi and N. Takahashi, “The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates. *EMBO J.* **9**, 2835 (1990).
- [7] H.H. McAdams and A. Arkin, “Stochastic mechanisms in gene expression,” *Proc. Natl. Acad. Sci. USA* **94**, 814 (1997).
- [8] A. Arkin, J. Ross and H.H. McAdams, “Stochastic kinetic analysis of developmental pathway bifurcation in phage  $\lambda$ -infected *Escherichia coli* cells,” *Genetics* **149**, 1633 (1998).
- [9] M.M. Domach and M.L. Shuler, “A finite representation model for an asynchronous culture of *Escherichia coli*,” *Biotechnol. Bioeng.* **26**, 877 (1984).
- [10] C.W. Gardiner, *Handbook of stochastic methods for physics, chemistry and the natural sciences* (Springer, Berlin, 1990), Chaps. 4.
- [11] J. Xing, H. Wang and G. Oster, “From Continuum Fokker-Planck Models to Discrete Kinetic Models,” *Biophysical Journal* **89**, 1551 (2005).
- [12] T.B. Kepler and T.C. Elston, “Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations,” *Biophysical J.* **81**, 3116 (2001).

- [13] D.T. Gillespie, “The chemical Langevin equation,” *J. Chemical Physics*. **113**, 297 (2000).
- [14] A.P.J. Jansen, An introduction to Monte Carlo simulations of surface reactions. <http://arxiv.org/abs/cond-mat/0303028> (2005).
- [15] P.J. Goss and J. Peccoud, “Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets,” *Proc. Natl. Acad. Sci. USA* **95**, 6750 (1998).
- [16] J.L. Peterson, *Petri net theory and the modeling of systems* (Prentice Hall, Englewood Cliffs, NJ, 1981).
- [17] W.H. Sanders, *Mobius User Manual. Version 1.5.0* (PERFORM Performability Engineering Research Group, University of Illinois at Urbana-Champaign, 2004).
- [18] R. Srivastava, M.S. Peterson and W.E. Bentley, “Stochastic kinetic analysis of the *Escherichia coli* stress circuit using  $\sigma^{32}$ -targeted antisense,” *Biotechnol. Bioeng.* **75**, 120 (2001).
- [19] T. Wilson and J.W. Hastings, “Bioluminescence,” *Ann. Rev. Cell. Dev. Biol.* **14**, 197 (1998).
- [20] M.B. Miller and B.L. Bassler, “Quorum sensing in bacteria,” *Ann. Rev. Microbiol.* **55**, 165 (2001).
- [21] W.C. Fuqua, S.C. Wintans and E.P. Greenberg, “Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators,” *J. Bacteriology.* **176**, 269 (1994).
- [22] J.M. Henke and B.L. Bassler, “Bacterial social engagements,” *Trends in cell biology.* **14**, 648 (2004).
- [23] B.A. Lazazzera and A.D. Grossman, “The ins and outs of peptide signaling,” *Trends Microbiol.* **6**, 288 (1998).
- [24] R. Hakenbeck and J.B. Stock, “Analysis of two-component signal transduction systems involved in transcriptional regulation,” *Methods Enzymol.* **273**, 281 (1996).

- [25] B.L. Bassler, M. Wright and M.R. Silverman, "Multiple signalling systems controlling expression of luminescence in *Vibrio harveyi*: sequence and function of genes encoding a second sensory pathway," *Mol. Microbiol.* **13**, 273 (1994).
- [26] K.C. Mok, N.S. Wingreen and B.L. Bassler, "*Vibrio harveyi* quorum sensing: a coincidence detector for two autoinducers controls gene expression," *EMBO J.* **22**, 870 (2003).
- [27] J.G. Cao and E.A. Meighen, "Purification and structural identification of an autoinducer for the luminescence system of *Vibrio harveyi*," *J. Biol. Chem.* **264**, 21670 (1989).
- [28] B.L. Bassler, E.P. Greenberg and A.M. Stevens, "Cross-species induction of luminescence in the quorum-sensing bacterium *Vibrio harveyi*," *J. Bacteriol.* **179**, 4043 (1997).
- [29] M.G. Surette, M.B. Miller and B.L. Bassler, "Quorum sensing in *Escherichia coli*, *Salmonella typhimurium*, and *Vibrio harveyi*: a new family of genes responsible for autoinducer production," *Proc. Natl. Acad. Sci. USA.* **96**, 1639 (1999).
- [30] S. Schauder, K. Shokat, M.G. Surette and B.L. Bassler, "The LuxS family of bacterial autoinducers: biosynthesis of a novel quorum-sensing signal molecule," *Mol. Microbiol.* **41**, 463 (2001).
- [31] M.G. Surette and B.L. Bassler, "Regulation of autoinducer production in *Salmonella typhimurium*," *Mol. Microbiol.* **31**, 585 (1999).
- [32] L. Wang, Y. Hashimoto, C.-Y. Tsao, J.J. Valdes and W.E. Bentley, "Cyclic AMP (cAMP) and cAMP receptor protein influence both synthesis and uptake of extracellular autoinducer-2 in *Escherichia coli*," *J. Bacteriol.* **187**, 2066 (2005).
- [33] M.G. Surette and B.L. Bassler, "Quorum sensing in *Escherichia coli* and *Salmonella typhimurium*," *Proc. Natl. Acad. Sci. USA.* **95**, 7046 (1998).
- [34] M.P. DeLisa, J.J. Valdes and W.E. Bentley, "Mapping stress-induced changes in autoinducer AI-2 production in chemostat-cultivated *Escherichia coli* K-12," *J. Bacteriol.* **183**, 2918 (2001).
- [35] A.L. Beeston and M.G. Surette, "pfs-dependent regulation of autoinducer 2 production in *Salmonella enterica* serovar Typhimurium," *J. Bacteriol.* **184**, 3450 (2002).

