Modelling Genetic Regulatory Networks: 
A New Model for Circadian Rhythms in *Drosophila* and 
Investigation of Genetic Noise in a Viral Infection Process

by

Z. Xie

A thesis submitted in partial fulfilment
of the requirements for the Degree of
Doctor of Philosophy
in Computational Systems Biology
at
Lincoln University
New Zealand

2007
Abstract of a thesis submitted in partial fulfilment of the requirements for the Degree of Doctor of Philosophy

**Modelling Genetic Regulatory Networks:**

A New Model for Circadian Rhythms in *Drosophila* and Investigation of Genetic Noise in a Viral Infection Process

By Z. Xie

In spite of remarkable progress in molecular biology, our understanding of the dynamics and functions of intra- and inter-cellular biological networks has been hampered by their complexity. Kinetics modelling, an important type of mathematical modelling, provides a rigorous and reliable way to reveal the complexity of biological networks. In this thesis, two genetic regulatory networks have been investigated via kinetic models.

In the first part of the study, a model is developed to represent the transcriptional regulatory network essential for the circadian rhythms in *Drosophila*. The model incorporates the transcriptional feedback loops revealed so far in the network of the circadian clock (PER/TIM and VRI/PDP1 loops). Conventional Hill functions are not used to describe the regulation of genes, instead the explicit reactions of binding and unbinding processes of transcription factors to promoters are modelled. The model is described by a set of ordinary differential equations and the parameters are estimated from the *in vitro* experimental data of the clocks’ components. The simulation results show that the model reproduces sustained circadian oscillations in mRNA and protein concentrations that are in agreement with experimental observations. It also simulates the entrainment by light-dark cycles, the disappearance of the rhythmicity in constant light and the shape of phase response curves resembling that of experimental results. The model is robust over a wide range of parameter variations. In addition, the simulated E-box mutation, *per* and *per* mutants are similar to that observed in the experiments. The deficiency between the simulated mRNA levels and experimental
observations in $per^{01}$, $tim^{01}$ and $clk^{brk}$ mutants suggests some differences in the model from reality. Finally, a possible function of VRI/PDP1 loops is proposed to increase the robustness of the clock.

In the second part of the study, the sources of intrinsic noise and the influence of extrinsic noise are investigated on an intracellular viral infection system. The contribution of the intrinsic noise from each reaction is measured by means of a special form of stochastic differential equation, the chemical Langevin equation. The intrinsic noise of the system is the linear sum of the noise in each of the reactions. The intrinsic noise arises mainly from the degradation of mRNA and the transcription processes. Then, the effects of extrinsic noise are studied by means of a general form of stochastic differential equation. It is found that the noise of the viral components grows logarithmically with increasing noise intensities. The system is most susceptible to noise in the virus assembly process. A high level of noise in this process can even inhibit the replication of the viruses.

In summary, the success of this thesis demonstrates the usefulness of models for interpreting experimental data, developing hypotheses, as well as for understanding the design principles of genetic regulatory networks.

Keywords: systems biology, genetic regulatory networks, mathematical molecular modelling, kinetic modelling, Hill function, stochastic modelling, stochastic differential equations, chemical Langevin equation, oscillations, circadian clock, circadian rhythms, Drosophila, intrinsic noise, extrinsic noise, viral infection, virus replication
My deepest gratitude belongs to my parents,

I dedicate this thesis to you.

谨以本文献给我最爱的父母，
感谢他们对我坚持不懈地教育和支持.
Acknowledgements

First and foremost, I would like to express my gratitude to Prof. Don Kulasiri, as a supervisor, mentor and friend. Beside the guidance, support and encouragement I needed to stay on track during the Ph.D. study, he also shared his philosophy of research, which has strongly influenced my own. I would also like to thank my associate-supervisors Dr. Sandhya Samarasinghe and Dr. Wynand Verwoerd for their suggestions and discussions about my research.

Prof. Kulasiri’s C-fACS (Centre for Advanced Computational Solutions) team is full of wonderful people who helped me in different ways. I would particularly like to thank Dr. Channa Rajanayaka for his mentoring and friendship as I began my graduate study, to Dr. Mu Lin and Khuyen Lan Nguyen for the continued sharing of ideas and discussions, and for being great table tennis and badminton opponents, and to Sean Richards, whom I have shared an office with for most of my postgraduate career. I have enjoyed enormously working with all these people.

Other extremely helpful and friendly people are Caitriona Cameron, whose workshop has improved my English writing skills significantly, and Flo Weingartner, who has always fixed my computer and network problems very quickly with endless patience.

I would also like to express my appreciation to Lincoln University for providing the Lincoln University Doctoral Scholarship during my three years of Ph.D. study.

Finally, I would never have started this work without years of encouragement and support from my parents, to whom I dedicate this thesis. And special thanks to my girlfriend, Jessica, for her love that has brought joy to every day of my life.
Figures

Figure 2-1 The 1970 version of the central dogma. Solid arrows show the general information flows, while dotted arrows represent the special information flows. ................................................................. 13

Figure 2-2 The fractional saturation of reactions without cooperative binding (h=1) and with positive cooperative binding (h=2). ............................................................................................................ 24

Figure 3-1 Schematic representation of the circadian systems, adapted from Eskin (1979). ................... 38

Figure 3-2 Interactions in the two loop model of Cyran et al. (2003). ....................................................... 43

Figure 3-3 The model proposed by Smolen et al. (2004). (A) The model accounts for three feedback loops. In the per loop, PER interacts with CLK forming a negative feedback loop. In the vri loop, vri is activated by CLK, and VRI in turn represses clk. In the pdp1 loop, pdp1 is activated by CLK, and PDP1 in turn activates clk. (B) PER undergoes two cytosolic phosphorylations and then enters the nucleus where PER interacts with CLK, suppressing CLK’s activation of per. Nuclear PER undergoes further phosphorylations before degradation. ..................................................................................... 47

Figure 3-4 The model proposed by Ruoff et al. (2005), where dCLK denotes Drosophila CLK, the subscript letters “c” and “n” denote cytoplasm and nucleus, respectively, and per/tim, as well as PER/TIM, are treated as one component in the system. In the core of this model, the transcription factor CLK is subjected to positive and negative regulation by the proteins PDP1 and VRI, whose transcription is activated by CLK. CLK also activates the clock genes per and tim and the PER/TIM complex binds to CLK and, thus, reduces the activity of CLK. ........................................................................................................ 48

Figure 4-1 The schematic diagram of the model. The model shows the regulatory relationships among genes, mRNAs and proteins in the negative and positive transcriptional feedback loops. Transcription of per, tim, vri and pdp1 genes are activated by CLK/CYC dimers binding to E-boxes in their promoter regions. In one loop, per and tim mRNAs are translated to PER and TIM proteins which form PER/TIM dimers. PER/TIM binds to CLK/CYC to form PER/TIM/CLK/CYC complex. In another loop, vri and pdp1 mRNAs are translated to VRI and PDP1 proteins. They compete to bind the V/P box in the promoter in clk gene. Transcription of clk gene is repressed by VRI and activated by PDP1. clk mRNA is translated to CLK which forms CLK/CYC dimers with CYC. Proteins, mRNAs, dimers and complexes are degraded at certain kinetic rates. CYC is assumed to be constant, therefore, there is no degradation process for CYC. Variable names used in the model are indicated in parentheses. The number of the E-boxes and the V/P box in the promoters is also shown here. ........................................................................................................ 50

Figure 5-1 An overview of the parameter estimation process .................................................................... 68

Figure 6-1 Sustained oscillations for the concentrations of the mRNAs and the proteins: (A) Oscillations for the mRNAs and (B) oscillations for the proteins. The time scale of clk in (A) and PDP1 in (B) has been enlarged to allow better visualisation. ................................................................. 74

Figure 6-2 Period variations of the circadian oscillations in respect to parameter variations, one parameter was increased or decreased by 20% each time while the other parameters were kept at the basal values. The most sensitive parameters are indicated. Parameter names corresponding to the parameter index are denoted in Table 5-1. The number of the E-boxes or V/P box is related to the structure of the model, and therefore these there parameters are not included in this figure. ........................... 76

Figure 6-3 (A). Entrainment by light dark cycles. klight is increased (0.8) during the light phase and remains at the original value (0.62) during the dark phase. Simulation was done with ZT0 lights on, ZT12 lights off. (B). The phases of oscillations after entrainment depended on the different values of klight. We plotted the 6th cycle after the cycles were stable to eliminate the transient effect of light. ......................... 79

Figure 6-4 (A) In vitro experiments showed that rhythmicity of per mRNA disappeared under the condition of constant light for three days, replotted from Qiu and Hardin (1996). (B) Computational simulations showed that rhythmicity disappeared under constant light condition when klight > 5. A klight value of 5 was used to produce this figure. ................................. 80
Figure 6-5 Phase response curve (PRC) obtained by using $k_{\text{light}} = 1.3$. The x-axis represents the time of onset of each light pulse, and on the y-axis positive values represent phase advances and the negative values represent phase delays. The means of experimental values for phase shifts from Konopka (1991) are denoted by diamonds. The simulated PRC is shown by the dashed curve and the shifted simulated PRC is shown by the solid curve. The shifted simulated PRC was obtained by advancing the simulated PRC by 5 h, and it is plotted here only for comparison purposes. ............................................................. 82

Figure 6-6 mRNA oscillations in E-box mutation simulations, where only one copy of E-box remains in each type of gene.............................................................................................................. 84

Figure 6-7 Simulation of arrhythmic mutants. Parameter values are as in Table 5-1, except for $t_{\text{per}}=0$ for per$^{01}$, $t_{\text{tim}}=0$ for tim$^{01}$ and $t_{\text{clk}}=0$ for clk$^{0k}$................................................................................................................................. 86

