Modelling signaling pathways on diverse scales

Zhu, Kaicheng

2017

Zhu, K. (2017). Modelling signaling pathways on diverse scales. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/73070>

<https://doi.org/10.32657/10356/73070>

NANYANG
TECHNOLOGICAL
UNIVERSITY

**MODELLING SIGNALING PATHWAYS ON
DIVERSE SCALES**

KAICHENG ZHU

**INTERDISCIPLINARY GRADUATE SCHOOL
NANYANG INSTITUTE OF TECHNOLOGY IN HEALTH
AND MEDICINE (NITHM)**

2017

MODELLING SIGNALING PATHWAYS ON DIVERSE SCALES

KAICHENG ZHU

**Interdisciplinary Graduate School
Nanyang Institute of Technology in Health and
Medicine (NITHM)**

**A thesis submitted to the Nanyang Technological
University in partial fulfilment of the requirement for
the degree of
Doctor of Philosophy**

2017

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

Aug 14th 2017
.....
Date

Karcheng ZHU
.....
Student Name

Abstract

The dynamic behavior of cell is a broad topic covering wide scope of subjects of varying scales and species. According to the great complexity and diversity involved in understanding cellular behavior, it is necessary to combine the experimental observations from different methods and scales to build up a thorough picture. Modeling techniques originating from physics and mathematics have helped to solve puzzles for complex systems over diverse space and time scales, including numerous biological related systems. In this thesis, we aim to apply the theoretical modeling methods to several biological systems scaling from molecular to ecological levels to develop a quantitative understanding of the interaction and dynamics involved in these subjects. On molecular level, we model the tubulin protein dimer as a feedback control system to show the rich dynamics ranging from picoseconds to hundreds of nanoseconds, as well as the sensitivity of such dimer structure on surrounding biophysical environment. Based on the experimental results of bacteria related study, we mainly focus on the quorum sensing pathway analysis to identify the key components and the robust topological motifs of the interaction network. We also analyze the bacterial ecology system of Ace lake of large scale and high complexity by reaction-diffusion theory and figure out the reason of spatial stratification of different bacteria species. With these modeling works, we are able to further both qualitative and quantitative understanding of molecular interactions and large scale observation of cellular behaviors.

Acknowledgements

I would like to thank my supervisors, Prof. Haibin SU and Prof. Zhaoxun LIANG, and my mentor, Dr. Sengbeng YEO, for their inspiring guidance of my PhD project. I appreciate the help from the discussion with my colleagues, Kathiresan NATARAJAN, Zhi PING, and Chaodong ZHANG. This PhD project is supported by Nanyang Institute of Technology in Health and Medicine (NITHM).

Table of Contents

Abstract	i
Acknowledgements	iii
Table of Contents	v
Table Captions	ix
Figure Captions	xi
Abbreviations	xv
 Chapter 1 Introduction	 1
1.1 Hypothesis /Problem Statement	2
1.2 Objectives and Scope	3
1.3 Dissertation Overview	4
1.4 Findings and Outcomes/Originality	6
 Chapter 2 Literature Review	 7
2.1 Overview	8
2.2 Biofilm	8
2.3 Quorum Sensing	11
2.3.1 Basic Mechanism of Quorum Sensing	12
2.3.2 Specific Quorum Sensing Systems	14
2.4 Microbial Stratification in Meromictic Lake	17
2.4.1 Meromictic Lake	17
2.4.2 Microecology in Meromictic Lake	19

2.5	Microtubule and Tubulin	20
2.6	Questions to answer	21
Chapter 3	Modelling Methodology	23
3.1	A Modelling View	24
3.1.1	Model Architecture	26
3.1.2	Deterministic Model	27
3.1.3	Stochastic Model.....	28
3.2	Mass Action Rate	30
3.3	Michaelis-Menten Mechanism.....	30
3.4	Reaction-Diffusion Theory	32
3.5	Stability Theory and Sensitivity Analysis.....	33
3.5.1	Functional Stability Theory	34
3.5.2	Sensitivity and Robustness of Biological Networks	36
3.6	Parameter Estimation	38
3.7	Numerical Simulation	39
Chapter 4	Gene Regulation Network Analysis of Quorum Sensing.....	43
4.1	Minimal Gene Regulation Network Dynamics.....	44
4.2	Stability of Gene Regulation Network.....	49
4.3	Modelling Quorum Sensing Network of Agrobacteria.....	51
4.4	Conclusion	55
Chapter 5	Modelling Tubulin Dimer Dynamics	57
5.1	Position Feedback Control Model of Sensor-Actuator System	58
5.2	Coupled Dynamic Model of Tubulin Dimer	58
5.3	Complexity Analysis of Tubulin Dimer Dynamics	60
5.4	Dynamical Robustness of Tubulin Dimer.....	63
5.5	Conclusion	65

Chapter 6	Reaction-Diffusion Model of Bacterial Stratification	67
6.1	Bacteria Distribution in Ace Lake	68
6.2	Metabolic Reaction Model for Bacteria-Nutrient Dynamics	69
6.3	Reaction-Diffusion Theory for a Microbial Relay Model	71
6.4	Conclusion	75
Chapter 7	Discussion and Future Work	77
7.1	Discussion	78
7.1.1	Discussion on Quorum Sensing Network Analysis	78
7.1.2	Discussion on Dynamic Model for Tubulin Dimer	79
7.1.3	Discussion on Microbial Stratification Model of Ace Lake	80
7.2	Future Work	81
7.2.1	Theoretical Improvement on Modelling Work	81
7.2.2	Prediction for Future Experimental Work	83
References		85
Publications		89

Table Captions

Table 3.1 Reaction propensity for Gillespie stochastic method. The concentrations come in the unit of molecules per cell.

Figure Captions

Figure 2.1 Biofilm development process 1: Initial attachment. Bacteria attach to surface and this initial attachment is reversible. 2: Irreversible attachment. In this state, new cells are produced and bacteria form monolayer biofilm. 3: Vertical development. Cell aggregation evolves into multilayer biofilm. 4: Maturation. Cells in mature biofilms present very different properties comparing with planktonic cells. Most biofilms are heterogeneous and resistant to antibiotics, antimicrobial agents and reduction of nutrients. 5: Dispersal, dispersal is an initiative process by the bacteria. (Image Credit: D.Davis).

Figure 2.2 Structures of homoserine lactone autoinducers produced by different Gram-negative bacteria. The top figure shows the typical structure of an AHL molecule. The bottom one lists the different structures of R-group in different bacteria.

Figure 3.1 The general modeling point of view of biological study. The experimental observations provide direct support for parameter evaluation and estimation, whereas numerical study offer explanation and prediction for current and future experiments. Model architecture is the presentation of the key elements and interactions of focused subject and analytic methods help to identify the robust motif of the model and solve the causality of numerical or experimental events.

Figure 3.2 Substrate dependence of reaction rate from Michaelis-Menten mechanism. The M-M kinetics describes the enzymatic reaction process generating a non-linear rate-concentration relation. The curve shows a saturation pattern of concentration dependence that when substrate concentration increases to infinite there is a saturation limit value of production rate. The value of M-M constant K_m can be examined directly from

experimental measurement of this curve, because when the reaction rate reaches half of the saturation maximum, the substrate concentration equals K_m .

Figure 4.1 General transfers of genetic information suggested by the central dogma. Assuming the DNA level stays constant, we just take these two main pathways into consideration for the models in this report.

Figure 4.2 Reaction networks of the three gene expression models. (a) shows the reactions involved in the minimal gene expression model, and (b) and (c) shows those from the protein monomer-TF model and the protein dimer-TF model.

Figure 4.3 The time evolutions of the three gene expression models. The plot is based on the ODE systems of their reaction networks. (a), (b) and (c) present the increase patterns of mRNAs, protein monomers and protein dimers. (a) and (b) show a clear similarity of mRNA and protein production between the minimal (no TF) model and the dimer-TF model at the beginning.

Figure 4.4 The solution for steady protein concentration in the dimer-TF model. Function y is the characteristic equation of protein level and the steady solutions are intersection points of curve y and the horizontal line $y=0$. The blue line is plotted with the same parameter values as in the above, and the green line is based on a modification of the mRNA transcription rate and degradation rate. After changing the parameters, we find that the system experiences a bifurcation from being monostable to being bistable, because two of the three steady solutions on green line are stable.

Figure 4.5 Quorum sensing network of an octopine-type *Agrobacterium* Ti plasmid. AAI is the key autoinducer of *Agrobacteria* quorum sensing network,

and the TraR^* is the regulator protein monomer, whereas the TraRd represents the dimer form of the regulator.

Figure 4.6 3D structure of the TraR –DNA complex. Domains in the two monomers are shown in different colours (light/dark orange and light/dark green respectively), the DNA is coloured blue and the OOHL is red. The N terminus and C terminus of the lower subunit are also labelled.

Figure 4.7 Transition to quorum sensing in the ODE system based on intracellular dynamics. The blue line on the bottom represents the off state of the QS switching, while the green line denotes the on state of the network.

Figure 4.8 The nullclines of the dynamic system with and without the anti-activator TraM . (a) is based on the exact parameter values of the work by Goryachev et al.¹⁸ with the TraM functioning as the anti-activator shown in figure 4.6. (b) shows the nullcline behavior with modified traR velocity. The value of y function at zero concentration in (b) is positive.

Figure 5.1 Scheme of feedback control model for tubulin heterodimer. The input signal stands for the surrounding biochemical environment and events, and the output signal is the corresponding mechanical states of the dimer. The beta subunit works as a signal sensor to track the input, while the alpha subunit plays the role of actuator to provide the feedback control.

Figure 5.2 Power spectrums of tubulin dynamics under different coupling strengths. From(a) to (h), coupling strengths decrease in the series of 5, 3, 1, 0.5, 0.3, 0.1, 0.05, 0.03, and 0. There is no coupling in (i), which corresponds to a decoupled case. The blue lines show the power spectrum of the beta subunit while the green lines denote the alpha subunit.

Figure 5.3 The sensitivity function of frequency and coupling coefficients. The top figure denotes the sensitivity function of 0 order (position of state) fluctuation and the bottom figure shows that of 1 order (momentum of state) fluctuation. Both present similar pattern for gamma dependence: during low frequency ($<10^8$ s) scale, the coupling raises the sensitivity, and when the signal fluctuation is around nanosecond scale, the increasing gamma creates a filtering channel to decrease the effect of these signal fluctuations.

Figure 6.1 A view of Ace lake environment condition. This photo is cited from ref 1. It shows the ice-covered state of Ace lake.

Figure 6.2 The qualitative concentration curves against depth. The blue curve denotes the primary resource level and the red curve denotes the secondary resource level. The first saturation layer takes the depth from 0 to 1 meter, while the second saturation layer takes from 1 to 2 meter.

Abbreviations

M-M Mechanism	Michaelis-Menten Mechanism
MT	Microtubule
ODE	Ordinary Differential Equation
PDE	Partial Differential Equation
R-D	Reaction-Diffusion
SDE	Stochastic Differential Equation
TF	Transcription Factor

Chapter 1

Introduction

Signaling pathways describe the cellular interactions on molecular level, which provide the fundamental information of cell dynamics. Because even the signaling pathways of one single cell is coupled with great complexity, it is difficult to completely address their influence on cell behaviour with current experimental techniques. We aim to apply modern theoretical methods to analyse the dynamics of several biological systems from microscale to macroscale for a systematic view of quantitative biology study. In this chapter, we explain our motivations, targets and major findings.

1.1 Hypothesis/Problem Statement

Cell, as the fundamental unit of most biological entities, shows rich dynamics and great complexity in modern biological researches. In a bottom-up point of view, to understand cell behaviour is the first step to solve the puzzles for biological systems of higher order of complexity. Since the recent developments in experimental cell biology have provided us with enormous new observations, the bacterial cells are not as simple as they were thought to be (1, 2, 3). In direct experimental measurements, it is always a challenge to keep both complexity and dynamics for targeted cells, but theoretical methodology provides us with the chance to model the complex cellular dynamics for diverse systems from microscale to macroscale.

Biological system is such a general and abstract concept that it may refer to either a tiny one as small as single cell system within micrometers, or a huge one as large as oceanic ecology. It is a major mission for theorists to explain the collective cell dynamics over such diverse scales. Since the very early of last century, people have witnessed rapid development in theoretical science, especially statistical physics and applied mathematics. These advancements have already helped to solve physical problems at different scales, and now the new challenge is to apply these methods for biology (4). Cells regulate themselves by intracellular dynamics and extracellular responses, in other words, the cellular signaling pathways. Thus, the key to understand cellular dynamics is theoretical modeling of these signaling networks.

Some newly reported bacteria behaviors in biofilms are correlated with quorum sensing mechanism (3, 5). To understand this basic but complex issues, it is essential to conduct molecular level study of bacterial dynamics and establish mathematical models for analytical calculations and predictions (6, 7). There are large number of biophysical and biochemical reactions involved in bacterial cells, which makes it hardly possible to exactly solve the complete kinetics of every elements; we need to find the dominant patterns of bacterial signaling

pathways that can explain the key function of the cells. Then, there is one more step further down to the molecular level: what is the crucial molecular structure responsible for this signaling function? To understand this relation between molecular structure and biological function, we are required to apply theoretical molecular methods for specific targets.

On the other hand, it is necessary for us to take the large-scale behavior of microbe into consideration. There are plenty of bacteria and viruses living on earth, especially in the oceans and lakes (8). These aqua environments provide ideal systems for theoretical modelling collective behavior of microbe. We plan to test the modelling techniques in these systems to link the large-scale cellular behavior with microscale signaling pathways.

1.2 Objectives and Scope

The aim of this research project is building mathematical models for cell regulation networks from intracellular and extracellular interactions on different scales: protein scale, cell scale, and ecology scale. This will lead to better understanding of bacterial life and help to propose methods of related anti-bacteria therapeutics.

On protein scale, we focus on the famous tubulin protein. It is the basic building block of microtubules, a functional organelle involved in cell motion and shaping. Tubulin monomers are not able to assemble themselves into a growing microtubule. Dimerization is an inevitable step before microtubule assembly. We plan to model the dynamics of tubulin dimer under different states to explain the biological advantage of this dimerization process.

On cell scales, we concentrate on the bacterial cells in biofilm. Biofilms are ubiquitous and essential for enormous bacterial colonies. They form prosperous environment for bacterial development and dispersal. There are diverse biological interactions involved in a typical biofilm system, from intracellular

biochemical reactions to intercellular signaling pathways and interspecies communications. Quorum sensing, the mechanism of cell regulation based on population density, plays an important role in explaining cell behavior (1, 3). By modeling quorum sensing in biofilms, we can search for and identify the crucial parts of the signaling networks, and even make testable predictions. With a well-established mathematical model (9, 10), not only can we have a deep understanding of how cells work in biofilm, but also find artificial methods to regulate cell behavior. Both kinetics and dynamics of signaling molecules will be modeled in this project, so that we can study the cell regulation system in biofilm in both qualitative and quantitative ways. We will also simulate cellular response to biological noise and external disturbance to make comparison with experimental results (11). On ecology scale, we study the microbe ecology in meromictic lakes. In this specific type of aqua system, the large-scale bacterial distribution patterns have been reported. By modeling cell dynamics with reaction-diffusion mechanism, we aim to explain this spatial pattern and predict future dynamics of this huge bacterial system.