- [36] L. Wang, J. Li, J.C. March, J.J. Valdes and W.E. Bentley, "luxS-Dependent Gene Regulation in *Escherichia coli* K-12 Revealed by Genomic Expression Profiling," *Journal of Bacteriology*, **187**, 8350 (2005).
- [37] B.M.M. Ahmer, "Cell-to-cell signalling in *Escherichia coli* and *Salmonella enterica*," *Mol. Microbiol.* **52**, 933 (2004).
- [38] X.D. Wang, P.A.J. de Boer and L.I. Rothfield, "A factor that positively regulates cell division by activating transcription of the major cluster of essential cell division genes of *Escherichia coli*," *EMBO J.* **10**, 3363 (1991).
- [39] Y. Wei, A.C. Vollmer and R.A. LaRossa, "In vivo titration of mitomycin C action by four *Escherichia coli* genomic regions on multicopy plasmids," *J. Bacteriol.* **183**, 2265 (2001).
- [40] S. Rahmati, S. Yang, A.L. Davidson and E.L. Zechiedrich, "Control of the AcrAB multidrug efflux pump by quorum sensing regulator SdiA," *Mol. Microbiol.* **43**, 677 (2002).
- [41] D.M. Sitnikov, J.B. Schineller and T.O. Baldwin, "Control of cell division in *Escherichia coli*, regulation of *ftsQA* involves both *rpoS* and SdiA-mediated autoinduction," *Proc. Natl. Acad. Sci. USA.* **93**, 336 (1996).
- [42] K. Kanamaru, I. Tatsuno, T. Tobe and C. Sasakawa, "SdiA, an *Escherichia coli* homologue of quorum-sensing regulators, controls the expression of virulence factors in enterohaemorrhagic *Escherichia coli* O157:H7," *Mol. Microbiol.* **38** 805 (2000).
- [43] P.V. Dunlap and J.M. Ray, "Requirement for autoinducer in transcriptional negative autoregulation of the *Vibrio fischeri luxR* gene in *Escherichia coli*," *J. Bacteriol.* **171**, 3549 (1989).
- [44] G.S. Shadel and T.O. Baldwin, "The *Vibrio fischeri* LuxR protein is capable of directional stimulation of transcription and both positive and negative regulation of the *luxR* gene," *J. Bacteriol.* **173**, 568 (1991).
- [45] J. Garcia-Lara, L.H. Shang and L.I. Rothfield, "An extracellular factor regulates expression of *sidA*, a transcriptional activator of cell division genes in *Escherichia coli*," *J. Bacteriol.* **178**, 2742 (1996).
- [46] M.P. DeLisa, C.-F. Wu, L. Wang, J.J. Valdes and W.E. Bentley, "DNA microarray-based identification of genes controlled by autoinducer 2-stimulated quorum sensing in *Escherichia coli*," *J. Bacteriol.* **183**, 5239 (2001).