Figure 6-8 Simulation of short and long mutants. Parameter values are as in Table 5-1, except for $d_{\text{pt}}=0.9$ for per$^{S}$ and $d_{\text{pt}}=0.08$ for per$^{L}$.................................................................................................................. 88

Figure 6-9 Time evolutions of proteins. The feedback loop was removed by making its corresponding gene expression constant. .................................................................................................................. 90

Figure 6-10 Parametric sensitivity results. The parameter indexes are as indicated in Table 5-1................. 91

Figure 7-1 The scheme of the viral replication cycle. .................................................................................................................. 103

Figure 7-2 Time evolution of the components of the system described by ODEs. Simulations are for low MOI. (A) plots of genome (G) and protein (P). (B) plots of mRNA (R) and virus (V)................................................. 106

Figure 7-3 Time evolution of mRNA in the stochastic simulations solved by the Gillespie algorithm. The rate constants used are the same as their deterministic counterparts. Low MOI is used as the initial condition. (A) A sample realisation of successful infection. (B) A sample realisation of abortive infection. ................................. 109

Figure 7-4 Stochastic simulation results of 1000 realisations. (A) The mRNA frequency distribution at 200 days post-infection. The x-axis indicates the percentage of mRNA present in the cells. The y-axis indicates the number of mRNAs at 200 days. (B) The average time evolution of 1000 stochastic realisations (dashed line) and the average time evolution of filtered cells (dotted line). The filtered cells are the cells which have only successful infections. For comparison, the time evolution from the deterministic solution is also plotted (solid line). ................................................ 110

Figure 8-1 Stochastic simulations of mRNA by the Gillespie algorithm and CLE for the low MOI (the first row) and the high MOI (the second row). 1000 realisations were obtained for each of the figure. (A, C) Average behaviour of the realisations during 200 days post-infection. (B, D) Frequency distribution at 200 days post-infection. .......................................................... 118

Figure 8-2 Frequency distribution of mRNA at 200 days post-infection. Each figure accounts for the intrinsic variable of the each reaction................................................................. 122

Figure 8-3 Time evolution of mRNA with different parameters perturbed under different noise levels. Black solid line denotes the average number of mRNA, and grey dashed line denotes the sample realisations. Five sample realisations are plotted. NC denotes the coefficient of noise ($c_i$). ....................... 127

Figure 8-4 The mean value and the extrinsic noise of mRNA at time 200 days over 1000 realisations against the coefficient of noise ($c_i$). The x-axis and y-axis in both figures are logarithmically scaled. Note in the figure A, the plots of reaction 2, 3, 5 and 6 are virtually identical................................. 128
Tables

Table 4-1 Biochemical meaning of the parameters ................................................................. 62

Table 5-1 Parameters of the model: The units of binding rates and association rates are nM⁻¹h⁻¹ and the units of the other parameters are h⁻¹. ......................................................................................................... 69

Table 5-2 Initial conditions. Abbreviations: CC – CLK/CYC, PT – PER/TIM, CCPT – CLK/CYC/PER/TIM. Constant values in the system are denoted by *. ................................................................. 71

Table 7-1 Reaction steps and descriptions. The units of $k_4$ are molecules⁻¹day⁻¹ and the units of other parameters are day⁻¹. .............................................................................................................................. 104

Table 8-1 Mean and intrinsic noise of all the viral components at 200 days post-infection. The values were obtained by the stochastic simulations via the Gillespie and CLE, respectively. .............................. 119

Table 8-2 Mean and intrinsic noise of mRNA at 200 days post-infection as well as the contributions of noise from each reaction. Reaction i represents that $a_i$ was set to be 1 in order to account for the intrinsic noise contributed from that reaction and the intrinsic noises from the other reactions were silenced..... 121
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic helix-loop-helix (bHLH)</td>
<td>PAR Domain Protein 1 (pdp1)</td>
</tr>
<tr>
<td>Casein Kinase 2 (CK2)</td>
<td>PER ARNT SIM (PAS)</td>
</tr>
<tr>
<td>Chemical Langevin equation (CLE)</td>
<td>PERIOD (PER)</td>
</tr>
<tr>
<td>Chemical master equation (CME)</td>
<td>(\text{period (per)})</td>
</tr>
<tr>
<td>CLOCK (CLK)</td>
<td>Phase response curve (PRC)</td>
</tr>
<tr>
<td>(\text{clock (clk)})</td>
<td>Ribonucleic acid (RNA)</td>
</tr>
<tr>
<td>Constant light (LL)</td>
<td>RNA polymerase (RNAP)</td>
</tr>
<tr>
<td>CYCLE (CYC)</td>
<td>Shaggy (SGG)</td>
</tr>
<tr>
<td>(\text{cycle (cyc)})</td>
<td>Slimb (SLMB)</td>
</tr>
<tr>
<td>Deoxyribonucleic acid (DNA)</td>
<td>Stochastic differential equation (SDE)</td>
</tr>
<tr>
<td>DOUBLETIME (DBT)</td>
<td>Systems Biology Markup Language (SBML)</td>
</tr>
<tr>
<td>Extensible Markup Language (XML)</td>
<td>TIMELESS (TIM)</td>
</tr>
<tr>
<td>(\text{Gene})nc regulatory network (GRN)</td>
<td>(\text{timeless (tim)})</td>
</tr>
<tr>
<td>Levenberg-Marquardt (LM)</td>
<td>Transcription factor (TF)</td>
</tr>
<tr>
<td>Linear noise approximation (LNA)</td>
<td>VIRLLE (VRI)</td>
</tr>
<tr>
<td>Messenger RNA (mRNA)</td>
<td>(\text{virlle (vri)})</td>
</tr>
<tr>
<td>Multiplicity of infection (MOI)</td>
<td>Weighted sum of squares (WSS)</td>
</tr>
<tr>
<td>Ordinary differential equation (ODE)</td>
<td>Wild-type (WT)</td>
</tr>
<tr>
<td>PAR Domain Protein 1 (PDP1)</td>
<td>(\text{Zeitgeber time (ZT)})</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

1.1 The challenge of systems biology

During the last fifty years, molecular biology has made remarkable progress in our understanding of biological systems at a molecular level. The components researched in molecular biology include DNA, the long linear molecules storing genetic information, RNA, a close relative of DNA, whose functions range from serving as a temporary working copy of DNA to structural and enzymatic functions, and proteins, the major structural and enzymatic type of molecules in cells. Traditionally, experimental techniques in molecular biology have mainly focused on identification of single components of a system. These kind of experimental techniques are often called “reductionist” in the sense of their ability to break down a system into parts and study one part of a process at a time. Although reductionist biology is useful to give basic information about components that make up cells and their individual chemical properties, it does not provide us with an understanding of cells as systems. The next major challenge is to combine the accumulated data from various sources to understand biological systems.

Since the first genome sequence of *Haemophilus influenzae* was published in 1995 (Fleischmann, Adams et al. 1995), many genome sequences have been completed, of which the sequencing of the human genome is the most important (Venter, Adams et al. 2001). Deciphering the genome sequences of many organisms is an important step towards understanding cells at the system level. However, knowing the information encoded in these sequences does not necessarily mean knowing how a living cell works. In order to arrive at biological properties and behaviours that arise from a list of components, we need to know not only the information about the genome, but also information about mRNA expression, interactions of DNA with protein, interactions of protein with protein, and other molecule interactions.
Several high-throughput experimental technologies have been developed recently that allow us to assess genome-wide expression. These methods include cDNA microarrays, which can be used to obtain thousands of temporal gene expression patterns for different cell types in response to specific stimuli simultaneously (Baldi and Hatfield 2002); proteome chips, which can be employed for global analysis of protein activities (Zhu, Bilgin et al. 2001); and two-hybrid screens, which enable the construction of protein interaction maps (Uetz, Giot et al. 2000). The development of these technologies gives us a golden opportunity to view a cell as a system, rather than focusing on its individual cellular components. This has opened up a new field in biology that aims to understand molecular biology as systems, called systems biology (Kitano 2002).

In fact, the study of the system-level understanding of biology has a long history. It started as early as the 1940s with the introduction of cybernetics, which aimed at describing animals and machines using control and communication theory (Wiener 1948). This was the first attempt to establish the idea of interactions between systems theory and biological sciences. Since then, several similar attempts have been made to describe and analyse biological systems at the physiological-level. The unique attributes of systems biology distinguishes itself from the previous attempts in that it connects system-level descriptions to molecular-level knowledge. Three major issues within systems biology are (1) to generate quantitative high-throughput data by experimentation, (2) to integrate various kinds of data by data processing, and (3) to build mathematical models based on the data. This thesis focuses on the third issue, and will investigate the components of cellular networks and their interactions by means of mathematical models.

The role of mathematical models in systems biology is multi-faceted. Firstly, mathematical models enable validation of current knowledge by comparing model predictions with experimental data. When discrepancies are found in these types of comparisons our knowledge of the underlying networks can be systematically expanded (Covert and Palsson 2002). Secondly, mathematical models can suggest novel experiments for testing hypotheses that are formulated from modelling experiences (Yuh, Bolouri et al. 2001). Thirdly, they enable the study and analysis of system properties that are not accessible through in vitro experiments (Pritchard and Kell 2002).
And, finally, mathematical models can also be used for designing desirable products based on existing biological networks (Arkin 2001).