1.3 Dissertation Overview

Chapter 1 provides a rationale for the research and outlines the goals and scope.

Chapter 2 reviews the literature concerning the background and recent developments in biofilm and quorum sensing system, microbes in meromictic lakes, and the microtubule and tubulin protein. This chapter provides a general and basic introduction to the targeted problems of following chapters.

Chapter 3 presents the major theoretical methods applied in the following chapters. We introduce the basic assumption and general formation of each method and discuss the principles and advantages of both analytic and

numerical techniques involved in this work. We introduce the basic mass action law and rate equations of biochemical reactions, the Michaelis-Menten kinetics about enzymatic catalysis, the reaction-diffusion theory, and the sensitivity analysis of dynamical models.

Chapter 4 discuss our study of gene regulation network of quorum sensing system. We model the quorum sensing network on molecular biology level and develop theoretical method to identify the critical component of such networking behaviour.

Chapter 5 discuss the modeling work of microecology in meromictic lakes. By modelling the metabolic dynamics of bacteria and the diffusive motion of cell chemotaxis, we explain the observed spatial pattern of microbe distribution on meter scales.

Chapter 6 elaborates our theoretical analysis on tubulin dimer dynamics. In this chapter, we create a feedback control model of coupled system to describe the robust feature of tubulin dynamics. We compared the mathematical modelling result with that from molecular dynamics simulation and discuss the generality of our method.

Chapter 7 is the overall discussion of the whole thesis. The modeling methods and results are discussed, as well as future applications of our modeling techniques. We provide several important predictions from the modeling analysis and mention plausible improvement of our methods to deal with the complex system with more available data.

1.4 Findings and Outcomes/Originality

Finding 1: We find that the dimerization of regulator protein is the key process for the switching behaviour of quorum sensing network of Agrobacteria and so on. This result can be generalized for other bacterial species with similar theoretical method.

Finding 2: The stratification of microbe distribution in meromictic lakes is the result of reaction-diffusion mechanism. So we can predict bacterial spatial distribution by analysing their metabolic networks with the help of our model.

Finding 3: With modeling analysis, we explain the complex dynamics and robustness of different tubulin dimer states. The GTP-bound state is more robust than the GDP-bound state, which supports its advantage as the building block of microtubules.

Chapter 2

Literature Review

Cell, the fundamental functional biological unit of all living organisms, remains a mystery for its complexity and diversity. Over the past decades, researchers have been building new and inspiring techniques to probe and analyze cellular structures and their functions. Thanks to the development of biofilm research, we can go one step further in observing and testing novel bacterial cell behaviors; and with the help of molecular experimental techniques, we are now able to see a clear view of cellular signaling pathways. However, it is still challenging to understand the dynamical behavior of cell over different space and time scales because of the limitation of experimental techniques.

2.1 Overview

Cell regulation network consists of both intercellular signaling and intracellular biochemical reactions. (1, 2, 12) System-level cell dynamic study is essential to identify the underlying mechanism of cell behavior and develop related biology research. (5, 13, 14) The dynamic pathway involves numerous biophysical interactions and biochemical reactions to form a huge network system, so mathematical model plays an indispensable role in analyzing different levels and parts of the biological systems, which naturally gives a systematic representation of the cell networking.

2.2 Biofilm

Biofilms are bacteria colonies that attach to surfaces, including biotic surfaces such as tissues and dental plaques as well as abiotic surfaces such as culinary facilities, air water interface, medical devices and inner side of water pipes. (15) Biofilms are commonly constituted of various bacterial species, but some are formed by single species. Most of bacteria have been found living in biofilm communities.

Many infectious diseases are caused by biofilms. For example, *Pseudomonas aeruginosa* biofilms can cause Cystic Fibrosis lung infections; *Vibrio parahaemolyticus* can cause gastroenteritis. Some diseases are caused by multi-species biofilms, such as chronic ear, nose and throat infections. Biofilms also form on medical devices which may cause surgery problems. (16) Biofilms on metal surface can either facilitate or inhibit the metal corrosion under different circumstances. (17) Biofilms in daily water-related facilities will play the roles of bacteria reservoirs, which may cause safety issues of drinking water. (18) Although planktonic bacterial colonies can be killed by water chlorination,

these biofilms cannot be easily eradicated by antimicrobial agents because of the antibiotic resistance of biofilms.

The common processes for biofilm development can be divided into five main steps: initial attachment, irreversible attachment, vertical development, maturation and dispersal (Figure 2.1). The initial attachment of the cell and the surface is transient or reversible. Cells can easily break the binding to the surface and move away. After initial attachment, bacterial cells gradually transfer to permanent or irreversible attachment by forming conditionally synthesized or specific adhesins. (15) The mature biofilms are often adsorptive and contain more than 95% of water. Most biofilms show porous structures, which make it easy for the transportation of nutrients and signaling molecules. It is common for biofilms to secrete extracellular polymeric substances (EPS), which contain polysaccharides, proteins and environmental DNA and protect the embedded cells. (19) The final stage of biofilm is dispersal, during which some motile cells leave the biofilm to form new colonies. Dispersal is regulated by many factors for different biofilms, such as the depletion of nutrients, temperature, inorganic ions, quorum sensing signals and other intercellular signals. (20) Biofilm dispersal enhances the survival of bacteria. If not for dispersal, all the bacteria within the biofilm will finally die from unfavorable environmental conditions, such as depletion of nutrients, accumulation of poisonous substances, temperature change and etc.

Within the biofilm, local environment, cell behaviors and EPS compositions are often heterogeneous. The heterogeneities of biofilms are due to chemical gradients, adaption to local environment, stochastic gene expression and genetic as signals and bacteria waste, may have high concentration at the bottom. (21) The chemical concentrations are determined by diffusion and reaction.

Nutrients and oxygen are depleted within a certain depth inside the biofilm because they are consumed by the bacteria cells near the surface. Chemicals that are secreted from bacteria cells, such as metabolic wastes and signals have higher concentrations inside the biofilm but lower near the surface. The

concentration of the “medium products” produced and used by bacteria is higher in the middle and lower at top and bottom. Bacteria cells interact with different local extracellular stimuli and result in different behaviors. There are cases when two cells of the same species present very different phenotypes. During the biofilm development, a considerable portion (sometimes as large as 10%) of cells differs from its parents. In summary, environmental and genetic variations contribute to the heterogeneity of biofilms.

Figure 2.1 Biofilm development process 1: Initial attachment. Bacteria attach to surface and this initial attachment is reversible. 2: Irreversible attachment. In this state, new cells are produced and bacteria form monolayer biofilm. 3: Vertical development. Cell aggregation evolves into multilayer biofilm 4: Maturation. Cells in mature biofilms present very different properties comparing with planktonic cells. Most biofilms are heterogeneous and resistant to antibiotics, antimicrobial agents and reduction of nutrients. 5: Dispersal, dispersal is an initiative process by the bacteria. (Image Credit: D.Davis).

Biofilms usually provide antibiotic-resisting environments for the bacterial colonies. It has been reported that biofilms can be much less susceptible to antibiotics compared with free-floating cells. (22) Cells, especially those deep inside the biofilm, grow very slow. This is one of the reasons for the increased resistance in biofilms. For some antibiotics that react with cell DNA, they primarily function as bacteria killers when bacteria are replicating their DNAs. At the bottom of biofilm, the ratio of DNA replication period is less than normal; so many cells can still survive. The EPS may also absorb or bind with antimicrobial agents and protect the cells. Additionally, some cells in biofilm may show increasing resistance due to individual phenotype transformation. It has been reported by the National Institute of Health that about 80 percent of all infections in humans are related to biofilms. (23) Within the human body, biofilms can grow in various positions from mouth, ears, and eyes to vessel walls, causing diverse infections on organs including ocular and acoustical. They can also aggregate on the surface of artificial implants leading to surgical site infections. By studying the biofilm formation mechanism, especially with quorum sensing modeling, we can find possible methods for effective cures of such infections.

2.3 Quorum Sensing

Quorum sensing is the term for mechanism dominating the cell response to population density. It is mostly observed in studies of bacteria: bacteria show certain behaviors such as biofilm formation, virulence, and antibiotic resistance, based on the changes of local population density. (3) This sensing happens within a single bacteria species and multispecies interconnections as a strategy-making process to determine the cooperation or defection behavior choice in evolution of competitive interactions.

2.3.1 Basic Mechanism of Quorum Sensing

Certain types of signaling molecules, called autoinducers, play key role in regulating quorum sensing network. (1) Quorum sensing bacteria usually intracellularly produce and extracellularly secrete autoinducers (Figure 2.2); these molecules could be detected by binding to specific receptors after entering the cells by free diffusion or membrane-protein-involved transportation. Autoinducer concentration increases with the population density of cell colony; the binding receptors will form complexes functioning as gene expression regulators after the concentration of autoinducers reaches a threshold value. (24, 25) Some of the receptor-autoinducer complexes are transcription factors themselves, while others relay signals to downstream transcription factors. (26) These transcription factors activate expression of some genes usually involving the autoinducer-producing proteins, so this mechanism constructs a positive feedback loop with the concentration of autoinducers as a switch of a bistable system. This network mechanism describes how one cell reacts and changes its behavior according to different surrounding biochemical environment when the cell population varies. (27)

Figure 2.2 Structures of homoserine lactone autoinducers produced by different Gram-negative bacteria. The top figure shows the typical structure of an AHL molecule. The bottom one lists the different structures of R-group in different bacteria.

2.3.2. Specific Quorum Sensing Systems

V. fischeri

Vibrio fischeri is a Gram-negative, motile, rod-shaped bacterium mainly found in seawater by symbiosis with marine animals such as squid. This bacterium is famous for its bioluminescence, and it is the first reported microbe species applying quorum sensing mechanism. When its population density grows over a critical value, the quorum sensing network turns itself out and the bacteria cells give off the light; if the density is not high enough, the quorum sensing network will remain in the off state and little light will be observed. Later molecular level researches of this quorum sensing behavior discovered that the lux operon regulated by *LuxR* and *LuxI* proteins control the bioluminescent function. (26) The *LuxI* produces a small signaling molecule named autoinducer which can freely diffuse in and out of the cell membrane, thus the density of the autoinducer represents local population density of *Vibrio fischeri*. *LuxR* is the regulator protein controlling the expression level of *LuxI*. If *LuxR* binds with the autoinducer, it is able to turn on the expression of *LuxI*; if not, the expression of *LuxI* stays inactive. This positive feedback mechanism establishes the on-off switching of bioluminescent quorum sensing network. *Vibrio fischeri* has been reclassified into the newly created genus *Aliivibrio* by rRNA comparison. Modeling analyses have been widely done for *V. fischeri*'s quorum sensing system. Goryachev (24) argued that the *LuxR* dimer is the key regulator for the robustness of the quorum sensing networking in *V. fischeri*.

P. aeruginosa

P. aeruginosa, or *Pseudomonas aeruginosa*, is a Gram-negative, aerobic (although classified as aerobic, it is often regarded as a facultative anaerobe, because it can adapt the conditions of partial or total oxygen depletion.), coccobacillus bacterium with unipolar motility. It uses quorum sensing to control biofilm formation, swarming motility, and virulence. (28) These

bacteria can grow within a host without harming it, until they reach a threshold concentration. Then they become aggressive, developing to the point at which their numbers are sufficient to overcome the host's immune system, and form a biofilm, leading to disease within the host as the biofilm is a protective layer encasing the bacteria population. Another form of gene regulation that allows the bacteria to rapidly adapt to surrounding changes is through environmental signaling. Recent studies have discovered that anaerobiosis can significantly impact the major regulatory circuit of quorum sensing. (29) This important link between quorum sensing and anaerobiosis has a significant impact on production of virulence factors of this organism. Garlic and ginseng experimentally block quorum sensing in *Pseudomonas aeruginosa*. It is hoped that the therapeutic enzymatic degradation of the signaling molecules will prevent the formation of such biofilms and possibly weaken established biofilms. Disrupting the signaling process in this way is called quorum sensing inhibition.

V. cholerae

V. cholerae is a Gram-negative, facultatively anaerobic, comma-shaped bacterium, etiological cause of the disease cholera. It was first isolated by Italian anatomist Filippo Pacini in 1854, but his discovery was not widely known until Robert Koch, working independently 30 years later, publicized the knowledge and the means of fighting the disease.

In *V. cholerae*, quorum sensing represses biofilm formation and virulence factor production and activates protease production, as well as causes genetic exchange among *V. cholerae* cells for serogroup conversion. (30) There is great deal of similarity of *V. cholerae* quorum sensing network to that of *V. harveyi* with only small difference: *V. cholerae* does not produce or detect HAI-1. Besides, one regulator in *V. cholerae* plays the roles of both activator and repressor of correlated gene expression.

X. campestris

Xanthomonas campestris pv. *campestris*, Xcc for short, is a rod-shaped, Gram-negative bacterium found in plant infection. It is characterized by double cell walls and yellow pigment. Xcc is the pathogen for black rot, the most destructive disease of crucifers, causing worldwide commercial loss every year. Xcc is now reported as a bacterial species using quorum sensing to regulate its biofilm formation. The signaling molecule of Xcc quorum sensing network is diffusible signal factor (DSF), a new signaling molecule type different from the early discovered quorum sensing signaling molecules. It shares similar quorum sensing mechanism with *Vibrio fischeri* in the central topology of signaling network. When the Xcc density is high enough, the biofilm formation process initiates and when its population drops, the biofilm disappears and the bacterium changes into planktonic state.

Agrobacterium

Agrobacterium is a big bacterial genus that cause disease in plants. This bacterial genus is Gram-negative, and it uses horizontal gene transfer for infection and pathogenesis. Among this genus, *Agrobacterium tumefaciens* is the most widely discussed species. This bacterial species also applies quorum sensing mechanism to regulate its colony behaviors. When its local population density passes the critical threshold, *Agrobacterium* turns out its biofilm formation and starts to invade the plant cells. Goryachev (25) established the regulatory network of *Agrobacteria* quorum sensing in 2005 and conducted numerical sensitivity analysis. In this work, we will study the quorum sensing signaling pathways of *Agrobacteria* in the network topology point of view.

2.4 Microbial Stratification in Meromictic Lake

Though recent years have witnessed rapid development in biological experimental techniques in wet lab, the quantitative measurements of microbe behaviour in outdoor environment become the new challenge for experimentalists. It is easy to build specific microbe community and spatial patterns of multispecies system with the help of all kinds of lab controls, but it is complicated to examine and explain the spatial dependence of microbe distribution on large scales. Latest experiments on meromictic lakes provide us with the chance to look into this issue with both qualitative and quantitative methods.

2.4.1 Meromictic Lake

Meromictic lake is a special type of lakes which have been reported and defined since last century (31, 32). Ordinary lakes, in other word holomictic lakes, are observed to have water mixing between surface layer and bottom layer at least once every year. However, in meromictic lakes, surface water and bottom layer do not have physical exchange for decades or even centuries. Such restriction on water dynamics creates separate physical and chemical environment for different microbe species to live in, thus it provides simple insight of spatially dependent microecology on large scale.

Usually there are three distinguishable layers of water in a meromictic lake. The top layer of water next to the lake-air surface is named mixolimnion, where water dynamics resemble that of an ordinary holomictic lake; the bottom layer is known as monimolimnion, where little water flow exists. The middle layer is called chemocline, where concentrations of some chemical species change dramatically.