- [47] D.L. Chopp, M.J. Kirisits and M.R. Parsek, “A mathematical model of quorum sensing in a growing bacterial biofilm,” *Journal of Industrial Microbiology & Biotechnology*. **29**, 339 (2002).
- [48] A.U. Viretta and M. Fussenegger, “Modeling the quorum sensing regulatory network of human-pathogenic *Pseudomonas aeruginosa*,” *Biotechnol. Prog.* **20**, 670 (2004).
- [49] A.G. Daga, *Quorum sensing in silico: kinetic modeling in Escherichia coli through stochastic Petri nets. Master of Science, Dept. of Chem. Eng., Univ. of Maryland, College Park (2001)*.
- [50] A.B. Goryachev, D.J. Toh, K.B. Wee, T. Lee, H.B. Zhang and L.H. Zhang, “Transition to quorum sensing in an *Agrobacterium* population: a stochastic model,” *PLoS Comput. Biol.* **1**, 265 (2005).
- [51] J. Li, L. Wang, Y. Hashimoto, C.-Y. Tsao, T.K. Wood, J.J. Valdes, E. Zafiriou and W.E. Bentley. “A Stochastic Model of *Escherichia coli* AI-2 Quorum Signal Circuit Reveals Alternative Synthesis Pathways,” *Molecular Systems Biology*. doi:10.1038/msb4100107 (December 2006).
- [52] J.M. Raser and E.K. OShea. “Noise in gene expression: origins, consequences, and control,” *Science* **309**, 2010 (2005).
- [53] B. Zhou, D. Beckwith, L.R. Jarboe and J.C. Liao. “Markov chain modeling of pyelonephritis-associated pili expression in uropathogenic *Escherichia coli*,” *Biophysical Journal*. **88**, 2541 (2005).
- [54] F. J. Isaacs, J. Hasty, C.R. Cantor and J.J. Collins. “Prediction and measurement of an autoregulatory genetic module,” *Proc. Natl. Acad. Sci. USA*. **100**, 7714 (2003).
- [55] A. Becskei, B. Seraphin and L. Serrano, “Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion,” *EMBO J.* **20**, 2528 (2001).
- [56] J.E. Ferrell, “Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability,” *Current Opinion in Chemical Biology*. **6**, 140 (2002).
- [57] M. Ptashne, *A genetic switch: Phage  $\lambda$  and higher organisms, 2nd edition* (Cell Press and Blackwell Scientific Publications, Cambridge, MA, 1992).

- [58] M.T. Record Jr., W.S. Reznikoff, M.L. Craig, K. McQuade, and P.J. Schlux. *Escherichia coli* RNA polymerase ( $E\sigma^{70}$ ), promoters, and the kinetics of the steps of transcription initiation, in *Escherichia coli and Salmonella*, 2nd ed, vol.2. F.C. Neidhart, R. Curtiss, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter, and H.E. Umbarger, editors (ASM Press, Washington DC, 1996), 792-816.
- [59] S. Pedersen and S. Reeh, "Functional mRNA half lives in *Escherichia coli*," *Mol. gen. Genet.* **166**, 329 (1978).
- [60] A. Varshavsky, "The N-end rule: functions, mysteries, uses," *Proc. Natl. Acad. Sci. USA* **93**, 12142 (1996).
- [61] W. Chen, J.E. Bailey, and S.B. Lee. "Molecular design of expression systems: comparison of different repressor control configurations using molecular mechanism models," *Biotechnol. Bioeng.* **38**, 679 (1991).
- [62] L.S. Weinberger, J.C. Burnett, J.E. Toettcher, A.P. Arkin and D.V. Schaffer, "Stochastic gene expression in a lantiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity," *Cell.* **122**, 169 (2005).
- [63] J.J. Tyson and B. Novak, "Regulation of the Eukaryotic Cell Cycle: Molecular Antagonism, Hysteresis, and Irreversible Transitions," *J Theor Biol.* **210**, 249 (2001).
- [64] W. Xiong and J.E.Jr. Ferrell, J.E.Jr. "A positive-feedback-based bistable 'memory module' that governs a cell fate decision," *Nature.* **426**, 460 (2003).
- [65] E. Bi and J. Lutkenhaus, "FtsZ ring structure associated with division of *Escherichia coli*". *Nature.* **354**, 161 (1991).
- [66] X. Ma, D.W. Ehrhardt and W. Margolin, "Colocalization of cell division proteins FtsZ and FtsA to cytoskeletal structures in living *Escherichia coli* cells by using green fluorescent protein," *Proc. Natl. Acad. Sci. USA.* **93**, 12998 (1996).
- [67] L. Rothfield, S. Justice and J. Garcia-Lara, "Bacterial cell division". *Annu. Rev. Genet.* **33**, 423 (1999).
- [68] C.E. Helmstetter and D.J. Cummings, "Bacterial synchronization by selection of cells at division," *Proc. Natl. Acad. Sci. USA.* **50**, 767 (1963).
- [69] C.E. Helmstetter and D.J. Cummings, "An improved method for the selection of bacterial cells at division," *Biochim. Biophys. Acta.* **82**, 608 (1964).