There are three major classes of cellular networks where significant modelling efforts are underway: metabolic pathways, signalling pathways and genetic regulatory networks (GRNs). A metabolic pathway is a series of chemical reactions occurring within a cell, resulting in either the formation of metabolic products or the initiation of another metabolic pathway. The dominant phenomenon in metabolism is enzymatic reactions. Scientists have characterised metabolism better than any other part of cellular behaviour due to more developed experimental techniques being available to quantify the network components. The typical mathematical modelling schemes are deterministic methods because, usually, a large number of molecules are involved in metabolic pathways. A signalling pathway is any process by which a cell converts one kind of signal or stimulus into another, where the dominant phenomenon is molecular binding. Signalling pathways normally have much fewer reactant molecules than metabolic systems, therefore, more efforts for modelling signal pathways have focused on stochastic methods. A GRN consists of a set of genes, proteins, metabolites (the intermediates and products of metabolism), and their mutual regulatory interactions. The dominant phenomenon in GRNs is molecular binding, polymerization and degradation. Like signalling pathways, they tend to contain a small number of molecular entities. Typical modelling schemes are deterministic and/or stochastic depending on the purpose of the modelling. Ultimately, all three pathways have to be integrated into a large network to generate whole-cell models. Due to the central role that genetic networks play in cellular function, mathematical modelling in GRNs is introduced next in detail, as it is the focus of this thesis.

1.2 Mathematical modelling in GRNs

Proteins are essential for the development and function of an organism. The inherited information embedded in DNA sequences has an ability to direct production of proteins and this process is called gene expression. Gene expression is highly regulated in cells. Only a fraction of genes in a genome are expressed under a given condition or in a particular cell type. There are complex networks that control where, when and which
genes are expressed in response to various environmental and developmental signals. Many interesting questions can be raised from gene regulation; for example, which gene is expressed in a certain cell at a certain time and how does gene expression differ with different stimuli? What makes a genetic network robust? Are there certain GRN architectures that are more likely to be compatible with life than others? To answer these questions, a deep understanding of mechanisms underlying GRNs is needed. The interactions of components in GRNs are, primarily, based on DNA-protein and protein-protein interactions, therefore, the networks of gene regulation can be very complex, where genes activate or repress one another’s activity, either directly or through their products, to form feedback loops.

Currently, two theoretical approaches are used to analyse GRNs, reverse engineering and forward modelling (Kauffman 2004). Reverse engineering is used to analyse data that are not \textit{a priori} known to contain any specific pathways (Armstrong and van de Wiel 2004). It analyses expression changes of thousands of genes in parallel over time and attempts to determine regulatory interactions based on the gene expression profiles (expression values of different genes under different experimental conditions). By searching for clusters and motifs, and eventually deducing functional correlations, reverse engineering methods seek to reconstruct underlying regulatory networks. The advantage of reverse engineering is that the gene expression data themselves are used to identify meaningful or informative gene dynamical behaviours, and normally a large fraction, or almost all genes, of a cell can be covered. However, the difficulty associated with this approach is that the data derived from the current experimental tools, such as gene expression arrays or proteomic arrays, are normally too noisy to provide insights into the underlying relations between the genes.

Forward modelling is also known as \textit{in silico} cell”, which tries to isolate some genetic pathways and build a detailed model that can be compared directly with experimental data (Bower and Bolouri 2001; Endy and Brent 2001). The basis of forward modelling is \textit{a priori} knowledge or hypotheses about the processes of the interactions taking place during gene expression. It starts with building a conceptual model where elements and their interaction are extracted from literature. The conceptual model is then converted into an appropriate computational model. Once the parameters are set, the model can produce the dynamics of the regulatory network. The advantage of forward modelling is
that the models can be compared with experimental reality directly and testable hypotheses for further experiments can be obtained. The drawback of this approach is that its scope focuses only on local dynamics but the target pathways are frequently influenced by genes from other pathways. Moreover, it often lacks specific kinetic parameters for the individual processes under consideration. Mathematical techniques used in forward modelling of GRNs will be introduced below, as they will be used to investigate GRNs in this thesis.

**1.2.1 Mathematical techniques for forward modelling of GRNs**

In forward modelling of GRNs, a GRN can be viewed as a cellular input-output device containing three components: (1) Inputs: proteins, such as transcription factors (TFs); (2) Nodes: genes are the nodes in the network. The nodes can also be viewed as a function that can be obtained by combining basic functions of inputs; and (3) Outputs: RNAs and proteins. The focus of forward modelling of a GRN is to determine input-output functions in order to summarise the current knowledge and hypothesise the behaviour of the GRN.

Before establishing a realistic and reliable input-output function, we have to ask ourselves some questions before choosing an appropriate abstraction – at what level does such detail become relevant, and at what level can one ignore it? The answer to this question is not obvious. Various modelling approaches have been used to describe GRNs including direct or undirected graphs, Boolean networks, Bayesian networks, continuous models based on ordinary differential equations (ODEs), partial differential equations and stochastic models. Each approach has certain advantages and disadvantages. The answer to the selection of an approach depends to a large extent on the purpose of the modelling exercise.

A comprehensive literature review of these techniques from a mathematical aspect is given by De Jong (2002). There are other reviews discussing various aspects of modelling in the literature. Smolen et al. (2000) concentrated on the Boolean networks and ODE models of prokaryotes. Bolouri and Davidson (2002) focused on the role of
modelling in understanding GRNs of eukaryotes. Schlitt and Brazma (2005) reviewed the modelling techniques in GRNs at different levels, from a genome-wide scale to dynamic models for a particular pathway. Longabaugh et al. (2005) reviewed developmental GRNs specifically; these are typically large-scale and multi-layered. Alves et al. (2006) gave an overview of the tools available for creating and exploring genetic networks.

1.3 Motivation of the study in the thesis

This thesis involves the study of two genetic networks, the circadian clock system in *Drosophila* (fruit fly) and an intracellular viral infection system. A detailed explanation of the reason for choosing these two particular systems is provided as follows.

1.3.1 Circadian clock system in *Drosophila*

Life on the earth is exposed to many different environmental influences and many of them follow a daily periodic change. The two most important changes are the daily changes of light and temperature. Consequently, many physiological processes in living beings follow a daily periodicity. In fact, all eukaryotes and some prokaryotes are capable of maintaining sustained oscillations in terms of gene activity, metabolism, physiology and behaviour with a period close to 24 h (Pittendrigh 1960; Panda, Hogenesch et al. 2002; van Gelder, Herzog et al. 2003; Nitabach 2005). These oscillations are known as circadian rhythms, where “circadian” comes from the Latin words, “circa” meaning about and “dies” meaning a day.

Circadian rhythms exist in nearly all species and affect all aspects of daily life. In recent decades, many components and molecular mechanisms comprising circadian clocks, the mechanisms in cells controlling circadian rhythms, have been uncovered. Among all the organisms used to study circadian clocks, *Drosophila* is the one most extensively researched because of its status as a central model organism in eukaryote biology. *Drosophila* is, therefore, emerging as one of the key model organisms for systems
biology where the aim is, eventually, to be able to build predictive models of all major cellular processes in a cell.

In this thesis, the *Drosophila* circadian clock is chosen as a modelling target because its molecular studies offer sufficient details to allow the assembly of a detailed mathematical model. The wealth of experimental information available makes modelling a feasible task. Even more importantly, the numerous mutant data enable the model to be reliably validated. The conventional method, ODEs, is proposed to model this biochemical system. The advantage of the description with ODEs is that we can take into account detailed knowledge about gene regulatory mechanisms such as individual kinetics, individual interactions of DNAs and proteins when reconstructing the model. The resulting ODEs can be solved by numerical integration; this is extremely useful in characterising the general features of pathway behaviour. Furthermore, various numerical tools, such as parameter fitness and sensitivity analysis, can be readily employed to explore the important properties of the system.

### 1.3.2 Intracellular viral infection system

Viruses infect major groups of organisms: vertebrates, invertebrates, plants, fungi, bacteria and human beings. Viruses, which mean “poison” in Latin, have caused some of the deadliest diseases in humans. For example, smallpox epidemics in the Middle Ages resulted in significant population losses, and the “Spanish flu” pandemic caused over 20 millions lives to be lost in 1918-1919. Now more than three million people die every year from AIDS-related illnesses (Quinn and Overbaugh 2005). Very recently, unexpected outbreaks of the Severe Acute Respiratory Syndrome (SARS) virus may become a pandemic threat (Li, Guan et al. 2004). The significant impact of viral infection has motivated numerous research efforts addressing diverse aspects of viruses (Evans and Kaslow 1997). Based on a wealth experimental data, viruses are important organisms for system biology because their relatively simple structures make the quantitative measurements of viral abundance and parameters possible.
Because of its significance in both molecular biology and systems biology, a viral infection system is also chosen as a modelling target in this thesis. The development of a mathematical viral infection model is hoped to provide valuable information about the basic mechanisms of molecular genetics and important frameworks for more efficient drug development and therapeutic intervention. The research interest is to observe the roles of noise in the viral replication processes. As will be discussed in Chapter 2, regulatory processes are stochastic processes in cellular systems that are subject to biochemical noise under some circumstances. This system is particularly suitable for modelling and gaining insights into the effects of noise because it has a relatively simple structure where only a few biochemical reactions and parameters are involved, while still being of considerable biological relevance.

However, it should be emphasised that although stochastic simulations are closer than their deterministic counterparts in approximating the underlying reality, this is not necessarily an advantage. The stochastic approach is computationally expensive and it is much more difficult to analyse the dynamics of biological systems than the deterministic one. The choice of mathematical approach for a particular system depends greatly on the experimental data available and research questions expected from the model. Ideally, the stochastic properties for the circadian clock system would be explored. However, that system containing a large number of biochemical reactions and parameters makes the analysis of noise difficult. For this reason, a simpler system, the viral infection model is used to illustrate the importance of noise in this thesis.

1.4 Objectives

Throughout this work, the major theme is to integrate our knowledge in mathematics and biology to construct in silico models which are then used to interpret experimental data and develop hypotheses. The specific objectives can be summarised as follows:

The circadian clock system

- To build a conceptual model for a mechanistically well-understood system.
• To apply system identification, parameter estimation and quantitative modelling approaches to develop a mathematical model from the conceptual model.

• To reconstruct a number of in silico phenotypes, such as the responses of the system to environmental changes and mutations.

• To investigate design principles of the circadian clock in Drosophila.