The mixolimnion, the top layer, has intensive thermal exchange by air-water surface. Thus, its temperature is likely to be affected by air temperature: in summer, it may grow to over 20°C and in winter it may drop to 1°C. In the mixolimnion, there is normally high concentration of dissolved oxygen since it is exposed to the open air. The oxygen level in this layer depends on the temperature and the air condition near the surface. In summer, it may increase to more than 10mg/L as temperature grows.

The monimolimnion, the bottom layer, is prohibited from heat convection with air by the screening effect of mixolimnion, so its temperature is much more stable than the top layer. The local temperature is constantly reported to be around 4°C because water of this temperature has highest density to ensure its stability at the bottom of the lake. There is very little oxygen dissolved in the monimolimnion. In some meromictic lake, the oxygen level may be less than 1mg/L. However, it is highly possible that we observe high concentrations of other chemical species, for example, sulfur compounds, because this bottom layer is usually saltier than the other. Together with the monimolimnion, the sediment layers on the bottom of the meromictic lake can remain undisturbed for decades, which results in the significance of the sediment cores to trace the climate history at the lake area.

The chemocline is indeed a thin layer where large physical or chemical gradient are presented. It is the middle layer between the mixolimnion and the monimolimnion, so it functions as a buffer layer. The oxygen concentration in chemocline may vary across one order of magnitude as well as the light intensity while there is also a temperature gradient.

Since the last century, meromictic lakes have been reported all over the world from tropical areas to polar regions. Despite the locations, they share the similar lake structure described above. This raises the need to conduct general theoretical study of meromictic lakes to explain their features and predict possible dynamics of both hydro conditions and biological events.

Ace lake is a meromictic lake located in Antarctica. It was first visited in 1974, and has been actively studied ever since. The chemocline separates the mixolimnion, the top layer supersaturated with oxygen, and the monimolimnion, the bottom layer of permanently cold, anoxic water. The sediment layer on the bottom of Ace lake supports a methane cycle that has been in existence for the past 3000 years. Recent experiments have isolated various bacteria and virus species, and their contributions to biogeochemical cycling has inspired active researches (26, 27).

2.4.2 Microecology in Meromictic Lake

The three-layer structure of meromictic lake provides distinguishable and stable environments for different biological species to live in (33-36). The top layer, mixolimnion, are filled up with oxygen supply, which enables the aerobic metabolism. So fish are seen in the areas near the water-air surface; many aerobic microbe species also exist in the mixolimnion. Besides, there is usually plenty of light intensity to ensure the photosynthesis. Since both respiration and photosynthesis are supported with enough supply, the metabolisms of microbe community are very active. According to recent study, we find that there is complex carbon cycle network built by biochemical reactions with high throughput in this layer.

However, the oxygen level starts to drop from the chemocline to the monimolimnion to the sediment layer, inhibiting the aerobic species. In addition the light intensity decreases dramatically below the chemocline, which makes photosynthesis hard to take place. Generally speaking, there will a drop in total biomass of microbes due to the limited nutrient resource, but people have found high levels of sulfur, nitrogen, and phosphorus in the areas below chemocline, these conditions create an ideal environment for anaerobic bacteria to grow in. Several major anaerobic microbe species are observed to exist in the

monimolimnion, including purple sulfur bacteria or green sulfur bacteria and sulfur-reducing bacteria. Purple sulfur bacteria (PSB) and green sulfur bacteria (GSB) are anaerobic bacteria species both apply photosynthesis to produce elemental sulfur from hydrogen sulfide, whereas sulfur-reducing bacteria (SRB) maintain their energy supply by reducing elemental sulfur to hydrogen sulfide. These species combine to establish the basic sulfur cycle in the monimolimnion. Although this sulfur cycle is general for most meromictic lakes, the species distributions show divergent patterns for different lakes (33, 34). The quantitative measurements introduce new question of understanding this specificity from basic biochemical laws.

2.5 Microtubule and Tubulin

Microtubules (MTs) are dynamic cytoskeletal polymers that play important roles in various cellular processes, such as mitosis and meiosis, motility and intracellular transport (37, 38). They also form as a intracellular structural matrix supporting the cell shape. Their involvement in various cellular processes makes them an attractive therapeutic target for cancer therapy. One microtubule has an outer diameter of about 24 nm and an inner diameter of about 12 nm and its length can vary across dozens of micrometers. MTs are constructed by tubulin heterodimers made of α and β tubulin subunits inside the cell. Two guanine-nucleotide binding sites are found in the tubulin dimer, a non-exchangeable (N) site at α -tubulin always bound with one GTP and an exchangeable (E) site at β -tubulin that can switch between GTP or GDP-bound state. Both α and β tubulins have three important domains, a nucleotide-binding N-terminal domain (N-domain), an intermediate Taxol/Epothilone binding domain (I-domain) and a C-terminal domain (C-domain) known to interact with motor proteins.

MTs inside the cells are in a dynamic process known as microtubule dynamic instability in which they toggle between growing and shrinking phases. The molecular mechanism involving a switch between growing and shrinking phases is due to the GTPase cycle which leads to hydrolysis of GTP in β -tubulin upon addition of tubulin dimers (38). MTs inside the cells are in a dynamic metastable equilibrium state by a layer of GTP-tubulin cap on the growing (+) end of the MTs. Due to the addition of tubulin dimers at the growing end during polymerization, GTP present in the E-site of β -tubulin undergoes hydrolysis. Loss of GTP-cap results in shrinking of MTs by shredding away the GDP-tubulins. This dynamic behavior of MTs is robustly controlled due to the structural scaffold of tubulin dimer and its GTPase cycle (39, 40). Although several works have provided many interesting insights into the tubulin dynamics and microtubule assembly, how the GTP-bound state of tubulin dimer can help to stabilize the whole microtubule structure has not been very well understood.

The role of nucleotide states affecting tubulin conformation and microtubule dynamics is still a largely debated issue. The molecular forces through which the different nucleotide states regulating tubulin dynamics remains unclear.

2.6 Questions to Answer

One major challenge to study biological systems is the great complexity and diversity. Not only there may exist huge number of elements involved in one biological network, but also the biological network may have a complicated topological structure. It is often the fact that to ensure the reliability and accuracy biological experiments have to focus on individual elements or events despite the complexity involved in nature environment. Collective behaviours on different scales from molecular level to ecology level may show divergent

patterns from intuition. Theoretical methods are thus necessary to help the analysis on diverse scales for a general and complete vision.

Based on the background of this work, we aim to improve the theoretical understanding towards two aspects of complex biological systems: identify the critical elements of signalling pathways on network level and understand how the interactions affect overall phenomenon of cellular behaviour. To be specific, we intend to conduct network analysis for quorum sensing system to see the critical part for the network on-off switching function, apply physics methods to model the large scale spatial patterns of microbe ecology in meromictic lakes and build theoretical model for the analysis of tubulin dimer dynamics. This is indeed a bottom-up technique in theoretical point of view to solve the puzzles in complex biological problem covering diverse space and time scales.

Chapter 3

Modelling Methodology

Cell, the fundamental functional biological unit of all living organisms, remains a mystery for its complexity and diversity. Over the past decades, researchers have been building new and inspiring techniques to probe and analyze cellular structures and their functions. Thanks to the development of biofilm research, we can go one step further in observing and testing novel bacterial cell behaviors; and with the help of molecular experimental techniques, we are now able to see a clear view of cellular signaling pathways. However, it is still challenging to understand the dynamical behavior of cell over different space and time scales because of the limitation of experimental techniques.

3.1 A Modelling View

It is obviously beneficial to build mathematical model for cell regulation system study. Not only can mathematical model further our understanding of cell regulation behavior, but also provide generate quantitative results with the help of its predicting power. As the computing devices and numerical algorithms have rapidly developed in recent years, we are able to conduct large-scale mathematical study of complex cell pathway by analysis and simulation (6, 27, 41). These studies of network dynamics can test and generate hypotheses about network architecture, the fundamental governing mechanisms, and the effect of modification and mutation. System-level mathematical network models can extend our research level of single genes and proteins to thinking of how perturbations translate into dynamic cell behavior and why different behavior occurs in diverse species despite their network homology.

To understand the cell regulation system, we must include over thousands of interactions among biophysical and cellular structures at separate time scales. A dynamic theory of such complexity must be simplified to build a solvable model. One attempt is motif identification (7). When referring to higher level of systematic behavior, we build network modules from interconnected motifs, representing the biological pattern with mathematical theory. It is very important for us to study the cell regulatory modularity, because it maintains robustness and evolvability, as well as reduce the potential lethality of mutations. It was suggested that modularity arises spontaneously with evolution in response to environmental variation.

Despite the advantages and necessity of mathematical modeling for cell research, there are still many challenges in forming and analyzing models. To build a well-defined model, we always need an appropriate assumption based on the goal of the research. On top of that, it is often the case that modelers are faced with lack of ‘detailed, quantitative biological information’ (42). Although biologists have been advancing our knowledge of diverse systems and behavior,

the parameters of related biochemical kinetics and spatial distribution of specific reactions are mostly unknown. Even high-throughput biological data can hardly maintain sufficiently high accuracy for theoretical analysis. As a result, the combination of mathematical modeling and experimental validation is required to address these issues. With the iteration of two, we can enhance the accuracy and predictive power of the theoretical model.

There have been distinguishing network models built for diverse purposes, and many of them can help develop similar system for cell regulation. Widely used in complex biochemical systems, kinetic models represent the networking structures as series of coupled chemical reactions (42, 43). These models make use of the chemical reaction equations involved in the cell regulation system to establish the relationships between components in the networks. Another prevailing type of mathematical modeling is Boolean network modeling (44). This method focuses on the discrete states of every component (point) and uses the regulatory connections to form logical network of the biological system. Since there are limited states of the whole system, the network will converge to certain cycles named attractors after reaching equilibrium, which reveal the possible phenotypes of cells. Besides, spatial dynamics is also important for detailed study of cell behavior. Models of this type often study spatial gradient of signaling activities in and out of cells (45).

In the following sessions, we will give a brief introduction of the basic mechanism of construction and analysis of dynamic biological models. Significant theoretical methodology is shown in Figure 3.1.

Figure 3.1 The general modeling point of view of biological study. The experimental observations provide direct support for parameter evaluation and estimation, whereas numerical study offer explanation and prediction for current and future experiments. Model architecture is the presentation of the key elements and interactions of focused subject and analytic methods help to identify the robust motif of the model and solve the causality of numerical or experimental events.

3.1.1 Model Architecture

For a complete mathematical representation of cell regulation system, we need build the theoretical network model based on reaction mechanisms. Both system architecture and parameter settings are essential to determine the model. However, since many of empirical discoveries have shown cells' robustness to varying conditions, the models should also stay robust to a varying range of parameters. When we are about to set up the network, model architecture always comes first.

There are two main issues involved in the architecture building: interaction between different molecular species and the kinetics of these interactions.

Pathway databases (EcoCyc, Kegg, ERGO, aMAZE, ExPASy, STKE, and Nature Signaling update) have been established for research based on specific data from GenBank, PDB, and EMBL (46-48). They store detailed experimental information of the molecular types and the related biochemical reactions, so we can work out the kinetic mechanisms coupled with the reaction networks.

3.1.2 Deterministic Model

According to mathematics and physics definition, deterministic systems are systems that there is no randomness shown in the development of the state. In other words, if the evolving mechanism and the initial state are given, a deterministic system will always have the same output. Physics models with evolutions built on differential equations are deterministic models; even for quantum mechanics, the system state described by Schrodinger equation or other representation, is deterministic, although the observables of the state follow certain distribution and are thus non-deterministic. Besides, a chaotic dynamic system is also deterministic and it will give the same result if the initial state remains the same. The uncertain property of chaos rises from the highly sensitivity of the initial state and the lack of precision in measuring the initial state.

In cell regulation system, the biochemical reactions can be described with a set of ordinary differential equations (ODE) representing the mass reaction rates. So we can establish a deterministic model for the biochemical reaction network. The general form of first-order ODE systems can be written as:

$$\frac{d\vec{x}}{dt} = \vec{f}(\vec{x}, t), \quad (3.1.1)$$

where $\vec{x}$ is a vector every component of which stand for the measurement of one molecule species in the system. Since time is not explicitly involved in the reaction rate functions, so this can be simplified as

$$\frac{d\vec{x}}{dt} = \vec{f}(\vec{x}) . \quad (3.1.2)$$

This representation implies that we assume the reactions happen in a well-stirred reactor with homogenous condition and that we qualify each component with its bulk concentration. After specifying the explicit form of $\vec{f}$ and key parameters as well as initial states, we can solve the ODE system analytically (for small system and simple system) or numerically (for large and complex system). Several efficient numerical algorithms have been developed on diverse widely accepted computational platforms, such as Matlab and Python platform and people have even modified the methods for biology research purpose in some numerical tools (43).

3.1.3 Stochastic Model

Different from the property of deterministic systems, stochastic systems refer to the ones with randomness: the system develops following specific probability distribution related to its structure, so the behavior cannot be determined without large scale numerical simulations.

ODE-based deterministic models are based on ideal systems where time and concentration variables are continuous and all chemical reactions are simultaneously measured. They prove great power in analytical study of network behavior in both physics and chemistry, but in some real biological systems, the dynamic evolutions are quite stochastic. There are two origins for the stochastic behavior in cell regulation system. First, small numbers of certain types of molecules lead to the intrinsic stochastic biochemical reactions.

Because it is often the case that the number of one specific reaction molecule species in a cell is less than 10, the value of the related concentration should be considered as discrete numbers and can no longer be treated as continuous variable in deterministic systems. The other factor contributing to randomness is the biological noise during the gene expression process. There is uncertainty

of the binding rate for enzymes and DNAs or mRNAs, and this will trigger fluctuations from upstream genes to other biochemical components in the network. Both of a challenge and a chance for modeling are presented by the presence of noise. To perform stable biological state, we should change the network architecture to minimize the impact of noise. In addition, noise can also explain the diversity of phenotype, evolutionary adjustment to environments and cell differentiation.

By now, we have enough motivation to build stochastic models for study concerning above issues. The dynamics of such a system with well-stirred homogenous condition can be captured by a generic chemical master equation (CME) (49):

$$\frac{\partial P(\vec{x}, t | \vec{x}_0, t_0)}{\partial t} = \sum_{j=1}^M a_j(\vec{x} - \vec{v}_j) \times P(\vec{x} - \vec{v}_j, t | \vec{x}_0, t_0) - a_j(\vec{x}) \times P(\vec{x}, t | \vec{x}_0, t_0), \quad (3.1.3)$$

where $\vec{x}$ is the numbers of each molecule species, t is the time when the system is measured, a_j is the propensity value for the reaction j , and $\vec{v}_j$ is a vector containing the changes in the system state caused by reaction j . The first term on the right side of the equation stands for the probability of a species reacting at time t to become a specific state $\vec{x}$, while the second describes the probability of a species of $\vec{x}$ at time t changing its state by reactions.

We can analytically solve the CME for very simple system with ideal approximations. The number of possible paths of reaction network increases exponentially with more molecule species and biochemical reactions.

Numerical tools using Monte-Carlo method have been developed to help with stochastic simulations (41), where occurrence of every elementary reaction leads to one direct formation of product and the probability of such reaction is simulated with the expression similar in ODE systems.