- [70] S. Copper, *Bacterial growth and division* (Academic Press, San Diego, CA, 1991).
- [71] S. Copper, "Analysis of bacterial division cycle using membrane-elution method," *Methods molec. Genet.* **3**, 234 (1994).
- [72] C.D. Kelly and O. Rahn, "The growth rate of individual bacterial cells," *J. Bacteriol.* **23**, 147 (1932).
- [73] A.L. Koch, The variability and individuality of the bacterium. In: *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* (F.C. Neidhardt, ed, American Society for Microbiology, Washington, DC, 1987), pp. 1606-1614.
- [74] E.O. Powell, "Growth rate and generation time of bacteria with special reference to continuous culture," *J. gen. Microbiol.* **15**, 492 (1956).
- [75] A.L. Koch and M. Schaechter. "A model for statistics of the cell division process," *J. Gen. Microbiol.* **29**, 435 (1962).
- [76] M.A. Hjortso, "Solution and properties of age population balance models which assume discrete division ages," *J. Biotechnol.* **42**, 271 (1995).
- [77] J.-J. Liou, F. Sreenc and A.G. Fredrickson, "Solution of population balance models based on a successive generations approach," *Chem. Eng. Sci.* **52**, 1529 (1997).
- [78] D.G. Kendall, "On the role of variable generation time in the development of a stochastic birth process," *Biometrika.* **35**, 316 (1948).
- [79] D.G. Kendall, "On the choice of mathematical models to represent normal bacterial growth," *J. R. Stat. Soc. B.* **14**, 41 (1952).
- [80] J.A. Smith and L. Martin, "Do cells cycle?" *Proc. Natl. Acad. Sci. USA.* **70** 1263 (1973).
- [81] R.F. Brooks, D.C. Bennett and J.A. Smith, "Mammalian cell cycles need two random transitions," *Cell.* **19**, 493 (1980).
- [82] J.D. Keasling, H. Kuo and G. Vahanian, "A monte carlo simulation of the *Escherichia coli* cell cycle," *J. theor. Biol.* **176**, 411 (1995).

- [83] N.V. Mantzaris, J.-J. Liou, P. Daoutidis and F. Srieenc, “Numerical solution of a mass structured cell population balance model in an environment of changing substrate concentration,” *J. Biotechnol.* **71**, 157 (1999).
- [84] M.A. Hjortso and J. Nielsen, “Population balance models of autonomous microbial oscillations,” *J. Biotechnol.* **42**, 255 (1995).
- [85] D. Ramkrishna, “Statistical models of cell populations,” *Adv. Biochem. Engng.* **11**, 1 (1979).
- [86] M.E. Taga, J.L. Semmelhack and B.L. Bassler, “The LuxS-dependent autoinducer AI-2 controls the expression of an ABC transporter that functions in AI-2 uptake in *Salmonella typhimurium*,” *Mol. Microbiol.* **42**, 777 (2001).
- [87] M.E. Taga, S.T. Miller and B.L. Bassler, “Lsr-mediated transport and processing of AI-2 in *Salmonella typhimurium*,” *Mol. Microbiol.* **50**, 1411 (2003).
- [88] K.B. Xavier and B.L. Bassler, “Regulation of Uptake and Processing of the Quorum-Sensing Autoinducer AI-2 in *Escherichia coli*,” *J. Bacteriol.* **187**, 238 (2005).
- [89] P. Mendes and D.B. Kell, “Non-linear optimization of biochemical pathways: Applications to metabolic engineering and parameter estimation,” *Bioinformatics.* **14**, 869 (1998).
- [90] K. Chen, A. Csikasz-Nagy, B. Gyorffy, J. Val, B. Novak and J.J. Tyson, “Kinetic analysis of a molecular model of the budding yeast cell cycle,” *Mol. Biol. Cell* **11**, 369 (2000).
- [91] J.W. Zwolak, J.J. Tyson and L.T. Watson, “Parameter Estimation for a Mathematical Model of the Cell Cycle in Frog Eggs,” *Journal of computational biology.* **12** 48 (2005).
- [92] M. Ronen, R. Rosenberg, B. Shraiman and U. Alon, “Assigning numbers to arrows: Parameterizing a gene regulation network by accurate expression kinetics,” *Proc. Natl. Acad. Sci. USA.* **99**, 10555 (2002).
- [93] C.G. Moles, P. Mendes and J.R. Banga, “Parameter Estimation in Biochemical Pathways: A Comparison of Global Optimization Methods,” *Genome Res.* **13**, 2467 (2003).
- [94] W.E. Bentley and D.S. Kompala, “A novel structured kinetic modeling approach for analysis of plasmid instability in recombinant bacterial fermentations,” *Biotechnology and Bioengineering* **33**, 49 (1989).

- [95] R. Erban, I.G. Kevrekidis, D. Adalsteinsson and T.C. Elston, “Gene regulatory networks: A coarse-grained, equation-free approach to multiscale computation,” the Journal of Chemical Physics **124**, 084106 (2006)