• To identify the gaps of our current knowledge of the system.

The viral infection system

• To understand fluctuations of gene expression in a viral infection model.

• To reveal the sources of intrinsic noise.

• To investigate the effects of extrinsic noise

• To develop hypotheses about how the cells will behave under defined noisy conditions.

• To better understand the complex mechanisms underlying the interaction of viruses with their host cells through simulations.

1.5 Overview of chapters

In the current chapter, an introduction is provided to systems biology, mathematical modelling in GRNs and their practical implications, which leads to the motivation for the thesis. In Chapter two, background information covering the fields touched in this thesis is given. In Chapter three, molecular and mathematical bases of the circadian clock are given. In Chapter four, a conceptual model from the molecular basis is developed, then the conceptual model is converted into a mathematical model. In Chapter five, the mathematical model is implemented into a computer solvable model, where the parameters are also estimated. In Chapter six, the in silico experiments are obtained from the model and compared with in vivo experiments. In Chapter seven, the biology of the viral infection system is given. In Chapter eight, two stochastic models are developed based on a deterministic model and the effects of intrinsic noise and
extrinsic noise are investigated using the two models. Finally, in Chapter nine, a retrospective look at the overall implications of this work is provided, as well as the contributions of the thesis and directions for future research.
Chapter 2: Background

Since the thesis embraces several disciplines, the background information discussed here covers the fields of molecular biology, biochemistry, mathematics and computing issues. Section 1 gives the relevant biological background of GRNs. Section 2 is a discussion of multi-scale issues in modelling biochemical systems. Section 3 provides a review of kinetic modelling of biochemical systems, particularly the Hill function. Sections 4 and 5 present stochastic modelling of intrinsic and extrinsic noise, respectively.

2.1 Biological background of GRNs

This section is written for readers who are unfamiliar with molecular biology. It aims to present some of the biological basis required for the rationale of the models formulated in the thesis. For a more comprehensive and detailed introduction, refer to two classic molecular biology books (Lodish 2003; Lewin 2006).

2.1.1 Cells and their molecular components

Cells are the fundamental working units of every living organism. Based on different cell structures, all living cells can be classified as either prokaryote or eukaryote. The structure of the prokaryotes is simpler than that of the eukaryotes. Organisms in the prokaryotic class constitute bacteria and cyanobacteria, characterised by the absence of nuclear membranes. Organisms in the eukaryotic class encompass some protists and all plants, animals and fungi, characterised by the presence of a nucleus and other membrane-enclosed structures. Knowing the difference between prokaryotes and eukaryotes is important for understanding genetic and biochemical mechanisms in cells, and this knowledge is essential for deciding what details to include when building a mathematical model for a genetic network.
The dynamic environment within a cell involves a highly complex interaction between three important classes of macromolecules: deoxyribonucleic acid (DNA), ribonucleic acid (RNA) and protein. DNA is a storage repository containing information of all the genes for protein synthesis and self replication. RNA acts as a bridge between DNA and proteins, and uses their genetic information to help cells produce proteins. Protein is the fundamental structural and functional unit in cells. Each protein is specialised to carry one of a variety of important roles, such as structural elements, enzymatic catalysts, or antibody and regulatory functions. In particular, there is a special class of proteins called transcription factors (TFs) that play an important regulatory role in the networks of genes, mRNAs and proteins. The details of how TFs function will be discussed in a later section.

2.1.2 Gene expression

DNA in a cell contains the complete genetic information that defines the structure and function of an organism, where each gene corresponds to the genetic information of one protein. The conversion from DNA to protein is determined by what, when and to what extent genes are expressed in DNA, resulting in the production of their respective proteins. This process is commonly known as gene expression. Gene expression decides the cellular development. For example, in a human, apart from reproductive cells (gametes) and mature red blood cells, every cell shares exactly the same DNA, but nerve cells and white blood cells have completely different shapes and functions resulting from the different gene expression.

The central dogma of molecular biology states the flow of gene expression. An early version of the central dogma states that DNA is first transcribed, or copied, into a short-lived messenger RNA (mRNA), and mRNA is then translated repeatedly into a protein (Crick 1958). Later, as the more pathways were revealed by researchers, the revised 1970 version, as shown in Figure 2-1, states that beside the standard pathway of information flow from "DNA to mRNA to protein", the pathway of replication of DNA also certainly exists. Furthermore, the pathways of reverse transcription from RNA to DNA, replication of RNA and direct translation from DNA to protein, rarely but
Figure 2-1 The 1970 version of the central dogma. Solid arrows show the general information flows, while dotted arrows represent the special information flows.
possibly, exist (Crick 1970). The flows from protein to protein and protein to DNA or RNA have not been discovered and are regarded as impossible.

### 2.1.3 Regulation of transcription

Gene expression is a tightly regulated process which gives the cell control over its structure and function. Any step of gene expression may be modulated, from transcription, RNA degradation, translation, post-translational modification of protein and protein degradation. Here, more detail about the control of transcription is given because this process is the predominant site for control of gene expression (von Hippel 2004).

Every gene consists of a coding region and a regulatory region. The coding region is the part that is transcribed into an mRNA, and the regulatory region is the part that contributes to the control of the gene. The transcription process begins when a RNA polymerase (RNAP), a catalytic protein, binds to the DNA upstream of the coding region, which is a part of the regulatory region called the promoter region. The RNAP separates the double-stranded DNA and then moves along a single strand, step by step, and transcribes the coding region into mRNA. As the mRNA is constructed, the RNAP peels away and the DNA strands are rejoined. The transcription process stops when the RNAP reaches a termination site at the end of the gene.

Typically, RNAPs do not bind to the regulatory region of DNA alone, but in complexes with TFs. In simple prokaryotes, the regulatory region is typically short (10-100 bases) and contains binding sites for a small number of TFs. In eukaryotes, the regulatory regions can be very long (up to 10,000 or 100,000 bases) and contain binding sites for multiple TFs. Sometimes TFs are called trans-regulatory elements, and regulatory sites where TFs bind are called cis-regulatory elements (Stamatoyannopoulos 2004).

The function of TFs is to control the rate of transcription. When TFs associate with the promoter regions of their target genes and affect the affinity of RNAP for the transcription initiation site of the gene, they can function to induce or repress synthesis
of the corresponding mRNA, and are called activators or repressors, respectively. Activators enhance the interaction between the RNAP and the promoter, therefore, increasing the gene expression rate. Repressors impede the RNAP’s progress along the DNA strand, therefore, hampering the expression rate of the gene. Accordingly, the binding site is called an enhancer if bound with an activator, and a silencer if bound with a repressor (Bower and Bolouri 2001).

### 2.1.4 Cooperativity

In biochemistry or molecular biology, cooperativity is a phenomenon displayed by enzymes or receptors that have multiple binding sites. In a genetic network there are always multiple binding sites in a promoter region where several TFs are able to bind. Therefore, transcriptional regulation tends to involve combinatorial interactions between several TFs, which allow for a sophisticated response to multiple conditions in the environment. Non-cooperativity occurs when TFs are independently bound to a promoter. Cooperative binding occurs when the affinity of the TF to a promoter depends on the amount of TFs already bound. The cooperative binding can be either positive or negative, indicating that the affinity is either increased or decreased by the binding of other TFs. Competition is also possible when two different TFs bind to one site.

### 2.1.5 GRNs

Because TFs are themselves the products of expressed genes, they too are under regulatory control, giving rise to complex networks of regulatory networks, commonly known as genetic regulatory networks. A GRN is a collection of DNA segments, proteins and other metabolites in a cell which interact with each other and form feedback loops in the cell. For a single-feedback system, there are two major kinds of feedback – positive and negative. For a multiple-feedback system, there are combinations of positive and negative feedbacks. Both single-feedback and multiple-feedback networks will be modelled in the thesis.
In a negative feedback, specifically, a TF inhibits the transcription of its own gene by blocking RNAP binding at the promoter region. Negative feedback loops are generally considered to provide stability. They are required for stable oscillations and some examples are the circadian rhythm (Goldbeter 1995), cell cycles (Novak, Pataki et al. 2001) and calcium waves (Bootman, Lipp et al. 2001).

In a positive feedback a TF promotes the transcription of its own gene by enhancing RNAP binding at the promoter region. Positive feedback loops in a resource-limited environment normally lead to a tendency to reinforce the growth of a species until it reaches a value that cannot be sustained. It is required for a permanent shift in behaviour, such as differentiation or evolution towards one of two states of a system (Becskei, Seraphin et al. 2001).

With multiple-feedback networks, biological systems display more complicated behaviours. For example, chaotic systems, deterministic but essentially unpredictable systems, frequently result from some form of positive feedbacks, usually mixed with negative feedbacks (Smolen, Baxter et al. 2000). The Elowitz and Leibler oscillator (Elowitz and Leibler 2000) is based solely on negative feedback loops but is unstable. This system could be made comparatively stable and robust by incorporating positive feedback loops (Pomerening, Sontag et al. 2003; Angeli, Ferrell et al. 2004). In this thesis, our first model, the circadian clock model, contains a mixture of positive and negative feedback loops. It will be shown that a negative feedback loop is responsible for producing the oscillation behaviour. However, the additional positive and negative feedback loops both increase the robustness of the system.

2.1.6 Noise in GRNs

It has long been recognised that genetically identical cells under the same environmental conditions can have significant variations in phenotypic characteristics (Delbruck 1945). Such variation has been observed in the cells of organisms ranging in complexity from bacteria to mammals, and is believed to be an important factor in the development and function of many living organisms, physiologically and evolutionarily.
It is, therefore, of great interest to study the implication of stochasticity in gene expression for cellular regulation and non-genetic individuality (Powell 1958; Singh 1969; Singh and Gupta 1971; Riney and Schieve 1977). Only in recent years, have new experimental techniques in molecular biology, such as fluorescent reporters, allowed stochastic gene expression to be quantified in vivo (Elowitz, Levine et al. 2002; Ozbudak, Thattai et al. 2002; Blake, M et al. 2003; Raser and O'Shea 2004; Pedraza and van Oudenaarden 2005; Austin, Allen et al. 2006; Dublanche, Michalodimitrakis et al. 2006). These elegant experiments, along with theoretical studies on stochasticity in gene expression (Kepler and Elston 2001; Swain, Elowitz et al. 2002; Paulsson 2004; Tao 2004; Austin, Allen et al. 2006), have greatly facilitated our understanding of the sources and consequences of such stochasticity in GRNs.