3.2 Mass Action Rate

In this work, the biochemical reactions we mainly deal with are classic chemical reactions involving massive particle processes in bulk solution environment. To describe the kinetics of these biochemical processes, we apply the mass action law of chemistry which proposes that the chemical reaction rate is proportional to the product of concentrations of all reactants. For the simple case, the reaction, $A + B \rightarrow C$, has a production rate of C of $k*[A]*[B]$, where the brackets denote the concentrations of individual chemical species, and k is the rate constant depending on the activity of this reaction.

3.3 Michaelis-Menten Mechanism

Complexity is one of the major problems of biological system, especially when we refer to biological reaction networks on molecular level. There are huge numbers of biological processes more complicated than basic chemical reactions, for example, enzyme-catalyzed reactions. Enzymes function as cell behavior regulators and thus follow mechanism different from simple chemical reaction. A common model to explain the effect of enzyme on reaction rate is the Michaelis-Menten mechanism (50, 51). This theory simply assumes that the enzyme is not consumed and thus total concentration of enzyme always stays constant. It is only the substrate that is consumed to produce the final product in the whole reaction chain:

where E , S , and P stand for enzyme, substrate and product. Here the forward and backward reaction rate constants of the left reaction are k_f and k_r respectively, and the reaction rate constant of the right one is k_{cat} .

With the quasi-steady-state assumption that the concentration of enzyme-substrate complex is steady, we can get the producing rate of P as:

$$\frac{dP}{dt} = \frac{V_{max}[S]}{K_M + [S]}, \quad (3.3.2)$$

where V_{max} is the maximal reaction rate which is proportional to the total concentration of enzyme and K_M is the Michaelis-Menten constant $\frac{(k_r + k_{cat})}{k_f}$.

This expression describes a saturation curve of rate-concentration relation (Figure 3.2): when we had infinite supply of substrate, the final production rate would reach the maximal reaction rate V_{max} . And the parameter K_M could be easily measured with experiments because when the production rate equals half of the maximum, the substrate concentration is the same as K_M . So the Michaelis-Menten mechanism not only helps to build a simple mathematical expression for enzymatic reaction rate, but also can be fitted to experimental observation with ease.

Figure 3.2 Substrate dependence of reaction rate from Michaelis-Menten mechanism. The M-M kinetics describes the enzymatic reaction process generating a non-linear rate-concentration relation. The curve shows a saturation pattern of concentration dependence that when substrate concentration increases to infinite there is a saturation limit value of production rate. The value of M-M constant K_m can be examined directly from experimental measurement of this curve, because when the reaction rate reaches half of the saturation maximum, the substrate concentration equals K_m .

3.4 Reaction-Diffusion Theory

Both theoretical and experimental studies on classical biological systems are based on homogeneous conditions, i.e., there are no concentration gradients or spatial distributions of molecules. However, recent developments in bio-imaging down to single cell level or even single molecule level unveil the detailed structures below 100nm scale. With the enhanced resolution, novel

observation of spatial related behaviors of biological process are constantly being reported. This triggers the need to explain and analyze the heterogeneous dynamics of complex biochemical processes within small systems. To understand the novel observations, we apply the reaction-diffusion theory to study the spatial distribution related biological events.

The basic mechanism of reaction-diffusion is quite simple that only two dynamical processes are considered: self-diffusion of individual molecular species and the chemical reactions between different chemical species. The assumption behind this simple idea is that different molecular species take no interactions in diffusion and the biochemical reactions happen at such a fast rate that will never affect the diffusion process, i.e., two processes are isolated. The mathematical expression of this theory can be written in a partial differential equation (PDE) system as follows:

$$\begin{aligned}\frac{\partial[A]}{\partial t} &= D_A \nabla^2[A] + \left. \frac{\partial[A]}{\partial t} \right|_{reaction} \\ \frac{\partial[B]}{\partial t} &= D_B \nabla^2[B] + \left. \frac{\partial[B]}{\partial t} \right|_{reaction} \\ \frac{\partial[C]}{\partial t} &= D_C \nabla^2[C] + \left. \frac{\partial[C]}{\partial t} \right|_{reaction}\end{aligned}\tag{3.4.1}$$

where A, B, and C denote three different molecular species involved in one simple reaction $A + B \rightarrow C$. The parameter Ds are the diffusion constant based on Fick's law, which assumes that the flux of particles is proportional to the gradient of concentration. The second terms on the right hand side of the equations are the rate contribution of chemical reactions between these species, based on the mass action law. This PDE system enables us to solve the dynamics of a real biological system and explain the spatial pattern of multi-molecular species.

3.5 Stability Theory and Sensitivity Analysis

Because of the great complexity of biological networks, it is always a challenge to quantify dynamic properties of one complete living system by simple and clear methods. For example, in one typical quorum sensing regulatory network on molecular interaction level, over 30 constant parameters and over 20 initial

conditions are essential inputs to determine the time evolution of the whole system. How the values of these parameters affect the overall dynamics and final state is one of the major questions of numerical analysis of complex networks. We mainly focus on the dynamic behaviors of several biological systems in this work, so we apply two well-defined mathematical theories to understand the dynamics of time-dependent properties and interactions of bio-systems: stability theory and robustness analysis method.

3.5.1 Functional Stability Theory

The interaction network on molecular level dominates cell behavior. Although the biological systems are expected to be highly dynamic even on very short time scale, the experimental observations focus on long-time properties rather than independent instant snapshots. Since cells may change their behaviors according to intracellular and extracellular conditions, we often require certain level of stability and steadiness to discuss the cell behaviors. So both the dynamics of cellular evolution and the stable steady state of cellular system are among the hottest topics in cell biology. After the mathematical expression is built upon the interaction network, it is easy to get the time-dependent evolution and solve the possible steady state. One simple theory to check whether the steady state can be truly observed is the stability theory of functional ODE system. If one steady solution of the ODE system is stable, then this steady state might be achieved; if not, this steady state might decay into oscillating motions or chaos depending on the attractors of phase diagram, since there always exist uncontrollable noise and fluctuation in living biological system to prevent pure thermal equilibrium to be reached.

In a continuous autonomous biological system, the state evolution is governed by the following equation:

$$\frac{d\vec{x}}{dt} = \vec{f}(\vec{x}, t), \quad (3.5.1)$$

where each component of vector $\mathbf{x}$ stands for the concentration or population of each biochemical species, and the functional $\mathbf{f}$ denotes the time derivative of the whole system. The steady state condition of the above ODE system is

$$\frac{d\vec{x}_s}{dt} = \vec{f}(\vec{x}_s, t) = 0, \quad (3.5.2)$$

where the subscript s indicates the steady state solution.

Now we assume that there is a small variation in the steady state, $\vec{\delta}(t)$, thus the dynamics of the disturbed state is

$$\frac{d[\vec{x}_s + \vec{\delta}(t)]}{dt} = \frac{d\vec{\delta}(t)}{dt} = \vec{f}[\vec{x}_s + \vec{\delta}(t), t] = \vec{f}(\vec{x}_s, t) + F\vec{\delta}(t) + O[\vec{\delta}(t)^2] \approx F\vec{\delta}(t), \quad (3.5.3)$$

in other words,

$$\frac{d\vec{\delta}(t)}{dt} = F\vec{\delta}(t), \quad (3.5.4)$$

where F is the Jacobian matrix,

$$F_{ij} = \frac{\partial f_i}{\partial x_j}. \quad (3.5.5)$$

The magnitude of variation follows an exponential dependence of time as shown in the above equation. It will decay exponentially if all eigenvalues of matrix F are negative, indicating a stable steady solution is achieved; otherwise it will take an exponentially increase to deviate from the steady state.

On the other hand, if we are dealing with a non-continuous autonomous system, the dynamics based on discrete time steps can be expressed as follows:

$$\vec{x}' = \vec{g}(\vec{x}), \quad (3.5.6)$$

where $\vec{x}$ and $\vec{x}'$ stand for the state vectors of the time steps before and after the evolution process g . Corresponding steady condition of this type is

$$\vec{x}_s = \vec{g}(\vec{x}_s), \quad (3.5.7)$$

And a disturbing in this state will generate variation in the subsequent time step:

$$\vec{g}[\vec{x}_s + \vec{\delta}(t)] = \vec{g}(\vec{x}_s) + G\vec{\delta}(t) + O[\vec{\delta}(t)^2] \approx \vec{x}_s + G\vec{\delta}(t), \quad (3.5.8)$$

where similarly,

$$G_{ij} = \frac{\partial g_i}{\partial x_j}. \quad (3.5.9)$$

It is obvious that if all eigenvalues of G have absolute value of less than 1, the variation will decrease with each time step and the state will exponentially converge to the steady solution.

The above algorithms have been developed to study the stability of autonomous dynamic system. Since most biological systems fall in this category, we could use these techniques to solve the puzzle of stability of complex biological networks.

3.5.2 Sensitivity and Robustness of Biological Networks

The system-level behavior of a mathematical model is governed by both the model architecture and the associated inputs. In biology, there are many network systems with high robustness to intracellular and extracellular environment changes, while others with ultra-sensitivity to certain parametric variations. However, most parameters involved in the dynamics are not precisely known, so it is often a key problem to understand the parameter dependence of a complex system.

Sensitivity analysis is the study of how a system's outputs change with different inputs. This system analysis technique has a number of benefits, including model modification and testing the functioning region of the model (52). With sensitivity analysis, we can test model response in the presence of input changes to modify the network architecture for corresponding biological system. Besides, to clarify the circumstance where the network model functions properly is also important to understand the mechanism of the system dynamics. When we refer to the inputs, we usually consider both the governing parameters of dynamics and the initial states. However, except for the chaotic cases, biological networks often require converged stable solutions for the cell homeostasis, so the parametric sensitivity analysis has interested more theoretical studies for a range of disciplines.

Consider a biochemical system described by $\vec{x}$ representing the concentrations of every molecule species for a deterministic model, thus we can ensure the dynamic equations have a unique solution, $\vec{x} = \vec{x}(\vec{\phi}, t)$, where $\vec{\phi}$ stands for the set of parameters for all the reactions. The general expression of the local sensitivity called *absolute sensitivity* $\vec{s}$ for a specific parameter ϕ_i can be written as

$$\vec{s}(\vec{x}; \phi_i) = \frac{\partial \vec{x}(\vec{\phi}, t)}{\partial \phi_i} = \lim_{\Delta \phi_i \rightarrow 0} \frac{\vec{x}(\vec{\phi} + \Delta \phi_i, t) - \vec{x}(\vec{\phi}, t)}{\Delta \phi_i}, \quad (3.5.10)$$

where the vector $\vec{s}$ consists of the parametric sensitivities of each molecule species. There is a more prevailing form named *relative sensitivity* applying normalization to the above form:

$$\vec{S}(\vec{x}; \phi_i) = \sum_j \frac{\phi_i}{x_j} \cdot \frac{\partial x_j(\vec{\phi}, t)}{\partial \phi_i} \cdot \hat{x}_j = \frac{\partial \ln \vec{x}(\vec{\phi}, t)}{\partial \ln \phi_i}. \quad (3.5.11)$$

In practical applications, we may focus on certain objective function determined by the goals of study rather than the concentrations, so with little algebra transformation, we can induce the objective sensitivity.

Although the above definition of sensitivity function helps us to reveal the dependence of complex system on input parameters, nature may even generate more complicated problems: sometimes the inputs include not only static parameters, but also time-dependent signals. For instance, if bacteria lived in a semi-closed environment with varying nutrient supply, the dynamic behavior of these microbe would show time-dependence due to the fluctuation in the nutrient. To be general, we should apply mathematical tools to quantitatively study this effect induced by the uncertainty in the input function. As a result, we apply the robustness analysis of functional dynamical system which uses time-dependent function as input.

When external function is needed as input, the ODE system is written as follows:

$$\frac{d\vec{x}}{dt} = \vec{f}(\vec{x}, \vec{y}, t) \quad (3.5.12)$$

where y denotes the groups of external input functions. To analyze the dynamic robustness of the system state, we need to apply linearization around steady solution by Fourier transformation.

$$-i\omega\widetilde{\Delta x_s} = \frac{\partial \tilde{f}(\widetilde{x_s}, \widetilde{y}, \omega)}{\partial \widetilde{x_s}} \cdot \widetilde{\Delta x_s} + \frac{\partial \tilde{f}(\widetilde{x_s}, \widetilde{y}, \omega)}{\partial \widetilde{y}} \cdot \widetilde{\Delta y} \quad (3.5.13)$$

where $\widetilde{\Delta x_s}$ denotes the system fluctuation around steady state, and $\widetilde{\Delta y}$ is the fluctuation of input signals. Following the general concept of input dependence, we derive the dynamic robustness against input fluctuation as:

$$\tilde{S} = \left| \frac{\partial \widetilde{\Delta x_s}}{\partial \widetilde{\Delta y}} \right| = \left| \frac{\frac{\partial \tilde{f}(\widetilde{x_s}, \widetilde{y}, \omega)}{\partial \widetilde{y}}}{i\omega + \frac{\partial \tilde{f}(\widetilde{x_s}, \widetilde{y}, \omega)}{\partial \widetilde{x_s}}} \right| \quad (3.5.14)$$

This ω dependent function describes the system response on different time scale when the fluctuation takes place. The smaller the norm of S function, the less the system is affected by the fluctuation. If S increases in around critical ω range, the system becomes vulnerable to the fluctuations with corresponding frequency; if S decreases, it shows high resistance to fluctuation around that time scale.

3.6 Parameter Estimation

Parameters often play the key role in complex mathematical systems, especially for the models for biological networks. Following the modeling architecture methods and dynamic analysis, we can build the equations to simulate the biophysical or biochemical events, which may involve large number of parameters. For a minimal gene regulation network, there are usually over 10 parameters, and to write a real complex network model with cross-talking between different signaling pathways, we may need more than 40 parameters. There is always the problem that even the most up-to-date experiment techniques are not able to provide explicit and accurate values of all parameters we need. The detailed biochemical reactions are difficult to measure at extreme short time scale and most biophysical processes take great complexity making it

hard to isolate different steps experimentally. As a result, it is highly likely that sometimes we are not able to get solid support of parameter values from experiments, so estimating parameter values without direct experiment measurement are inevitable in modeling biological system.

There are two major types of parameters involved in modeling: absolute parameters (bare parameters) and relative parameters (normalized parameters). For absolute parameter estimation, we usually set their bare value based on comparable quantities of similar biophysical or biochemical interactions. To be specific, when we are about to modeling the dynamics of Cyclic di-GMP regulation network, we find that direct measures missing; then we estimate the kinetic parameters from those of Cyclic AMP, a chemical species has similar molecular structure and alike function in signaling pathway. If there is no available experimental data of high similarity, we need to estimate the parameters from fundamental physics quantities and processes. In our work on tubulin dimer modeling, some elastic constants are not available from experimental resources. We calculate their values from the protein structure and raw data of biochemical bonding strength. In study based on normalized parameters, we mainly conduct qualitative analysis and comparison, so as long as the model is robust under moderate uncertainty, we are allowed to vary the parameter values within reasonable ranges. The specific restriction on plausible estimation of normalized or relative parameters depends on both physics background of the kinetics and the key quantity we concentrate upon. After the primary estimation and model simulation, it is necessary to check the result by comparison with related experiments to ensure the reliability and provide reference support.

3.7 Numerical Simulation

A complete modeling study requires both analytic method and computational simulation. Numerical method plays key role in solve complex or difficult

mathematical models, which is often the case in biological related researches. On top of the fact that previous mathematics and physics development provide us enormous analytic techniques for modeling, current growth in computational technology have inspired objective interpretation and visualization. Not only mature algorithms have been built but also convenient platforms and packages have been made available for public scientific access. In this work, all computational scripts are written for Python with famous scientific calculation packages including SciPy, NumPy, and etc..

Classic models for biochemical network analysis are mostly based on deterministic mathematical system. Finite difference methods are the major tool to solve these models. The specific methods involved in this study are described in the references about the above numerical packages we mentioned.

To improve the mathematical model based on ideal statistical assumptions for comparison with real situations, we must introduce uncertainty to build stochastic simulations. Different algorithms and software have been developed for such research purpose.

Gillespie (53) established a Monte-Carlo solution to the CME (equation 3.1.3) with a simple algorithm. The probability of every reaction depends on its reaction propensity, which is defined similarly as the rate expression of ODE-system. The reaction propensity describes the probability of one reactant molecule colliding with another reactant molecule (Table 3.1). Although it is a simple method to simulate stochastic kinetics, Gillespie algorithm requires drastically increasing cost when the number of reactions and molecules reaches the level around a real system. Gibson (54) developed the ‘next reaction model’ to improve computational efficiency of the Gillespie method. He used a tree data structure to store each reaction time, and a map data structure to minimize recalculation of the reaction propensity at every time step.

Several other algorithms were also announced to provide efficient calculation methods for stochastic simulation. Most of these are not exact and require users to set a parameter that defines the accuracy of numerical solution. Tauleap

algorithms (55) assume multiple firing of fast reactions and reduce the total number of time steps. A hybrid algorithm (56) calculates fast reactions using either ODEs or Langevin equations, while treating slow reactions with the stochastic algorithms.

Reaction	Propensity
$A \rightarrow B$	cx_A
$A+B \rightarrow C$	cx_Ax_B
$2A \rightarrow B$	$cx_A(x_A - 1)/2$

Table 3.1 Reaction propensity for Gillespie stochastic method. The concentrations come in the unit of molecules per cell.

Chapter 4

Gene Regulation Network Analysis of Quorum Sensing

Quorum Sensing is a cell-cell communication mechanism for population density response in bacteria. This mechanism has been reported to be used by diverse bacterial species for different biological functions. In this chapter, we model the quorum sensing network on molecular interaction level for Agrobacteria and develop theoretical method to identify the key network topology responsible for the switching of quorum sensing, which can be applied for future studies in other species.

4.1 Minimal Gene Regulation Network Dynamics

Gene expression is a rather complicated biological process, but it is always helpful to start from the simplest case of a single gene expression without any regulator: the minimal gene expression model. According to the central dogma, biological sequential information transfers from DNA through mRNA, and finally to protein (Figure 4.1). Although there are special transfers and potentially unknown transfers, here we just consider three general transfers in this model: DNA replication (from DNA to DNA), transcription (from DNA to mRNA), and translation (from mRNA to protein).

Figure 4.1 General transfers of genetic information suggested by the central dogma. Assuming the DNA level stays constant, we just take these two main pathways into consideration for the models in this report.

We assume that the DNA level remains relatively stable in the cell phase of interest and thus ignore the replication of DNA molecules. So the involved reactions can be written as:

The first and second processes are based on the transcription and translation for mRNA and protein; and the third and fourth ones show their self-degradations. According to the reactions, mRNA is produced by transcription of DNA and the

translation rate of protein depends on the concentration of mRNAs (Figure 4.2a). The stable DNA concentration in the cell results in a constant transcription rate of mRNA molecules. The ODE representation of this dynamic system based on these reaction equations is

$$\frac{d[mRNA]}{dt} = k_m - d_m[mRNA] , \quad (4.1.1)$$

$$\frac{d[Protein]}{dt} = k_p[mRNA] - d_p[Protein] , \quad (4.1.2)$$

where $[x]$ stands for the concentration of molecular species x , and the other parameters follow the convention mentioned at the beginning of this part. In real experiments, the reaction constants can vary heavily depending on the specific gene and internal or external cell environment, but for the minimal gene expression model we can just adapt the widely accepted estimation to show the time-dependent evolution pattern (57).

Figure 4.2 Reaction networks of the three gene expression models. (a) shows the reactions involved in the minimal gene expression model, and (b) and (c) shows those from the protein monomer-TF model and the protein dimer-TF model.

Next, let us study the basic feedback behavior of the protein functioning as the transcription factor (TF) of its own gene. TF works as gene regulator by binding to certain DNA sequences and then altering the binding chance of RNA polymerase to the DNA. There are two types of TFs, those promoting (activators) and the other blocking (repressors) the recruitment of RNA polymerase. Here we mainly discuss the effect of activators and study the consequences of the positive feedback loop in gene expression models. Protein may work as activator alone or by binding with other biological molecules to form a complex; a very common case is that one protein binds with another protein of the same species to form a protein dimer as transcription factor. We expand the minimal gene expression model discussed above with protein monomer as TF and protein dimer as TF separately (Figure 4.2b and 4.2c). Based on the networking topology, the reactions involved in the monomer-TF model are presented below:

And those in the dimer-TF model are summarized as:

From these biochemical reactions, we have the following two sets of ODE systems:

$$\frac{d[mRNA]}{dt} = k_m - d_m[mRNA] + V_m \cdot \frac{[Protein]}{K_m + [Protein]}, \quad (4.1.3)$$

$$\frac{d[Protein]}{dt} = k_p[mRNA] - d_p[Protein], \quad (4.1.4)$$

$$\frac{d[mRNA]}{dt} = k_m - d_m[mRNA] + V_m \cdot \frac{[Dimer]}{K_m + [Dimer]}, \quad (4.1.5)$$

$$\frac{d[Protein]}{dt} = k_p[mRNA] - d_p[Protein] - 2k_D[Protein]^2 + 2d_D[Dimer], \quad (4.1.6)$$

$$\frac{d[Dimer]}{dt} = k_D[Protein]^2 - d_D[Dimer]. \quad (4.1.7)$$

Equations 4.1.3 and 4.1.4 are based on the monomer-TF model, and 4.1.5-4.1.7 based on the dimer-TF model. With widely accepted parameter values, we plot the time-dependent evolution curves of these two models and make a comparison with the minimal gene expression model (Figure 4.3). It is the same in all three models that the concentration of mRNAs increases and reaches steady state earlier than protein. Dimerization delays the equilibrium of the system, and causes two stages in the time development. In the first stage from the beginning to 30 minutes for mRNA, since there is very few protein dimers in the cell (statistically less than 1 molecule per cell), the mRNA and protein increasing rate and is almost the same for minimal gene expression model and the model of dimer-TF. (It seems that the first stage for protein lasts much longer than for mRNA, but it results from the delay of responding time of protein level to mRNA level caused by time of translation.) After this period, the accumulated dimers keep raising the mRNA level in the dimer-TF network, entering the second stage of this system, while the mRNA concentration remains in the model without any TF. During the second stage, the concentration of dimers increases quickly and comes to its steady state with other molecular species. So the dimer regulation mechanism mainly affects the system in a longer time scale than the basic transcription mechanism.

Figure 4.3 The time evolutions of the three gene expression models. The plot is based on the ODE systems of their reaction networks. (a), (b) and (c) present the increase patterns of mRNAs, protein monomers and protein dimers. (a) and (b) show a clear similarity of mRNA and protein production between the minimal (no TF) model and the dimer-TF model at the beginning.

4.2 Stability of Gene Regulation Network

In minimal gene expression network, the steady state can be achieved with intrinsic stability. It is easy to get the analytic solution under steady condition for these two linear functions and there exists only one solution:

$$[mRNA]_s = \frac{k_m}{d_m}, \quad (4.2.1)$$

$$[Protein]_s = \frac{k_m}{d_m} \cdot \frac{k_p}{d_p}, \quad (4.2.2)$$

Following the stability analysis method of ODE system, we can calculate the Jacobean matrix of the dynamic evolution,

$$F = \begin{bmatrix} -d_m & 0 \\ k_p & -d_p \end{bmatrix}, \quad (4.2.3)$$

and it is obvious to find that two eigenvalues of F are $-d_m$ and $-d_p$, both negative regardless of parameter estimation range and the steady state, which indicates the steady state in this case is always robust to biological noise.

By similar analytic derivation, the steady state of the monomer-TF model can be calculated by solving a quadratic characteristic equation,

$$y = -\frac{d_m \cdot d_p}{k_p} [Protein]^2 + \left(k_m + V_m - \frac{d_m \cdot d_p \cdot K_m}{k_p} \right) [Protein] + k_m K_m = 0, \quad (4.2.4)$$

and there is only one positive solution of steady protein concentration:

$$[Protein]_s = \frac{-d_m d_p K_m + (k_m + V_m + \sqrt{\Delta}) \cdot k_p}{2 d_m d_p}, \quad (4.2.5)$$

where

$$\Delta = \left(k_m + V_m - \frac{d_m d_p K_m}{k_p} \right)^2 + \frac{4 d_m d_p k_m K_m}{k_p}.$$

This is the only physically possible solution of the quadratic equation, because the other solution is less than zero. The Jacobean matrix at this steady point can be written as:

$$F = \begin{bmatrix} -d_m & \frac{v_m K_m}{(K_m + [Protein]_s)^2} \\ k_p & -d_p \end{bmatrix}. \quad (4.2.6)$$

The fact that F is a negative definite matrix over the whole parameter space ensures that this steady state is intrinsically stable, the same as the minimal gene expression model.

However, we find that things are quite different in the dimer-TF model. The cubic behavior results in a different cubic characteristic function y with three possible steady solutions, among which two are stable, or only one stable steady state, depending on the specific ranges of parameter estimation (Figure 4.4).

The dimerization introduces essential parameter dependence into the gene expression model, and causes bifurcation of bistable switching. Further analyses of more complicated intracellular regulations also prove that dimerization of TF proteins is necessary for the existence of the two-state switching, which will be discussed in the following part about quorum sensing network.

In this numerical simulation, we estimate the values of previously mentioned rate parameters from Ref.57.

Figure 4.4 The solution for steady protein concentration in the dimer-TF model.

Function y is the characteristic equation of protein level and the steady solutions are intersection points of curve y and the horizontal line $y=0$. The blue line is plotted with the same parameter values as in the above, and the green line is based on a modification of the mRNA transcription rate and degradation rate. After changing the parameters, we find that the system experiences a bifurcation from being monostable to being bistable, because two of the three steady solutions on green line are stable.

4.3 Modeling Quorum Sensing Network of Agrobacteria

A model has been established to investigate the quorum sensing in the soil-dwelling plant pathogen *Agrobacterium tumefaciens* (25). The bacterial species often harbor Ti (tumor-inducing) plasmids that endow their hosts with the unique ability to genetically modify susceptible plants through a cross-kingdom DNA transfer. In this study, we mainly consider the octopine type Ti plasmids. The layout of the QS network for them is shown in Figure 4.5 based on the experimental data. We assume that octopine is constitutively available at the saturation concentration that results in the maximal rate of *occ* transcription. The last open reading frame of this operon codes for the QS transcription activator TraR. Due to structural similarity with other typical LuxR-type

proteins, TraRs will form protein dimers after they bind with autoinducers (Figure 4.6) (58). From the networking scheme, we have the following kinetic equations. With the same quasi-steady assumption of Michaelis-Menten mechanism for TF-involved transcription, we can reduce the full model to two dynamic equations:

$$\frac{dD}{dt} = A_i \frac{k_6}{2k_{12}} \left(k_7 + \frac{k_{28} \cdot D}{k_M^{msh} + D} \right) - \frac{k_{19}k_{15} \cdot D^2}{k_{18}(k_M^{traM} + D)(\frac{k_{20}}{k_{11}} + D)} - k_{10} \cdot D, \quad (4.3.1)$$

$$\begin{aligned} \frac{dA_i}{dt} = & \frac{k_1k_{25}}{k_{26}k_{24}} \left(k_{27} + \frac{k_{21} \cdot D}{k_M^{trb} + D} \right) + k_2(A_e - A_i) - A_i \frac{k_6}{k_{12}} \left(k_7 + \frac{k_{28} \cdot D}{k_M^{msh} + D} \right) + \\ & \frac{k_{13}k_3 \cdot A_e}{k_{12}k_{14}(k_M^{imp} + A_e)} \left(k_7 + \frac{k_{28} \cdot D}{k_M^{msh} + D} \right). \end{aligned} \quad (4.3.2)$$

Here we define the parameters in the same way as Goryachev did in Ref. 25, and adopt the same parameter values.

When the external autoinducer concentration is less than 40um, the QS network remains off as the intracellular autoinducer concentration stays in relatively low expression level. When A_e increases, the system takes an “off-on” switching among the 40-60um range (Figure 4.7). In the range, there are two possible stable states for the system and the actual stable state dependent on the initial condition. This bistable behavior provides the QS network with robustness to tiny biological stimulus, ensuring the bacteria’s survival when the environment experiences an instant change. After A_e passes 60um, the QS is turned on and the cell starts to produce high level of signaling molecules.

Figure 4.5 Quorum sensing network of an octopine-type *Agrobacterium* Ti plasmid. AAI is the key autoinducer of *Agrobacteria* quorum sensing network, and the TraR* is the regulator protein monomer, whereas the TraRd represents the dimer form of the regulator. (Cited from ref 25.)

In the work of Goryachev et al. (25), they announced that traM is necessary for the existence of the off state. However, after our analytic derivation, we find that this necessity is only based on the reaction parameters, and we can have bistable switching behavior without traM in the network model if some parameters involved in the traR mRNA production change. Following the same routine as we use for characteristic study of single-gene expression model, we calculate the nullclines of the system for the traR dimer concentration (Figure 4.8). There are two stable states out of the three steady solutions in the original

model when Ae enters the critical switching area; so is the same for the model without traM.

Figure 4.6 3D structure of the TraR–DNA complex. Domains in the two monomers are shown in different colours (light/dark orange and light/dark green respectively), the DNA is coloured blue and the OOHL is red. The N terminus and C terminus of the lower subunit are also labelled.

Figure 4.7 Transition to quorum sensing in the ODE system based on intracellular dynamics. The blue line on the bottom represents the off state of the QS switching, while the green line denotes the on state of the network.

Figure 4.8 The nullclines of the dynamic system with and without the anti-activator TraM. (a) is based on the exact parameter values of the work by Goryachev et al.¹⁸ with the TraM functioning as the anti-activator shown in Figure 4.6. (b) shows the nullcline behavior with modified *traR* velocity. The value of *y* function at zero concentration in (b) is positive.

4.4 Conclusion

We apply analytical techniques to study the signaling pathways of quorum sensing network by starting with the minimal gene regulation network. From

the modeling and stability analysis, we find the dimerization process is the critical reaction process for bistability. Then we generalized this method for a specific quorum sensing system of *Agrobacteria*. We confirm that the dimerization of regulator proteins also plays the key role in the switching behavior of the quorum sensing network. This result may be applied to other bacterial species with similar signaling network topology.

Chapter 5

Modelling Tubulin Dimer Dynamics

Microtubule is the basic functional organelle consisting of tubulin dimers, responsible for cell motion and reshaping, so microtubule dynamics is the key to understand cell motion. As the fundamental building blocks of microtubule, the tubulin protein heterodimers contribute to the complexity and robustness in microtubule structure. In this chapter, we build a theoretical model for tubulin dimer to explain the rich dynamics and robustness covering different time scales.