Stochasticity in the dynamics of molecular and cellular behaviour, in principle, stems from two sources, intrinsic noise and extrinsic noise. Although the definition of both is somewhat relative, in general, the intrinsic noise is confined in the system and the extrinsic noise is due to the changes in the surrounding environment (Swain, Elowitz et al. 2002; Paulsson 2004). Research has been carried out by different groups trying to separate one type of noise from another theoretically (Swain, Elowitz et al. 2002; Paulsson 2004), and experimentally (Ozbudak, Thattai et al. 2002; Blake, M et al. 2003). In many cases, the evidence has shown that the extrinsic noise dominates the intrinsic noise and sets cell-to-cell variation in both prokaryotes and eukaryotes (Elowitz, Levine et al. 2002; Raser and O'Shea 2004). The sources of the intrinsic and extrinsic noises are described separately below.

- **Intrinsic noise**

Intrinsic noise is inherent in the dynamics of any chemical or biochemical system. In a GRN, specifically, reacting molecules must first find each other through diffusion in a cell and their motion is driven by random collisions. Even if we ignore the diffusion process and assume that all molecules are well-mixed in the cell, reactions occur with a finite probability per unit time, instead of continuously and deterministically. Such stochastic effects are especially important when mean numbers of the molecules are low, which is always the case in GRNs. Unlike metabolic processes, biochemical processes in GRNs are generally in a small volume and have low concentrations of molecular species (McAdams and Arkin 1999). For example, only ten molecules of the
Lac repressor, on average, are present in *E.coli* cells (Lewin 2004). Therefore, the stochastic nature inherited from GRNs in the cells often leads to intrinsic noise which cannot be negligible.

- **Extrinsic noise**

Beside the intrinsic noise, recent studies have demonstrated that a significant component of gene expression variability also arises from external factors (Elowitz, Levine et al. 2002; Raser and O'Shea 2004). The external factors leading to fluctuations in kinetic parameters in biochemical reactions, which, in turn, influence the expression of gene of interest are referred to as “extrinsic noise”. One of the most obvious external factors is the random variation of environmental conditions, such as thermal fluctuations (Blake, M et al. 2003). External factors could also come from the internal processes of the cell, including the variance of number of RNAPs, ribosomes and degradosomes, the timing of gene expression in different stage in cell cycles, the quantity of proteins, and energy demand (Swain, Elowitz et al. 2002). Therefore, the sources of the extrinsic noise arise independently from the components of interest in the system and do not depend on changes in system size.

- **Roles of noise**

Unlike the role of noise in engineering networks in which noise mostly causes destructive effects so that system stability declines, noises have both positive and negative aspects in biological networks. On the one hand, noise is found to be harmful. For example, it disrupts the fine-tuned process of development so that developmental switches have evolved so as to minimise the disruptive effect of such fluctuations (von Dassow, Meir et al. 2000). On the other hand, there are also numerous theoretical studies showing that noise plays important beneficial roles in biological networks. For example, noise can enhance the functioning of biochemical networks, by increasing the sensitivity (Paulsson, Berg et al. 2000) or by driving oscillations (Vilar, Kueh et al. 2002). Noise can also induce stochastic switching which provides a mechanism for phenotypic and cell-type diversification (Arkin, Ross et al. 1998; Hartwell, Gill et al. 1999; Kussell and Leibler 2005).
2.2 Multi-scale issues in modelling biochemical systems

In modelling systems, including biochemical systems, simplifications are unavoidable. Depending on the level of detail that a model intends to capture, certain assumptions should be made to ignore the effects of some unnecessarily detailed processes without a significant loss of higher level of knowledge that can be acquired in a system. In physics, there are some well-defined rules. For example, the macroscopic object obeys the laws of classical mechanics, whereas these laws no longer hold true in mesoscopic and microscopic physics, which obey the laws of quantum mechanics. Modelling a biochemical system is similar; models have to concentrate on a particular focus due to the multiple time and space scales and, frequently, the lack of low level data. Sometimes computational cost is also a factor to account for. We can clarify the scales involved in biochemical reactions as follows:

- Macroscopic scale: In this scale, we assume that the system is a well-mixed solution or, equivalently, is homogeneous. The behaviour of every particle is assumed to be the average behaviour of its kind. Therefore, particles are treated as concentrations (the number of molecules per unit volume) and models in this level are normally expressed by differential equations. Because the chemical reaction is described by increasing or decreasing concentration levels, the changes in state of the system are continuous.

- Microscopic scale: This is the lowest level of reactions, where atom-atom, atom-molecule or molecule-molecule collisions take place. The Avogadro number is the number of formula units in a mole and it describes the fundamental quantitative relationship between macroscopic and microscopic levels: one mole of atoms or molecules = 6.022×10^{23} atoms or molecules. The system in the microscopic level is represented by single molecules, each with a position and a momentum. Hence, the dynamics are stochastic in contrast to macroscopic computation where the dynamics can be described through averaging theorems.
• Mesoscopic scale: This intermediate description of chemical reactions incorporates the information between the microscopic and macroscopic scales in a suitable way. The boundaries are not sharp but can be roughly indicated. In the mesoscopic level we eliminate some irrelevant or poorly understood variables. For example, we assume the solution is well-mixed, therefore, we only count the molecules in a system, rather than keeping track of their individual properties. Because every particle is treated as an individual in this level, the dynamics of the system is stochastic with states changing discretely.

2.3 Kinetic models and the Hill function

Deterministic kinetic modelling of a chemical reaction is in the macroscopic scale which describes the dynamic behaviour of concentrations of reactive components. The rate of a reaction representing concentration change per unit time can be usually written as a function of the concentrations of reactants and products.

As mentioned previously, the chemical systems are assumed to be spatially homogeneous at the macroscopic level. This assumption enables the reaction rate at a time to be a unique function of the concentrations of all participating chemical species. There exist a number of kinds of rate laws corresponding to different types of reaction mechanisms. Here, one rate law, the Hill function, is described that is used to describe ligand-receptor interactions and the cooperativity between the ligands. Two other commonly used rate laws, the mass action rate law and Michaelis-Mention kinetics are reviewed in Appendix A.

The Hill function was first proposed by A.V. Hill (1910) to describe the binding of oxygen to haemoglobin. From then, it has been widely used to analyse the binding equilibrium in ligand-receptor interactions. In a GRN, the binding of proteins (TFs) to a promoter region in a DNA can be viewed as ligand-receptor interactions. As mentioned previously, TFs binding at the promoter region affect the transcription initiation by RNAP. Therefore, it is important to calculate the DNA binding activity of TFs which is
directly linked to its ability to regulate transcription. The Hill function provides a possible way for this calculation using some assumptions.

For simplicity, we first consider the case of binding one TF to a promoter with only one binding site. It can be formulated as

\[ [T] + [P] \xrightleftharpoons[k_2]{k_1} [TP], \]

(2.1)

where \([T]\) denotes the concentration of TFs which are not bound to a promoter. \([P]\) denotes the concentration of an unbound promoter, \([TP]\) is the concentration of a promoter bound with a TF, and \(k_1\) and \(k_2\) are the binding and unbinding rates at the unbound or bound site.

Assuming the binding and unbinding processes of TFs to a promoter are very fast, we can consider that the promoter is always in a specific state, either bound or unbound with the TF. We define \(v\) to be the fractional saturation of the TF,

\[ v = \frac{[TP]}{[P_T]}, \]

(2.2)

where \([P_T]\) is the total concentration of the promoter. We know that \([P_T] = [P] + [TP]\) and \(k_2 = ([T][P])/[TP]\) according to the mass action rate law. In addition, if we define \(K_{div}\) as the equilibrium dissociation constant following tradition, where \(K_{div} = k_2\), then \(v\) can be expressed in terms of free TFs (T) in the following way,

\[ v = \frac{[T]}{K_{div} + [T]}, \]

(2.3)
In some more complex cases where there are multiple binding sites in a promoter, there are three possible cooperativities existing between the binding sites, named non-cooperativity, positive cooperativity and negative cooperativity, as discussed in Section 2.1.4. For mathematically modelling the cooperativity, we consider two special cases first.

In the first case, if we assume \( n \) TFs bind to \( n \) binding sites independently of each other, then all the \( k_1 \) and \( k_2 \) values are the same. The expression of \( v \) can be derived from Eq. (2.3) by summing all the \( n \) binding sites,

\[
v = \frac{n[T]}{K_{dis} + [T]}, \tag{2.4}
\]

In the second case, if we assume \( n \) TFs bind to \( n \) binding sites with infinite cooperativity among the TFs, then there are only possibilities for the binding of \( n \) TFs to bind the promoter, either all the binding sites in the promoter are bound or none of the binding sites in the promoter are bound. We can write the reaction as Eq. (2.1),

\[
[P] + n[T] \rightleftharpoons [PT_n], \tag{2.5}
\]

where \([PT_n]\) is the concentration of a promoter where all the \( n \) binding sites are bound to \( n \) transcription factors. The expression for \( v \) can be expressed as

\[
v = \frac{n[T]^n}{K_{dis} + n[T]^n}, \tag{2.6}
\]

where the dissociation constant \( K_{dis} = \frac{[P][T]^n}{[PT_n]} \) according to the mass action rate law.
For modelling the cooperative binding to $n$ equivalent sites over part of the saturation range (between the two special cases), we can write an equation analogous to Eqs. (2.4) and (2.6), which is known as the Hill function (Hill 1910),

$$v = \frac{n \,[T]^h}{K^h + n \,[T]^h},$$

(2.7)

where $h$ is the Hill constant or the Hill coefficient with $1 < h < n$.