5.1 Position Feedback Control Model of Sensor-Actuator System

The basic position feedback control model of the sensor-actuator system is expressed with two coupled second-order ordinary differential equations:

$$\ddot{X}_a + k_{a1}\dot{X}_a + k_{a2}(X_a - \gamma X_b) = 0 \quad (5.1.1)$$

$$\ddot{X}_b + k_{b1}\dot{X}_b + k_{b2}(X_b - X_a) = 0 \quad (5.1.2)$$

where the X_a and X_b are the generalized state coordinates of sensor and actuator respectively, the k parameters are the damping and elastic constants, and the γ is the coupling coefficient of the feedback control between two subunits (59, 60). To extend this basic form to describe the dynamics in tubulin heterodimer, we assume that the equilibrium position is an external parameter dependent on biochemical signal and other molecular properties than elasticity.

5.2 Coupled Dynamic Model of Tubulin Dimer

Microtubule is an important component of cytoskeleton, a structural intracellular matrix supporting cell shape and function. One microtubule has an outer diameter of about 24 nm and an inner diameter of about 12nm and its length can vary across dozens of micrometers. Microtubules are constructed by tubulin dimers, each of which consists of an alpha and a beta tubulin to form a heterodimer. Alpha tubulin always binds with one GTP molecule and protects it from hydrolysis; beta tubulin may come with GTP molecule or GDP molecule, and the GTP molecule within beta complex can be hydrolyzed to GDP. The GTP-beta type tubulin dimer helps the polymerization of microtubule, whereas the hydrolysis of GTP to GDP may lead to instability and depolymerization of microtubule. As long as the positive (+) end of microtubule has one GTP-beta type tubulin dimer, the whole microtubule can keep growing and even the hydrolysis inside the microtubule will not cause total collapse (61). Molecular dynamic (MD) simulations help us to analyze this scenario by proving the

dimer structure with GTP-beta tubulin is more stable than one with GDP-beta tubulin, which indicates that the bonding is stronger for a dimer in GTP-beta state. Although previous works presented many inspiring thoughts, how the GTP-beta state dimer can help to stabilize the whole microtubule structure has not been well understood.

The tubulin heterodimer resembles the general sensor-actuator system, a simple model which we could study for dynamics of a wide range of time scales. In a tubulin heterodimer, the alpha subunit plays the role of actuator responsible for the feedback control, while the beta subunit works as a signal sensor to coordinate the dimer state based on biochemical signal, for example, the GTP-GDP transition. The topology of this feedback control system is shown in Figure 5.1. The cooperation of alpha and beta tubulins, just like the sensor and actuator, generates rich dynamics from picosecond scale to nanosecond scale, and builds up a trade-off mechanism for the robustness of the heterodimer.

Applying the basic feedback control theory for the coupled system, we model the tubulin dimer as a sensor-actuator system with generalized mechanical coordinates based on elastic properties to study the dynamical behaviors during configuration changes. With the above assumption, we can write the expression of this heterodimer system as follows.

$$\ddot{x}_a + k_{a1}\dot{x}_a + k_{a2}[(x_a - x_{a0}) - \sqrt{\gamma}(x_b - x_{b0})] = 0 \quad (5.2.1)$$

$$\ddot{x}_b + k_{b1}\dot{x}_b + k_{b2}[(x_b - x_{b0}) - \sqrt{\gamma}(x_a - x_{a0})] = 0 \quad (5.2.2)$$

where x_a and x_b denote the abstract mechanical coordinates of the alpha and beta subunits with a coupling coefficient γ , x_{a0} and x_{b0} denote the equilibrium states based on biochemical conditions, and the k s are damping and elastic constants of individual subunits. This modified symmetric expression comes from equations 5.1.1 and 5.1.2 and an additional transformation $X_a \rightarrow x_a$, $X_b \rightarrow \frac{x_b}{\sqrt{\gamma}}$. Since the interaction between the alpha and beta monomers are weaker than the peptide bonding within individual monomers, the coupling coefficient γ falls in the range of (0,1) and it is the same for the normalized symmetric expression $\sqrt{\gamma}$. When γ equals 0, the equations describe a decoupled feedback

control mechanism; and when $\gamma=1$, the linkage between two monomers is as strong as the peptide bond inside of one protein molecule, which describes an ideal strong coupling limit.

Figure 5.1 Scheme of feedback control model for tubulin heterodimer. The input signal stands for the surrounding biochemical environment and events, and the output signal is the corresponding mechanical states of the dimer. The beta subunit works as a signal sensor to track the input, while the alpha subunit plays the role of actuator to provide the feedback control.

5.3 Complexity Analysis of Tubulin Dimer Dynamics

In order to analyze the dynamics of the tubulin dimer, we conduct the Fourier transformation of the equations (5.2.1 and 5.2.2) and then find that the dynamics of the whole system is a coupling of two damping-oscillating motions: one is always underdamping, but the damping type of the other depends on the coupling strength γ . Because the damping and elastic constants of alpha and beta monomers are highly similar, we use the approximation that the $k_{a1} \approx$

$k_{b1} \approx k_1$ ($k_1 = 4.067 \times 10^7 s^{-1}$) and $k_{a2} \approx k_{b2} \approx k_2$ ($k_2 = 4.067 \times 10^{22} s^{-2}$)

to get a simple analytical solution of the two damping-oscillating modes:

$$\begin{aligned}\omega_{1,2} &= \frac{ik_1 \pm \sqrt{\Delta_1}}{2} \\ \omega_{3,4} &= \frac{ik_1 \pm \sqrt{\Delta_2}}{2}\end{aligned}\tag{5.3.1}$$

where $\Delta_1 = 4(1 + \sqrt{\gamma})k_2 - k_1^2$, and $\Delta_2 = 4(1 - \sqrt{\gamma})k_2 - k_1^2$. These delta values distinguish the damping type based on the value of coupling strength: when delta equals zeros, the omegas have repeated roots for critical damping case leading to the fastest relaxation rate of the heterodimer. For tubulin dimers, $k_2 \sim 1e18-1e20$ and $k_1 \sim 1e8-1e9$, so $\omega_{1,2}$ are always underdamping modes but the behavior of $\omega_{3,4}$ depends on the value of γ . The critical damping condition requires a critical γ within (0.9,1). When γ is smaller than the critical value, the system relaxes through two coupled underdamping processes, and when γ is larger than the critical value, we can observe an underdamping process coupled with an overdamping one. Both cases cost longer time for the system to converge than the case when the system is coupled through the critical γ . On top of that, the rich dynamic behavior involved in the conformational change of tubulin dimer requires us to take the dynamics of coupling strength into consideration rather than study the static γ cases. The coupling strength indicates the linkage interaction between alpha and beta subunits, so we assume that this interaction follows the Arrhenius relation with the explicit form of

$\gamma_d = \gamma \cdot e^{-\frac{E}{k_B T}}$, where the E is the energy of the coupling linkage. Although it is hard to get a simple and explicit result out of this dynamical coupling system, it generates rich dynamics over diverse time scales. From the power spectrum analysis, we observe two different motions. In the picosecond scale from 1e-12s to 1e-10, the system oscillates with a cluster of high frequency vibration modes due to single monomer elasticity. On the other hand, another time scale above sub-nanosecond is retrieved, which corresponds to the energy transition

dynamics between two monomers. As shown in the power spectrum (Figure 5.2), the dense cluster of high frequency peaks characterizes the damping-oscillating motion of individual monomers, and the low frequency peak shows the energy transition rate between two monomers. Large γ values around strong coupling limit push the low frequency peaks to high frequency ranges, making two vibration motions indistinguishable. This leads to failure in the relaxation of total energy, thus cause a loss in robustness. A decreasing γ elongates the time cost of energy transition between the subunits and the time cost goes to infinity in the decoupling limit. In conclusion, the GDP-beta configuration helps build a slow and robust dynamic motion for energy transition whereas the NF states nearly forbid the energy exchanges.

Figure 5.2 Power spectra of tubulin dynamics under different coupling strengths. From(a) to (h), coupling strengths decrease in the series of 5, 3, 1, 0.5, 0.3, 0.1, 0.05, 0.03, and 0. There is no coupling in (i), which corresponds to a decoupled case. The blue lines show the power spectrum of the beta subunit while the green lines denote the alpha subunit.

5.4 Dynamical Robustness of Tubulin Dimer

Robustness is one of the most significant principle of biological system, especially when the scale goes down to molecular level. How a cell and the organelles can sustain their normal functions against a wide range of signal fluctuations remains a mystery for both theorists and experimentalist. It has been one of the major concerns of biological robustness that how the heterodimer remains its normal function when exposed to a variety of fluctuations (62). With the simple expression of ODE system (equation 5.2.1 and 5.2.2), we are able to analyze the dynamic robustness without the strict limit of linearization around steady state. We calculate the robustness as defined in previous works with Fourier transformation on time and get the frequency dependent form of sensitivity function S on external signal fluctuations.

$$S_0(\omega) = \frac{k_{12}(-\omega^2 - i\omega k_{21} + k_{22})}{(-\omega^2 - i\omega k_{11} + k_{12})(-\omega^2 - i\omega k_{21} + k_{22}) - \gamma k_{12} k_{22}}$$

$$S_1(\omega) = \frac{i\omega k_{11}(-\omega^2 - i\omega k_{21} + k_{22})}{(-\omega^2 - i\omega k_{11} + k_{12})(-\omega^2 - i\omega k_{21} + k_{22}) - \gamma k_{12} k_{22}}$$
(5.4.1)

S_0 and S_1 stand for the sensitivities of the mechanical state of beta monomer when come up against 0-order and 1-order fluctuations, in other words, the noise from position variation and from momentum variation. The moduli of S functions show the effect of external fluctuation on final equilibrium state, and the imaginary part is responsible for a phase shift when the system is exposed to an oscillating signal. When the moduli grow, the dynamics of whole tubulin dimer is more prone to be affected by the fluctuation. The behavior of sensitivity functions over the parameter space shows a trade-off mechanism for the overall robustness. The decoupled case (with no coupling, γ equals 0) gives smaller sensitivity in low frequency range ($0-10^9/s$) compared with the coupled case; however, in the nanosecond and picosecond scale (corresponds to $10^{10}-10^{12}/s$), the decoupled system has larger sensitivity. When the signal frequency

decreases, both $|S_0|$ and $|S_1|$ have fast increasing trend with coupling approaching 1, indicating that a solid coupling raises the sensitivities of the heterodimer of low frequency fluctuations. However, when the signal frequency is larger than certain critical value (around $10^{10}/s$), the increasing coupling helps to build a fluctuation-filtering channel to lower the sensitivities. Thus, the scheme here establishes the trade-off mechanism of coupling strength between two time scales: a vanishing coupling closes the filtering channel and makes the heterodimer vulnerable on nanosecond variations; an extreme strong coupling leads to the divergence of sensitivity in low frequency range, so the optimized coupling should depend on the major time-scale the system is designed to track. This trade-off mechanism provides an intuitive understanding of the advantage of the heterodimer structure. When the dimer is in the GTP-beta state with a strong coupling between two subunits, it can catch up with external slow signal to function as the building block of microtubule and filter out the nanosecond variations to ensure the success of assembly process; after hydrolysis triggers GTP to GDP transition and the coupling becomes weaker, the dimer closes the filtering channel against nanosecond events, thus the noise from molecular level will lead to instability of GDP-beta state and finally result in the dissociation.

Figure 5.3 The sensitivity function of frequency and coupling coefficients. The top figure denotes the sensitivity function of 0 order (position of state) fluctuation and the bottom figure shows that of 1 order (momentum of state) fluctuation. Both present similar pattern for gamma dependence: during low frequency ($<10^8$ s) scale, the coupling raises the sensitivity, and when the signal fluctuation is around nanosecond scale, the increasing gamma creates a filtering channel to decrease the effect of these signal fluctuations.

5.5 Conclusion

We model the tubulin dimer dynamics as a coupled feedback control network based on its structural similarity with a sensor-actuator system. By this modeling method, we successfully generate the multi-timescale dynamics that fits numerical simulation results. Different tubulin dimer states show diverse behaviours in their power spectrums, which explains the time evolution patterns observed from molecular dynamics study. Also, we quantitative analyse the robustness feature of tubulin heterodimer structure. We discover that the GTP-

tubulin dimer is the optimal for microtubule assembly process, whereas the GDP-tubulin dimer is responsible for the maintenance of microtubule when kept away from external fluctuations.

Chapter 6

Reaction-Diffusion Model of Bacterial Stratification

Stratification pattern of bacterial distribution has been observed in meromictic lakes. Ace lake, a meromictic lake located in Antarctica, provide us with an ideal environment for quantitative study of this bacterial spatial pattern on large scale. In this chapter, we apply reaction-diffusion theory to simulate the bacteria dynamics in Ace lake and show that this mechanism coupled with bacterial metabolic reactions can successfully explain the layering structure of microbe ecology.

6.1 Bacteria Distribution in Ace Lake

Ace lake is a great system for modeling analysis due to its remarkable stability. As a meromictic lake, Ace lake is permanently stratified (8, 33, 34). It has an ice cover of 1-2m thickness during 11-12 months of the year (Figure 6.1), which ensure that Ace lake is a closed system for most time of the year. Thus, the long-time behaviour of Ace lake can be regarded as the dynamics govern by its own biochemical cycle. The max depth of Ace lake is 25m, and the chemocline appears at around 10-12m deep. The zone above the chemocline is an aerobic mixolimnion from 0-10m, and the zone below is a anoxic monimolimnion from 12-25m. The average temperature of lake water is around 0°C. The bottom water is permanently 1-2°C and the water near the surface can increase to around 10°C in summer. The sulphate trace amounts at 18-24m, increasing to 8-9mM at the chemocline, and decreasing to 2-3mM below the ice sheet. The total microbial concentration is around 1×10^5 cells/mL at 4m, increasing to around 1×10^6 cells/mL at 24m and in sediment layer, and up to 8×10^6 cells/mL in blooms of cyanobacteria (33).

Improvement in genomic technology has helped us to analyse the huge data of the mixed environment consisting of diverse microbe species. Both the diversity and the population density of bacteria and virus species have been quantified against the depth of Ace lake (34). After successfully identifying the species, people figure out the biochemical cycle of individual elements and build the biochemical networks of major contributors. Since both the biochemical networking and spatial observation of microbe distribution have become clear and quantified, we are now able to apply theoretical methods for modeling this semi-closed cold aqua system for its specific microbe distribution pattern. In this chapter, we mainly focus on the sulfur cycle and sulfur related bacteria for generality. The green sulfur bacteria (GSB) shows up at the chemocline and monotonically decreases as depth grows. The sulfur reducing bacteria (SRB) has little concentration at the chemocline but dramatically

increases with depth in a thin layer below the chemocline and then starts to drop. These two species build the basic sulfur oxidation-reduction cycle in the monimolimnion.

Figure 6.1 A view of Ace lake environment condition. This photo is cited from ref 1. It shows the ice-covered state of Ace lake.

6.2 Metabolic Reaction Model for Bacteria-Nutrient Dynamics

To model the whole bacterial ecology system, we need to start with the fundamental metabolism of bacterial cells. Here we model the basic bacteria metabolism as a eating-growing-dying dynamics. We only take three factors into account: the nutrient eating process, the growing process based on the fulfilling of nutrient supply, and the self-dying process. The reaction equations are shown below:

where B stands for the bacterium, N stands for the nutrient, and W stands for the waste, the product from B metabolism that cannot be used by itself.

Parameter alpha shows the growing ratio of per metabolic cycle, which corresponds to the nutrient utility rate. After applying the mass action law of biochemistry, we derive the following rate equations for the metabolic dynamics.

$$\frac{d[N]}{dt} \Big|_{reaction} = -k_f[B][N] + k_r[B \cdot N], \quad (6.2.1)$$

$$\frac{d[B]_t}{dt} \Big|_{reaction} = \frac{d[B]}{dt} + \frac{d[B \cdot N]}{dt} = (\alpha - 1)k_{cat}[B \cdot N] - k_d[B]_t, \quad (6.2.2)$$

$$\frac{d[W]}{dt} \Big|_{reaction} = k_{cat}[B \cdot N], \quad (6.2.3)$$

where the brackets denote the concentration of each species, k_f and k_r are the reaction rate constants for forward and backward eating process, k_{cat} is the catalytic reaction rate constant for the digestion process, and k_d shows the dying rate of the bacterial cell.

Assumption 1: Fast digestion equilibration

$$k_f[B][N] - (k_r + k_{cat})[B \cdot N] = 0 \quad (6.2.4)$$

To simplified the coupled dynamics of these molecular species, we apply our first assumption here that the digestion process reaches fast equilibrium compared with the dying process. So we get

$$\frac{d[N]}{dt} \Big|_{reaction} = -\frac{1}{\alpha-1} \left\{ \frac{d[B]_t}{dt} \Big|_{reaction} + k_d[B]_t \right\}, \quad (6.2.5)$$

$$\frac{d[W]}{dt} \Big|_{reaction} = -\frac{d[N]}{dt} \Big|_{reaction}. \quad (6.2.6)$$

These two equations describe the relations between the dynamics of three individual species. Since there is no explicit molecular reaction rate parameter involved, these equations set up a scale-independent correlation which can be applied to all bacteria-nutrient dynamic systems. The only parameters required in 6.2.5 and 6.2.6 are the nutrient utility rate alpha and the dying rate of the bacterium.

6.3 Reaction-Diffusion Theory for a Microbial Relay Model

Theoretical modelling in previous subchapter describes the metabolic reactions in homogenous environments or local conditions. To quantitatively explain the stratification pattern, the spatial dependence needs to be introduced by the reaction-diffusion mechanism (63). This mechanism assumes that the spatial-dependent concentration dynamics is based on the linear addition of local chemical reactions and self-diffusion process of each molecular species. Besides, in order to model the multi bacterial species network, it is necessary to build their linkage through signaling network consisting of biochemical cycles. In Ace lake system, these factors contribute to a microbial relay model where there are two bacterial species and the second bacterium (SRB) feeds on the metabolic product (elementary sulfur or sulfur oxide) of the first bacterium (GSB). Since the metabolism of GSB relies on the sunlight intensity, we regard this factor as an external supply.

The dynamics of reaction-diffusion system for the relay model is written as follows:

$$\frac{\partial[N_1]}{\partial t} = D_{N_1} \nabla^2[N_1] + \frac{\partial[N_1]}{\partial t}|_{reaction}, \quad (6.3.1)$$

$$\frac{\partial[B_1]}{\partial t} = D_{B_1} \nabla^2[B_1] + \frac{\partial[B_1]}{\partial t}|_{reaction}, \quad (6.3.2)$$

$$\frac{\partial[N_2]}{\partial t} = D_{N_2} \nabla^2[N_2] + \frac{\partial[N_2]}{\partial t}|_{reaction}, \quad (6.3.3)$$

$$\frac{\partial[B_2]}{\partial t} = D_{B_2} \nabla^2[B_2] + \frac{\partial[B_2]}{\partial t}|_{reaction}, \quad (6.3.4)$$

where B1 and B2 stand for the two bacterial species, N1 denotes the primary resource from external supply, and N2 denotes the product from B1 metabolic reaction and also the food to maintain B2 metabolism. The parameters Ds are the diffusion constants of each species, which can be estimated from experimental observations about bacterial chemotaxis. The second terms on the right hand side of the equations come from the basic metabolic reactions defined in previous subchapter.

With the PDE (partial differential equation) system given above, we can set the steady state condition as all the time derivatives vanishes. So for the steady state,

$$D_{N_1} \nabla^2 [N_1] = \frac{1}{\alpha_1 - 1} \{-D_{B_1} \nabla^2 [B_1] + k_{d1} [B_1]\}, \quad (6.3.5)$$

$$D_{N_2} \nabla^2 [N_2] = \frac{1}{\alpha_1 - 1} \{D_{B_1} \nabla^2 [B_1] - k_{d1} [B_1]\} + \frac{1}{\alpha_2 - 1} \{-D_{B_2} \nabla^2 [B_2] + k_{d2} [B_2]\}. \quad (6.3.6)$$

Equation 6.3.5 shows the space distribution correlation of primary nutrient and first bacteria species, while equation 6.3.6 shows the correlation between the intermediate nutrient and both bacteria.

Assumption 2: Total available volume for bacteria living is limited, i.e., there exists a maximal limit on the concentrations of bacteria, or the total numbers of bacteria cells.

$$V_{B1} [B_1] + V_{B2} [B_2] \leq V_0$$

Here we introduce our second assumption of the theory. This assumption is based on the physical intuition that every bacterial cell needs spatial volume to maintain their basic biological existence. It is different from a simple point particle model where every particle occupies no volume.

Applying the second assumption, we directly have the saturation pattern depending on the resource supply level. When the primary resource supply exceeds a critical value, one or more saturation layers where the total biological living space is filled up with bacterial cells can be observed.

For the analytical solution of the steady state distribution of the resource relay model, we use the Laplace transformation for the steady state correlations (equations 6.3.5 and 6.3.6). Finally we obtain three possible spatial areas: B1 saturation layer, B1&B2 saturation layer, and free space. Following equations express the concentration-depth relation in all three possible ranges.

$$[B_1] = \begin{cases} \frac{V_0}{V_{B1}} & x < x_{s1} \\ \frac{V_0}{V_{B1}} \cdot \exp[-k_1(x - x_{s1})] & x \geq x_{s1} \end{cases} \quad (6.3.7)$$

$$[N_1] = \begin{cases} \frac{k_{d1}}{2D_{N1}(\alpha_1-1)} \cdot \frac{V_0}{V_{B1}} (x - x_0)^2 + C_{01} & x < x_{s1} \\ \frac{1}{A_1} \cdot \frac{V_0}{V_{B1}} \cdot \exp[-k_1(x - x_{s1})] & x \geq x_{s1} \end{cases} \quad (6.3.8)$$

$$[B_2] = \begin{cases} 0 & x < x_{s1} \\ \frac{V_0}{V_{B2}} \cdot \{1 - \exp[-k_1(x - x_{s1})]\} & x_{s1} \leq x < x_{s2} \\ \frac{V_0}{V_{B2}} \cdot \{\exp[-k_1(x - x_{s2})] - \exp[-k_1(x - x_{s1})]\} & x \geq x_{s2} \end{cases} \quad (6.3.9)$$

$$[N_2] = \begin{cases} -\frac{k_{d1}}{2D_{N2}(\alpha_1-1)} \cdot \frac{V_0}{V_{B1}} x^2 + C_{02} & x < x_{s1} \\ \frac{1}{D_{N2}k_1^2} \cdot \frac{V_0}{V_{B1}} \cdot \left\{ \left[\frac{D_{B1}}{\alpha_1-1} + \frac{D_{B2}}{\alpha_2-1} \cdot \frac{V_{B1}}{V_{B2}} \right] k_1^2 - \left[\frac{k_{d1}}{\alpha_1-1} + \frac{k_{d2}}{\alpha_2-1} \cdot \frac{V_{B1}}{V_{B2}} \right] \right\} \cdot \exp[-k_1(x - x_{s1})] + \\ \frac{k_{d2}}{2D_{N2}(\alpha_2-1)} \cdot \frac{V_0}{V_{B2}} (x - x_3)^2 + C_{03} & x_{s1} \leq x < x_{s2} \\ \frac{1}{A_2} \cdot \frac{V_0}{V_{B2}} \cdot \{\exp[-k_1(x - x_{s2})] - \exp[-k_1(x - x_{s1})]\} & x \geq x_{s2} \end{cases} \quad (6.3.10)$$

where the decay rate constant k_1 is defined as:

$$k_1 = \sqrt{\frac{k_{d1}}{\left(\frac{\alpha_1-1}{A}\right)D_{N1}+D_{B1}}}. \quad (6.3.11)$$

x_{s1} and x_{s2} are the ending depth of the first and second saturation layer; and x_0 and x_3 are two depth parameters to be determined in following analysis.

The three layers are theoretical possible, the actual existences of this stratification structure is up to the supply level of the first resource species. To quantitatively show this dependence, we manage to calculate the saturation layer thickness. Only if the supply flux ensures that one layer thickness is larger than zero, that specific layer can be observed.

The saturation layers' thicknesses are determined by the supply flux of primary resource at the system boundary($x=0$). Then we calculate the boundary condition at x_{s1} and x_{s2} for the values of x_0 , x_{s1} , x_{s2} , and x_3 .

$$x_0 = \frac{(\alpha_1 - 1)}{k_{d1}} \cdot \frac{V_{B1}}{V_0} \cdot f \quad (6.3.12)$$

$$x_{s1} = \frac{(\alpha_1 - 1)}{k_{d1}} \cdot \left[\frac{V_{B1}}{V_0} \cdot f - \frac{D_{N1} k_1}{A_1} \right] \quad (6.3.13)$$

$$x_3 = x_{s1} \cdot \left\{ \frac{k_{d1}(\alpha_2 - 1)V_{B2}}{k_{d2}(\alpha_1 - 1)V_{B1}} + 1 \right\} - \frac{(\alpha_2 - 1)V_{B2}}{k_1 k_{d2} V_{B1}} \cdot \left\{ \left[\frac{D_{B1}}{\alpha_1 - 1} + \frac{D_{B2}}{\alpha_2 - 1} \cdot \frac{V_{B1}}{V_{B2}} \right] k_1^2 - \left[\frac{k_{d1}}{\alpha_1 - 1} + \frac{k_{d2}}{\alpha_2 - 1} \cdot \frac{V_{B1}}{V_{B2}} \right] \right\} \quad (6.3.14)$$

$$\frac{D_{N2} k_1}{A_2} \cdot \frac{V_0}{V_{B2}} \cdot \{1 - \exp[-k_1(x_{s2} - x_{s1})]\} = \frac{1}{k_1} \cdot \frac{V_0}{V_{B1}} \cdot \left\{ \left[\frac{D_{B1}}{\alpha_1 - 1} + \frac{D_{B2}}{\alpha_2 - 1} \cdot \frac{V_{B1}}{V_{B2}} \right] k_1^2 - \left[\frac{k_{d1}}{\alpha_1 - 1} + \frac{k_{d2}}{\alpha_2 - 1} \cdot \frac{V_{B1}}{V_{B2}} \right] \right\} \cdot \exp[-k_1(x_{s2} - x_{s1})] - \frac{k_{d2}}{\alpha_2 - 1} \cdot \frac{V_0}{V_{B2}} (x_{s2} - x_3) \quad (6.3.15)$$

Equations 6.3.12, 6.3.13, and 6.3.14 give the explicit expression of x_0 , x_{s1} , and x_3 , whereas the value of x_{s2} can be calculated by numerical solution of equation 6.3.15, which could be simplified as $C'' \cdot \exp(-k_1 x_{s2}) + C' \cdot x_{s2} + C = 0$. A qualitative concentration profile is shown in Figure 6.2. All parameters involved in the plot is normalized to the standard scale for simplicity. The thickness of the first saturation layer and that of the second are both scaled to 1 meter.

Figure 6.2 The qualitative concentration curves against depth. The blue curve denotes the primary resource level and the red curve denotes the secondary resource level. The first saturation layer takes the depth from 0 to 1 meter, while the second saturation layer takes from 1 to 2 meter.

6.4 Conclusion

We build theoretical model to explain the stratification pattern of bacterial distribution in Ace lake. Combining basic metabolic reactions and diffusion process to simulation the chemotaxis and cell life cycle, we apply the reaction-diffusion theory for the dynamics. One key assumption introduced here is the space limit for bacteria growth, which is from physical intuition that living cells need to take spatial volume for maintenance. With the help of modern theoretical methods, we succeed in explain the stratification pattern in a quantitative fashion. Also with the simple expression of the saturation layer

thickness, our theory can predict the location and thickness feature of the microbial stratification pattern. This model and calculation method can be applied to other meromictic lake system for its generality and prediction power.

Chapter 7

Discussion and Future Work

This thesis includes original studies on theoretical modeling works on three different biological scales. In this chapter, we address the significance of each subject and discuss the possible applications based on our theoretical methods. Due to the prediction power of mathematical models, we also provide testable results for future experiments.

7.1 Discussion

Biology is such a broad discipline that it essentially required interdisciplinary contributions from other fields. In my thesis, we apply several theoretical methods to establish dynamic models for biological systems on diverse scales. These general and systematic ways help to build fundamental understanding of the biological interactions and correlations. In this session, we will discuss the impact of these modeling works.

7.1.1 Discussion on Quorum Sensing Network Analysis

Quorum sensing network is responsible for the bacterial regulation based on the population density. A typically quorum sensing signalling pathways may involve more than 20 molecular species. To identify the crucial network motif or topological feature of signaling network, we build the mathematical model for the quorum sensing pathways and use theoretical analysis to reveal the criticality out of the complexity.

We start the modeling analysis with the minimal gene regulation network with transcription factor. After making the basic assumption and approximation on the reaction rate and DNA level, we derive the explicit dynamic equations of regulatory network, which are proved to have stable steady solutions depending on the dimerization of transcription factor (TF). The dimerization process of TF is the key factor for the on-off switching of gene regulation. If there is no such dimerization process, then the system will lose its distinguishable on-off gene expression states.

Then we apply the general ideas developed by the minimal gene regulation model for the quorum sensing system. We focus on the quorum sensing network of *Agrobacteria*, a Gram-negative bacterium found in plants. Based on the regulation network on molecular level, we build the ordinary differential

equation (ODE) dynamics of *Agrobacteria* quorum sensing system, by which we are able to solve both the time-dependent evolution of the whole network and the steady states for cellular homeostasis. Analytic treatments generate similar result with that from the minimal gene regulation model: the dimerization of regulatory protein *traM* is the key for the bistability of quorum sensing network. Our model proves that the TF dimerization leads to not only two distinguishable steady states for on and off switching but also the stability of these two states. The non-contiguous steady changing is responsible for the biofilm-planktonic state switching of *Agrobacteria*, and if the dimerization is missing, the bacterium will have no biofilm state. The stability feature of both states is also important for state maintenance since there exist thermal and external fluctuations which may destroy a unstable steady state. Besides, we argue that our theory is not highly sensitive to systematic parameters, indicating that even when certain parameters varies over 1 order of magnitude, the conclusion of our model have not be harmed.