To illustrate the effect of the Hill constant on the fractional saturation of reactant, $v$, in the reactions, we plotted $v$ against the increasing concentration of reactant, $T$, from 0 to 10 using an $h$ value of 1 (no cooperativity) and 2 (positive cooperativity), respectively (Figure 2-2). For simplicity, a $K$ value of 1 is taken. It is clearly shown that the fractional saturation of reactant reaches its maximum more rapidly with positive cooperativity.

### 2.4 Stochastic modelling of intrinsic noise

For the deterministic approach, molecule populations are described by continuous state variables as concentrations. Two important underlying assumptions are: (1) there are a large number of molecules of interest and (2) the system is a continuously predictable process. These assumptions are reasonable for large-scale chemical reactions where the behaviour of each molecule can be viewed as the average of the whole system. However, they fail to capture the discrete and stochastic behaviour of chemical processes when the system becomes smaller. The underlying reason is that chemical species exist in discrete numbers on microscopic and mesoscopic levels, and reaction events only occur when molecules randomly collide. The principle of stochastic modelling of chemical reactions is that the molecular reactions are, essentially, random processes and the state of the system changes discretely.
Figure 2-2 The fractional saturation of reactions without cooperative binding ($h=1$) and with positive cooperative binding ($h=2$).
This section reviews three types of stochastic modelling. Two of them are exact* and rigorous ways to predict the evolution of a chemical system: (1) chemical master equations (CME), and (2) numerical simulation for constructing sample realisations through the probability distributions. In addition, in order to speed up the simulation, a modicum of exactness can be sacrificed by using (3) the chemical Langevin equation (CLE). A brief note on stochastic mathematical constructs relevant to this thesis is provided in Appendix B.

### 2.4.1 Chemical master equation

At the mesoscopic level, we assume that molecular species in a chemical reaction are spatially homogeneous, so the reaction rate depends only on the current numbers of molecules. Mathematically, a Markov process describes the probability of transitions between states as depending only on the current state of the system, therefore, at this level, chemical reactions can be described as Markov processes (Kampen 2001).

Suppose there is a spatially homogeneous reaction system of a fixed volume at constant temperature $T$. This system consists of $N$ molecule species $S_i (i = 1,\ldots,N)$ and $M$ possible different reactions $R_j (j = 1,\ldots,M)$. The state of the system $X$ is defined by $X(t) \equiv (X_1(t),\ldots,X_N(t))$, where $X_i(t)$ is the number of $S_i$ molecules in the system at time $t$ ($i = 1,\ldots,N$).

The fundamental hypothesis of the stochastic formulation of chemical kinetics is the definition of the stochastic rate constant $c_j$ associated with the reactant, as follows,

$$c_j \, dt \equiv \text{average probability that a randomly chosen combination of } R_j \text{ reactant molecules will react accordingly in the next infinitesimal time interval } dt. \quad (2.8)$$

---

* Exactness of a stochastic approach is ‘in the sense that it takes full account of the fluctuations and correlations’ of reactions (Gillespie 1977).
Gillespie (1977) explained the microphysical meaning of chemical reactions based on basic Newtonian physics and thermodynamics. Next, we define the probability $a_j$ as the probability density of reaction $R_j$, also called the propensity function,

$$a_j(x) \, dt \equiv \text{probability that one reaction } R_j \text{ will happen somewhere inside the system in the next infinitesimal time interval } [t, t+dt) \text{ given the state } X(t) = x. \quad (2.9)$$

The probability that a reaction $R_j$ will occur somewhere in the system in the time interval $dt$ can be described by

$$a_j(x) \, dt = c_j \, h_j(x) \, dt, \quad (2.10)$$

where $h_j$ denotes the number of possible combinations of reactant molecules involved in reaction $R_j$. In the case of a first order reaction, for example $X_1 \xrightarrow{a_1} X_2$, $h_1(x) = x_1$. In a second order reaction, for example, $X_1 + X_2 \xrightleftharpoons[a_1]{a_2} 2X_1$, we have $h_1(x) = x_1 \, x_2$. and for the inverse of this reaction, we have $h_2(x) = x_1 (x_1 - 1)/2$.

The state-change vector $v_j$ whose $i$th component is defined by

$$v_{ji} \equiv \text{the change in the number of } S_i \text{ molecules produced by one } R_j \text{ reaction } (j = 1, \ldots, M; i = 1, \ldots, N) \quad (2.11)$$

The state of vector $x$ will change to $x + v_j$ after an occurrence of reaction $R_j$. If we take a time increment $dt$ that is small enough so that the probability for two or more reactions to occur in $dt$ is negligible compared to the probability for one, we can derive a time evolution equation to describe the system state $x$ at time $t + dt$ as the sum of the
probabilities of all the possible precursor states at time $t$. The probability $P(x,t)$ that the system is in state $x$ at time $t$ obeys,

$$P(x,t + dt) = P(x,t)(1 - \sum_{j=1}^{M} a_j(x) dt) + \sum_{j=1}^{M} (P(x-v_j,t) a_j(x-v_j) dt). \quad (2.12)$$

The first term on the right hand side of Eq. (2.12) represents the probability that $x$ retains its state unchanged over $dt$, whereas the second term is the probability that one reaction occurs over $dt$ which leads the state change to $x$. If we take the limit as $dt \to 0$, we have arrived at the CME,

$$\frac{\partial P(x,t)}{\partial t} = \sum_{j=1}^{M} a_j(x-v_j) P(x-v_j,t) - a_j P(x,t). \quad (2.13)$$

### 2.4.2 Gillespie algorithm

CME attempts to write all the possible transition states and solve them simultaneously. Although the inherent stochasticity of the system is mathematically formalised intuitively, the number of possible trajectories of the state transition increases exponentially when the dimension of a system increases. The dimension of the system depends not only on the number of chemical species $N$ but also on any possible number of molecules of any species in the $M$ equations. In this case, the analytical solution of CME is difficult to achieve and generally impractical. Subsequently, Gillespie developed an effective algorithm to simulate the CME numerically (Gillespie 1976; Gillespie 1977).

The basis of the Gillespie algorithm is the next-reaction density function, which is defined by,
\[ P(\tau, j) \, d\tau \equiv \text{the probability at time } t \text{ that the next reaction will occur in an infinitesimal time interval } (t + \tau, t + \tau + d\tau) \text{ and will be a reaction of } R_j, \text{ where } 0 \leq \tau < \infty. \] (2.14)

The main idea of the Gillespie algorithm is to construct sample paths or realisations of \( X(t) \) using the next-reaction density function \( P(\tau, j) \, d\tau \). Suppose a system commences at \( t_0 \) with some initial state, and the time to the next reaction and the index of that reaction are chosen randomly according to \( P(\tau, j) \, d\tau \). After an occurrence of the reaction \( R_j \), the new state of the system is generated for time \( t + \tau \), and the reaction density function \( P(\tau, j) \, d\tau \) is recalculated. This process is executed repeatedly, eventually forming a complete evolution of the system.

From the definition of \( P(\tau, j) \, d\tau \), we can note that this density function is equal to the probability of no reaction over time interval \( (t, t + \tau) \) multiplied by the probability that \( R_j \) will occur over time interval \( (t + \tau, t + \tau + d\tau) \). The first multiplicative term is defined as \( P_0(\tau) \) and the second term is simply equal to \( a_j \, d\tau \) according to the definition of propensity function (Eq. (2.9)), Thus

\[ P(\tau, j) \, d\tau = P_0(\tau) a_j \, d\tau. \] (2.15)

To find an expression for \( P_0(\tau) \), we note its time evolution, given by

\[ P_0(\tau' + d\tau') = P_0(\tau') (1 - \sum_{k=1}^{M} a_k \, d\tau'), \] (2.16)

where \( 1 - \sum_{k=1}^{M} a_k \, d\tau' \) is the probability that no reaction will occur in time \( d\tau' \). Eq. (2.16) leads to

\[ P_0(\tau) = \exp(-\sum_{k=1}^{M} a_k \, d\tau). \] (2.17)
Inserting (2.17) into (2.15), we obtain the next-reaction density function,

$$P(\tau, j) \, d\tau = a_j \exp(-a_0 \tau) \, d\tau,$$  \hspace{1cm} (2.18)

where $0 \leq \tau < \infty$, $j = 1, \ldots, M$, and $a_0 \equiv \sum_{k=1}^{M} a_k$ (i.e., the next-reaction density function for all the possible reactions in the system). The propensity function can be further decoupled into two probability distributions using Bayes’ rule,

$$P(\tau, j) = P(\tau) \, P(j|\tau).$$  \hspace{1cm} (2.19)

$P(\tau)$ accounts for the duration $\tau$ at which the reaction occurs and $P(j|\tau)$ accounts for the probability that reaction $R_j$ will occur next. Since the probability distribution of reactions and the probability distribution for times are independent, $P(j|\tau)$ is equal to $P(j)$. We conclude that

$$P(\tau) = a_0 \exp(-a_0 \tau),$$  \hspace{1cm} (2.20)

and

$$P(j) = a_j / a_0.$$  \hspace{1cm} (2.21)

These two distributions lead to the Gillespie’s direct algorithm. The algorithm generates “unbiased realisations” of the stochastic time evolution of a chemically reacting system. Such realisations are fully consistent with the CME (2.13), since both Eqs. (2.13) and (2.18) are based on the same microphysical premise (propensity function). It should be noted that one simulation provides only one realisation of the total ensemble of possible time evolutions of the system starting from a given initial state. Therefore, a number of simulations are required if we want to estimate any of the moments of $X_i^{(k)}(t)$, where $k$ is the order of the moments. Any moment $X_i^{(k)}(t) \equiv \langle X_i^{(k)} \rangle$, may be estimated directly as the average of the $k$th power of the numbers found for $X_i$ at time $t$ in these runs.
Typically, the mean value $\langle X_i^{(1)} \rangle$, the second moment $\langle X_i^{(2)} \rangle$, and the covariance functions, such as $\langle X_i^{(1)} X_j^{(1)} \rangle - \langle X_i^{(1)} \rangle \langle X_j^{(1)} \rangle$, are of interest. The Gillespie algorithm is given below:

1. Initialization: set initial numbers of molecules $S_1, \ldots, S_N$ for time $t_0$, reaction rate $c_j$ and time $t = 0$.
2. Calculate the propensity function of reaction, $a_j (j = 1, \ldots, M)$.
3. Generate $\tau$ from the exponential probability distribution in Eq. (2.20).
4. Generate $j$ from the discrete probability distribution in Eq. (2.21).
5. Update the $S_1, \ldots, S_N$ values according to the change of reaction $R_j$ set $t \rightarrow t + \tau$.
6. Go to step 2 or Stop if $t_{\text{max}}$ is reached.

### 2.4.3 Chemical Langevin equations

CME and the Gillespie algorithm are both exact consequences of the propensity function, which defines the temporal probability density function for a single molecular reaction. Although the exact stochastic simulation has been proved as the most relevant to the realistic behaviour of chemical reaction systems, such completeness comes at a high computational cost (Haseltine and Rawlings 2002; Rao and Arkin 2003; Salis and Kaznessis 2005). For example, the Gillespie algorithm takes time steps of variable length, based on the rate constants and the numbers of each chemical species. The probability of one reaction occurring relative to another is obtained by multiplying the rate constant of each reaction with the numbers of its substrate molecules. The time step has to be small enough so that only one reaction occurs in the time interval. Therefore, the exact stochastic simulation is computationally inefficient when the number of molecules or the propensity function is large.
To derive an approximate solution for the “semi-exact” simulation, a random variable $K_j(x, \tau)$, the number of $R_j$ reactions that occur in a time interval $[t, t + \tau]$, is introduced. By using the definition of the state change vector, the state of system at time $t + \tau$ will be

$$X_j(t + \tau) = X_j(t) + \sum_{j=1}^{M} K_j(X(t), \tau) v_{ji}, (i = 1, ..., N).$$  

(2.22)

The jump Markov Process $X(t)$ in CME (2.12) can be approximated by the continuous Markov process defined by the standard form of the multivariate Langevin equation (Appendix B) under the following conditions:

1. The system possesses a macroscopically infinitesimal time scale, which means that the propensity functions do not change too rapidly over small time intervals and, in this case, the propensity functions can be approximated as

$$a_j(X(t')) \cong a_j(X(t)), \text{ where } t \leq t' \leq t + \tau.$$  

(2.23)

2. Over those same intervals, there are significant activities in all reaction channels. Since the propensity function does not appreciably change during the time interval $[t, t + \tau]$, all reactions occurring in this time interval are statistically independent of each other. As proved by Gillespie (2000), $\{K_j(X(t), \tau)\}$ can be approximated by a Poisson random variable, $P_j(a_j(X(t)), \tau)$ with the mean of

$$\langle P_j(a_j(X(t)), \tau) \rangle = a_j(X(t)) \tau.$$  

(2.24)

If the mean $a_j(X(t)) \tau$ is large, the Poisson random variable $P_j$ can be further approximated by a Gaussian random variable,
\[ P_j(a_j(X(t)), \tau) \approx \mathcal{N}_j((a_j(X(t)) \tau, a_j(X(t)) \tau) \]
\[ = a_j(X(t)) \tau + \sqrt{a_j(X(t)) \tau} \mathcal{N}_j(0,1). \] (2.25)

The third part in Eq. (2.25) is derived according to the linear combination theorem for the normal variable. Thus, using Eq. (2.25), Eq. (2.22) can be converted into the form,

\[ X_j(t + \tau) = X_j(t) + \sum_{j=1}^{M} v_{ji} a_j(X(t)) \tau + \sum_{j=1}^{M} v_{ji} \sqrt{a_j(X(t))} \mathcal{N}_j(0,1). \] (2.26)

Now, we simply replace the time interval \( \tau \) with \( dt \), and replace \( \mathcal{N}_j(0,1) \) with \( \mathcal{N}_j(t) \), where \( \mathcal{N}_j(t) \) are \( n \) statistically independent, temporally uncorrelated random variables, to obtain the standard form of CLE,

\[ X_j(t + dt) = X_j(t) + \sum_{j=1}^{M} v_{ji} a_j(X(t)) dt + \sum_{j=1}^{M} v_{ji} a_j(X(t))^{1/2} \mathcal{N}_j(t) dt^{1/2}. \] (2.27)

Through a few simple algebraic rearrangements and a replacement of \( \mathcal{N}_j(t) \) with Gaussian white noise \( \xi(t) \), where \( \xi(t) \) is rapidly fluctuating random terms with zero mean \( \langle \xi(t) \rangle = 0 \), we arrive at the white noise form of CLE,

\[ dX_j(t) = \sum_{j=1}^{M} v_{ji} a_j(X(t)) dt + \sum_{j=1}^{M} v_{ji} a_j(X(t))^{1/2} \xi_j dt. \] (2.28)

For numerical treatment, it is also often written in the conventional form of the Wiener process term. The introduction of the Wiener process was motivated by its connection with white noise, where \( dW_j = \xi dt \) or equivalently \( W_j = \int \xi_j dt \). Accordingly, Eq. (2.28) is equivalent to
\[ dX_i(t) = \sum_{j=1}^{M} v_j a_j(X(t)) \, dt + \sum_{j=1}^{M} v_j a_j(X(t))^{1/2} \, dW_j, \quad (i = 1, \ldots, N). \tag{2.29} \]

where \( W_i \) is a standard Wiener process whose increment is a Gaussian random variable with \( W_0 = 0, \quad \langle W_0 \rangle = 0 \) and \( \text{var}[W_{t_b} - W_{t_a}] = t_a - t_b \) for \( 0 \leq t_a \leq t_b \).

### 2.4.4 Linear noise approximation

Besides the numerical simulation methods that have been reviewed above, there exists another approximation method which is also used in practice to estimate the effects of intrinsic noise in the literature, called the linear noise approximation (LNA) approach, as proposed by van Kampen (1976).

Before explaining this approach, we first need to introduce a system size parameter, \( \Omega \), that connects the units of volume of the system and the number of molecules. If the concentration of each chemical species is fixed, then changing \( \Omega \) alters the number of molecules of every chemical species. The key assumption of the LNA is that the deterministic evolution of the reactant concentration can be meaningfully separated from the fluctuations, and the fluctuations scale roughly as the square-root of the number of molecules. Now the number of molecules for a species in the system can be written as the sum of its deterministic concentration with a coefficient of \( \Omega \) and its fluctuation with a coefficient of \( \sqrt{\Omega} \). Using the assumption above, the CME can be written in a convenient manner. The LNA is based on a systematic expansion of the master equation in \( \Omega^{-1} \). This leads to a Fokker-Planck like equation that has a set of nonlinear differential equations that govern the deterministic evolution of the system and a partial differential equation that characterises the probability distribution of the fluctuations. Therefore, the expanded equation can accurately describe small fluctuations around the stable attractor of the system. Unfortunately, the LNA becomes intractable once the number of chemical species in the system reaches more than three. Then we need analytical inversions of \( 4 \times 4 \) matrices or calculation of their eigenvalues (Swain and Longtin 2006). Therefore, CLE will be used, instead of LNA, to
approximate the intrinsic noise for our model in this thesis because CLE can be practically applicable for large non-linear systems.

2.5 Stochastic modelling of extrinsic noise

The stochastic approach discussed in the preceding section assumes the intrinsic fluctuations to be Markovian and describes them in the form of a discrete birth and death process via the master equation, which can be numerically solved by the Gillespie algorithm or approximated by CLE or LNA. In chemical or biochemical systems, there is another kind of noise which originates outside the system, called extrinsic noise. The extrinsic noise reflects the random character of the environment. Because the origin of the extrinsic noise stems from the outside of the system, it is completely independent of the system size, whereas the intrinsic noise tends to vanish in the thermodynamic limit. Therefore, the extrinsic noise can be important even for large volume systems.

Extrinsic noise can be studied by a means of stochastic differential equations (SDEs). The SDE, or the equivalent Langevin equation, is obtained from a deterministic approach where the constant value of parameters is replaced by a stochastic process (Horsthemke and Lefever 1984). SDEs have been widely used to study random systems in biology, chemistry, finance and physics. An extensive list of applications of SDEs in sciences is given by Kloeden and Platen (Kloeden, Platen et al. 1997). Some examples of applications are population dynamics (Kiester and Barakat 1974), stock market (Talay and Rogers 1997) and solute transport in porous media (Kulasiri and Verwoerd 2002). A short introduction to the mathematical formalism in the form of SDEs is given below for accounting for irregular rate constants in biological and biochemical systems.

We first consider that one of the parameters $p$ in an ODE is perturbed by some stochastic noises. If, for simplicity, it is assumed that the parameter is perturbed by white noise only, then $\tilde{p}$ has the form,

$$\tilde{p} = p + \alpha W_t$$  \hspace{1cm} (2.30)
where $W_t$ is a standard Wiener process and $\alpha$ is its amplitude coefficient. In general, a stochastic process obeys an Itô SDE which has a general form,

$$dX_t = F(X_t)\,dt + G(X_t)\,dW_t,$$

(2.31)

or in the form of an integral equation,

$$X_t = X_0 + \int F(X_s)\,ds + \int G(X_s)\,dW_s,$$

(2.32)

where $X_t$ represents the macroscopic variable, $F(X_t)\,dt$ is the continuous deterministic component with $F(X_t)$ usually a nonlinear function. $G(X_t)\,dW_t$ is the continuous random component and $G(X_t)$ decides whether the coupling is additive or multiplicative depending on whether it is constant or otherwise. $W_t$ is an multi-dimensional stochastic process which has Wiener process components. Generally, SDEs are not analytically solvable for most practical purposes (Kloeden, Platen et al. 1997). Numerical simulations are necessary to determine the stability and convergence of sample paths, and to compute the probability distribution and statistical measures of the solution.