7.1.2 Discussion on Dynamic Model for Tubulin Dimer

Dynamical analysis of protein complex is always a challenge for both experimentalists and theorists. Tubulin, the basic functional unit of microtubule, has attracted intensive researches for its significance in structural feature and complex dynamics. In this work, we build a biophysical model to explain the complexity and robustness in tubulin heterodimer dynamics. The heterodimer structure of alpha-beta tubulin complex, as well as its state switching behaviour between the GTP-bound state and the GDP-bound state in beta subunit, inspires us to model the tubulin dimer as a sensor-actuator system for their structural and functional similarity. Based on the simple sensor-actuator feedback control network, we derive the governing dynamics of the coupled alpha-beta tubulin system. The dynamic system is expressed in equations 5.2.1 and 5.2.2. In these

equations, we model the two subunits as elastic oscillators coupled through an interaction force indicated by the nucleotide bound condition, where we assume that the effective elastic property comes from the bonding strength within one tubulin monomer. We apply power spectrum analysis to show the rich dynamics over different time scales of the heterodimer. These dynamical behaviors are highly dependent on the coupling strength, in other words, the nucleotide bound state. When the dimer is in GTP-bound state, it has a strong coupling, resulting in high frequent motions around picosecond scale; when it is in GDP-bound state, the heterodimer's coupling strength is relatively weak thus it has more low frequency motions around nanosecond scale. The extreme case is that of the nucleotide free state when no nucleotide is bound with the beta subunit, the dynamics of two subunits are totally separate and each decay into a monotonic motion.

On top of that, the robustness analysis enlightens the functional behaviors of different tubulin dimer states. The large coupling strength of the GTP-bound state results in high sensitivity for low frequent signals which corresponds to the assembly process of the microtubule triggered by local signals of GTP concentration. On the other hand, the GDP-bound state, indicating a small coupling constant, is more sensitive for specific signalling around 1 to 100 picosecond. This explains the observation that the microtubule consisting of the GDP-bound tubulin dimers have a high tendency to collapse due to the fluctuations on molecular level.

7.1.3 Discussion on Microbial Stratification Model of Ace Lake

Recent years have witnessed a burst in experimental studies about meromictic lakes and their ecology. The observations that the stratification structure of both water and microbe layers have not been well understood. Lake level observations provide biological data across meters, which is hardly possible to

be achieved within modern labs. Ace lake, due to its specific location and condition, makes an ideal system for theoretical analysis. In this work, we focus on the general microbial families and biochemical cycle of sulfur in Ace lake. We start our modeling work with the fundamental analysis on bacterial metabolic dynamics. Our model provides a simple and deep understanding of the nutrient-based metabolism which remains valid for different biological scales from cellular scale to colony scale. To understand the spatial pattern of bacteria distribution inside the lake, we build the reaction-diffusion mechanism into the model and succeed in obtaining the layering structure. With detailed analytical work, we calculate the supply need for this stratification pattern and get the thickness of each saturation layer. With the provided explicit expressions, people can easily make quantitative predictions from our theoretical methods. Since our model is mainly analytical, its mathematical results are not very sensitive to parameter variation, which supports its application for other meromictic lake systems.

7.2 Future work

This PhD project covers broad subjects about theoretical methods for biology on different scales. However, to have a complete view of biology, it is essential for us to combine the theory with the experiment. In this session, we will discuss the future work from both the theoretical and the experimental angles.

7.2.1 Theoretical Improvement on Modelling Work

Theoretical modeling of signaling pathways in diverse biological systems is a promising topic during recent years. People have started to seek theoretical

interpretation and prediction for experimental observation and design. This work concentrates on fundamental modeling techniques on specific subjects. First, our models by themselves are built with good application potential for similar biological systems. The quorum sensing network model introduced in this thesis is based on the virulence behaviour of *Agrobacteria*, but the modeling analysis can be directly applied to other quorum sensing system. For example, *V. fischeri* use quorum sensing response for bioluminescence with the famous LuxI/LuxR regulation network which has similar regulator protein structural and functional features. So we can apply the same assumption and thus get similar dynamic expression for *Vibrio fischeri* quorum sensing. Together with the numerical methods, our model can predict the critical autoinducer level where the state change occurs. In addition to *V. fischeri*, *P. aeruginosa*, *V. cholera*, and *Xcc*, lay in our future work scope due to the similarity of the key regulation motifs of individual quorum sensing network. At the same time, the specific features of each new bacterial species will add the modeling content and improve our understanding of new signaling pathways. Early work of this generalization has been in process.

Second, the theoretical works discussed in this thesis still focus on the simplicity of ideal biological subjects. When we further our research into real conditions of complex biological networks with varying noise and fluctuation, we need to take more of complexity into account (49, 53-56). In our model for tubulin dynamics, we build the physical understanding of a single tubulin heterodimer. However, to fully understand the dynamics of microtubule, we are required to push the complexity of our model to a high level where multiple tubulin heterodimers show assembly and disassembly process governed by the surrounding nucleotide states and concentrations. This higher complexity will not only help us to analyse the dynamics of microtubules but also make one more step further towards the dynamics on cell level.

Third, we want to expand our modeling scope to different levels of cell biology. The models in this thesis cover the biological levels of the DNAs, the proteins

and protein complexes, and the cellular motion. We plan to include the different RNA species into theoretical analysis. RNA has the wide molecular families including those of small and large molecular weights and simple and complex 3D structures (64-67). They play different roles in gene regulation, and some functions of non-coding RNA have not been understood. We believe that the analytical and numerical methods will help the development of modern concepts about RNA families.

Last but not the least, we aim to introduce more numerical methods into the future modeling work. Modern advancements in computational theory and technology have sped up the quantitative study in most theoretical fields from physics to chemistry. Bioinformatics is such a discipline that created by the combination of big data from biological experiments and numerical interpretations from computational methods. It is beneficial to introduce similar ideas into the modeling works and we plan to expand our modeling scope with two main topics: modeling analysis coupled with biological big data and stochastic dynamics of complex signaling networks.

7.2.2 Prediction for Future Experimental Work

Biology is never a pure theoretical discipline, so when we continue to advance our theoretical understanding, we have to keep in mind that the theory should give both explanation for real observations and predict for experiments. In this thesis, the explanations from modeling point of view have been discussed, but the predictions of future experiments have not. We intend to improve our parameter estimation and generate more accurate quantitative predictions. For example, in quorum sensing model, we argue that the dimerization of the regulator is the essential factor for bistability, so we plan to further our study to identify the key residues involved in the dimerization process to provide mutation suggestion for future experiments on the bistability behaviour of quorum sensing network.

Also, for the stratification model of microbial distribution, we focus on the Ace lake. In the future, we plan to put more of numerical analysis into this part for quantitative predictions which can be applied for outdoor examinations in other meromictic lakes. We will seek the best chances to combine our theories with modern experimental techniques.

References

References

- [1] Fuqua, C., Parsek, M. R. & Greenberg. *Annual review of genetics* **35**, 439-468 (2001).
- [2] Elias, S. & Banin. *FEMS microbiology reviews* (2012).
- [3] Miller, M. B. & Bassler, B. L. *Annual Review of Microbiology* **55**, 165-199 (2001).
- [4] Kholodenko, B. N. *Nature reviews. Molecular cell biology* **7**, 165-176 (2006).
- [5] Neves, S. R. & Iyengar, R. *Bioessays* **24**, 1110-1117 (2002).
- [6] Kang, H.-W. & Kurtz, T. G. *The Annals of Applied Probability* **23**, 529-583 (2013).
- [7] Wolf, D. M. & Arkin, A. P. *Current Opinion in Microbiology* **6**, 125-134 (2003).
- [8] Cavicchioli. *Nature Reviews Microbiology* **13**, 691-706 (2015).
- [9] Kanehisa, M. *et al. Nucleic Acids Research* **34**, D354-D357 (2006).
- [10] Gibson, M. A. & Bruck, J. *Journal of Physical Chemistry A* **104**, 1876-1889 (2000).
- [11] Romond, P. C., Rustici, M., Gonze, D. & Goldbeter, A. in *Tempos in Science and Nature: Structures, Relations, and Complexity* Vol. 879 *Annals of the New York Academy of Sciences* (eds C. Rossi, S. Bastianoni, A. Donati, & N. Archettini) 180-193 (New York Acad Sciences, 1999).
- [12] Waters, C. M. & Bassler, B. L. in *Annual Review of Cell and Developmental Biology* Vol. 21, 319-346 (2005).
- [13] Lee T, T. C., Tu D, You L. in *Bioinformatics: An Engineering Case-Based Approach* Ch. 6, (Artech House Publishers, 2007).
- [14] Ideker, T., Winslow, L. R. & Lauffenburger, D. A. *Annals of Biomedical Engineering* **34**, 1226-1233 (2006).
- [15] Karatan, E. & Watnick, P. *Microbiology and Molecular Biology Reviews* **73**, 310 (2009).

References

- [16] Costerton, J. W., Stewart, P. S. & Greenberg, E. P. *Science* **284**, 1318-1322 (1999).
- [17] Beech, W. B. & Sunner, J. Biocorrosion. *Current Opinion in Biotechnology* **15**, 181-186 (2004).
- [18] Wingender, J. & Flemming, H. C. *International Journal of Hygiene and Environmental Health* **214**, 417-423 (2011).
- [19] Characklis, W. G. *Biotechnology and Bioengineering* **23**, 1923-1960 (1981).
- [20] McDougald, D., Rice, S. A., Barraud, N., Steinberg, P. D. & Kjelleberg, S. *Nature Reviews Microbiology* **10**, 39-50 (2012).
- [21] Stewart, P. S. & Franklin, M. J. *Nature Reviews Microbiology* **6**, 199-210 (2008).
- [22] Davies, D. *Nature Reviews Drug Discovery* **2**, 114-122 (2003).
- [23] Berk, V. *et al. Science* **337**, 236-239 (2012).
- [24] Goryachev, A. B., Toh, D. J. & Lee, T. *Bio Systems* **83**, 178-187 (2006).
- [25] Goryachev, A. B. *et al. PLoS Computational Biology* **preprint**, e37 (2005).
- [26] Fuqua, W. C., Winans, S. C. & Greenberg, E. P. *Journal of Bacteriology* **176**, 269-275 (1994).
- [27] Ward, J. P. *et al. Ima Journal of Mathematics Applied in Medicine and Biology* **18**, 263-292 (2001).
- [28] Kiratisin, P., Tucker, K. D. & Passador, L. *Journal of Bacteriology* **184**, 4912-4919 (2002).
- [29] Dockery, J. D. & Keener, J. P. *Bulletin of mathematical biology* **63**, 95-116 (2001).
- [30] Ng, W. L. & Bassler, B. L. *Annual review of genetics* **43**, 197-222 (2009).
- [31] Lewis, William M., Jr. *Canadian Journal of Fisheries and Aquatic Sciences*. 40 (10): 1779–1787 (1983).

References

- [32] Wetzel, Robert G. *Limnology: Lake and River Ecosystems* (Third ed.). New York: Academic Press (2001).
- [33] Cavicchioli. *Nature Reviews Microbiology*, 4, 331-343 (2006).
- [34] Lauro, etc. *The ISME Journal*, 5, 879-895 (2011).
- [35] Lampert, Winfried & Sommer, Ulrich. *Limnoecology: The Ecology of Lakes and Streams*. Translated by James F. Haney. Oxford University Press (1997).
- [36] Fry, Brian. *Limnology and Oceanography*. 31 (1): 79–88 (1986).
- [37] Desai A. & Mitchison TJ. *Annu Rev Cell Dev Biol*. 13: 83–117 (1997).
- [38] Walker RA, O'Brien ET, Pryer NK, Soboeiro MF, Voter WA, et al. *J. Cell Biol*. 107: 1437–48 (1988).
- [39] Gunning PW, Ghoshdastider U, Whitaker S, Popp D, Robinson RC. *Journal of Cell Science*. 128 (11): 2009–19 (2015).
- [40] Pilhofer M, Ladinsky MS, McDowall AW, Petroni G, Jensen GJ. *PLoS Biology*. 9 (12): e1001213 (2011).
- [41] Arkin, A., Ross, J. & McAdams, H. H. *Genetics* **149**, 1633-1648 (1998).
- [42] Lee T, T. C., Tu D, You L. in *Bioinformatics: An Engineering Case-Based Approach* Ch. 6, (Artech House Publishers, 2007).
- [43] Alves, R., Antunes, F. & Salvador, A. *Nature Biotechnology* **24**, 667-672 (2006).
- [44] Orlando, D. A. *et al. Nature* **453**, 944-947 (2008).
- [45] Kholodenko, B. N. *Nature reviews. Molecular cell biology* **7**, 165-176 (2006).
- [46] Kanehisa, M. *et al. Nucleic Acids Research* **34**, D354-D357 (2006).
- [47] Overbeek, R. *et al. Nucleic Acids Research* **31**, 164-171 (2003).
- [48] Gasteiger, E. *et al. Nucleic Acids Research* **31**, 3784-3788 (2003).
- [49] Gillespie, D. T. *Physica A* **188**, 404-425 (1992).
- [50] Bintu, L. *et al. Current Opinion in Genetics & Development* **15**, 125-135 (2005).

References

- [51] Bintu, L. *et al.* *Current Opinion in Genetics & Development* **15**, 116-124 (2005).
- [52] Hedges, R. M. & Rabitz, H. *Journal of Chemical Physics* **82**, 3674-3684 (1985).
- [53] Gillespie, D. T. *Journal of Physical Chemistry* **81**, 2340-2361 (1977).
- [54] Gibson, M. A. & Bruck, J. *Journal of Physical Chemistry A* **104**, 1876-1889 (2000).
- [55] Cao, Y., Gillespie, D. T. & Petzold, L. R. *Journal of Chemical Physics* **124** (2006).
- [56] Haseltine, E. L. & Rawlings, J. B. *Journal of Chemical Physics* **117**, 6959-6969 (2002).
- [57] Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. *Nature genetics* **31**, 69-73 (2002).
- [58] Zhang, R. G. *et al.* *Structure of a bacterial quorum-sensing transcription factor complexed with pheromone and DNA* (vol 417, pg 971, 2002).
- [59] J. L. Fanson and T.K. Caughey. *AIAA Journal* VOL.28, NO. 4, 717-724 (1990).
- [60] A. S. Zeiger and B. E. Layton. *Biophysical Journal*, 95, 3606 (2008).
- [61] Nogales, E. *Cell. Mol. Life. Sci.* 56, 133- 142 (1999).
- [62] F. A. Chandra, *et al.* *Science* 333,187(2011).
- [63] P. Grindrod. *Patterns and Waves: The Theory and Applications of Reaction-Diffusion Equations*, Clarendon Press (1991).
- [64] Morozova, N. *et al.* *Rna-a Publication of the Rna Society* **18**, 1635-1655 (2012).
- [65] Eulalio, A. *et al.* *Rna-a Publication of the Rna Society* **15**, 21-32 (2009).
- [66] Bazzini, A. A., Lee, M. T. & Giraldez, A. J. *Science* **336**, 233-237 (2012).
- [67] Lim, L. P. *et al.* *Genes & Development* **17**, 991-1008 (2003).

Publications

Publications

Publications including works in submission:

- [1] Kathiresan Natarajan, Zhu Kaicheng, and Haibin Su. *Physical Principles involved in the Robustness of Tubulin Dynamics*, 2017.
- [2] Zhu Kaicheng, Haibin Su, and Federico Lauro. *Metabolism-driven Bacterial Stratification in the Ace Lake*, 2017.
- [3] Le Quang Luan, Zhu Kaicheng, and Haibin Su. *Sequential Pathway with Transient States of Ribosome during Protein Synthesis*, 2017.