It should be noted that a CLE is also a SDE which approximates the intrinsic noise by the explicit noise term. In spite of the mathematical similarity, it is distinguished from the SDE biochemically in the sense that the rate constants of reactions represented by the CLE are assumed to be constants whereas those of the SDE are perturbed by fast fluctuating noise.
Chapter 3: From components to systems: biology and models of circadian clock

In the following four chapters, the problem under investigation is the circadian clock in *Drosophila*, which is responsible for maintaining circadian rhythms. The purpose is to develop a detailed mathematical model incorporating the current knowledge of molecular mechanisms about the circadian clock and compare computational simulations with experimental results. This chapter aims to provide a detailed background of the molecular components and interactions in the system and mathematical models.

3.1 Circadian rhythms and circadian clocks

Circadian rhythms affect all aspects of daily life and have long provided a unique point from which to address fundamental and wide-ranging questions of physiology and behaviour. It has been hypothesised that circadian rhythms arose during evolution by anticipating the 24 h rotation of the earth and its consequent light and temperature cycles (Dunlap 1999; Panda, Hogenesch et al. 2002). Anticipating the daily changes enables an organism to prepare itself for the conditions occurring with the highest probability and help itself to save resources for energy production, synthesis of proteins or uptake of nutrients. For example, in *Arabidopsis* production of photo-system I and photo-system II before sunrise, a circadian change allows photosynthesis to start as soon as sun energy is available (Harmer, Hogenesch et al. 2000). Therefore, organisms with circadian rhythms matching the cycling environment are favoured by nature in comparison to those who do not match the daily environmental cycle (Pittendrigh 1959).

The most obvious explanation given for these rhythms is that organisms passively follow the environment periodicity. However, when isolated from the periodic environmental influences, for example, by maintaining them under constant darkness, the large majority of eukaryotes and some prokaryotes still show self-sustained
circadian oscillations (Kondo, Mori et al. 1997). Such experiments demonstrate the existence of an internal biological clock that can function independently of environmental influences. Thus, a critical question has driven circadian biology for decades: what are the mechanisms of circadian timekeeping?

Now it is experimentally established that self-sustaining circadian clocks controlling circadian rhythms regulate hundreds of genes and allow organisms to anticipate daily changes in environmental influences (Pittendrigh 1993; van Gelder, Herzog et al. 2003). In recent decades, many components and molecular mechanisms comprising circadian clocks have been uncovered, largely due to advances in molecular biology experiments (Dunlap 1999). The principal way that molecular components of the clock have been identified is by forward mutagenesis screens. Specifically, mutations in each of these genes result in circadian rhythm abnormalities, ranging from alterations in the circadian period to complete arrhythmicity. The analysis of these proteins and their genes has produced a wealth of information and insight.

Circadian clocks have become one of the most attractive models to study cellular and molecular mechanisms connecting genes and behaviour, and much has been learned about the molecular mechanisms of oscillating clocks in different organisms. The model organisms include unicellular eukaryotes, fungi, plants, invertebrates and mammals (Young and Kay 2001). Because of its ease of genetic manipulation and the property of being suited to large-scale mutant screening, *Drosophila* has contributed most to the timing mechanism studies of the central circadian clock (van Gelder, Herzog et al. 2003).

### 3.2 Circadian system

There are three major components of circadian systems in most organisms, shown in Figure 3-1. The circadian input pathway transmits information from external stimuli such as light and temperature to the internal clock and allows the internal biological clock to synchronise with the environment. The circadian output pathway conveys information from the internal clock to signal the daily changes into the biochemical and
Environmental cycles

Input pathway

Internal clock

Oscillating molecular clock in cells

Output pathway

Biochemical and physiological changes

Sleep/wake cycles, Photosynthesis, Hormonal control, etc.

Figure 3-1 Schematic representation of the circadian systems, adapted from Eskin (1979).
physiological behaviours in the cell. Examples of behaviours controlled by the clock are sleep-wake cycles, photosynthesis, and hormonal control (Schoning and Staiger 2005). The internal clock comprises a number of clock molecules, and autonomously produces circadian oscillations of the clock molecules, with or without external stimuli. It should be noted that the picture of three components is oversimplified because there are numerous overlaps where the different components and pathways can utilise the same molecules for different roles.

Recently, many findings have shown that the molecular mechanisms of the internal clocks among the different organisms share a common theme (van Gelder, Herzog et al. 2003). At the core of all the circadian clocks there is a network of positive and negative elements. The positive elements (i.e. activators) activate the transcription of the negative elements, whereas the negative elements (i.e. repressors) block their own transcription by eliminating the positive elements. Examples of activators are KaiA in *Synechococcus*, WCI-2 in *Neurospora*, CLK and CYC in *Drosophila*, and CLK and BMAL in mice, and examples of repressors are KaiB and KaiC in *Synechococcus*, FRQ in *Neurospora*, TIM and PER in *Drosophila* and PER1, 2 and 3 in mice (Dunlap 1999). The genes and proteins of the activators and the repressors form transcriptional regulatory networks in the circadian clocks with feedback loops.

### 3.3 Molecular basis of the *Drosophila* circadian clock

#### 3.3.1 Molecular components

The *period* (*per*) gene was identified as the first clock component in *Drosophila* by the isolation of its mutants (Konopka and Benzer 1971). Three types of circadian rhythms were found based upon the period of rhythmicity under free running conditions in constant darkness: arrhythmicity in the *per* null allele (*per*⁰), a shortened daily rhythm in the “short” period allele (*per*⁴) and an elongated rhythm in the ‘long’ period allele (*per*¹) (Konopka and Benzer 1971). Following the molecular cloning of the *per* gene, it was found that the PER protein shares a domain with a family of proteins including
ARNT (Arhl hydrocarbon receptor nuclear translocator) and SIM (Single Minded Protein) (Bargiello, Jackson et al. 1984; Crews, Thomas et al. 1988; Hoffman, Reyes et al. 1991). This region is named the PAS domain (PER ARNT SIM) and is thought to be involved in protein-protein interactions (Huang, Edery et al. 1993). Later biochemical studies showed that the PER protein forms heterodimers with the TIM protein, the product of the second clock gene identified in *Drosophila, timeless (tim)* (Gekakis, Saez et al. 1995; Myers, Wager-Smith et al. 1995). As per, a null mutation of tim resulting in arrhythmicity indicates, tim is also vital to generating rhythmic behaviour (Sehgal, Price et al. 1994).

Although the early genetic evidence suggested that the per and tim genes play crucial roles in the circadian cycling machinery, no insight had been given into how these genes participated in the clock mechanism until their discovery by Hardin et al. (1990). They showed that PER feeds back to regulate its own mRNA levels though the observation that mutants in per affected the quantity and quality of per mRNA cycling. Moreover, experiments showed that the molecular levels of per mRNA fell when PER levels rose, and per mRNA levels rose when PER levels fell (Hardin, Hall et al. 1990). These findings suggested that PER directly or indirectly represses its own gene expression. This concept was later incorporated with TIM, which forms a heterodimer with PER, and together these results suggested the oscillating PER and TIM levels result from the oscillating levels of per and tim mRNAs (Sehgal, Rothenfluh-Hilfiker et al. 1995).

However, the mechanism of feedback regulation of per and tim transcription remained poorly understood because neither PER nor TIM was found to have a DNA-binding domain. The breakthrough came with promoter dissection studies which identified a ~70-basepair (bp) enhancer sequence, found ~ 500 bp upstream of the per transcription initiation site, as a circadian regulatory sequence. This fragment, which contains a consensus E-box element (CACGTG), was required for transcriptional activation (Hao, Allen et al. 1997). Similarly, a consensus CACGTG E-box was found in the tim upstream sequence that is also necessary for tim transcription (McDonald and Rosbash 2001). Subsequently, the mechanisms underlying PER and TIM transcription feedbacks became clear when genetic screening for mutations identified that two genes, clock (clk) and cycle (cyc), played critical roles in circadian rhythmicity. In flies with clk and cyc mutations, per and tim expression is arrhythmic and low, suggesting that clk and cyc act
as positive regulators for \textit{per} and \textit{tim} transcription (Allada, White et al. 1998). Since E-box elements are known targets for basic helix-loop-helix domain (bHLH) TFs, and the products of \textit{clk} and \textit{cyc} expression, CLK and CYC, are bHLH-PAS TFs, these suggest a model for a transcription feedback loop: CLK and CYC activate \textit{per} and \textit{tim} transcription by binding to E-box elements within their promoters, and PER and TIM inhibit their transcription by binding CLK and CYC through their PAS domains (Darlington, Wager-Smith et al. 1998).

The model described above explains the generation of PER and TIM oscillation, but it does not explain the oscillation of \textit{clk} mRNA and CLK proteins. In \textit{Drosophila}, \textit{clk} mRNA levels peak just after dawn, roughly in anti-phase with \textit{per} and \textit{tim} mRNA levels (Darlington, Wager-Smith et al. 1998). Moreover, \textit{clk} mRNA levels are constitutively low in \textit{per}^{01} and \textit{tim}^{01} mutants which produce non-functional PER and TIM, suggesting that PER and TIM are required for activating \textit{clk} expression (Bae, Lee et al. 1998). However, \textit{clk} mRNA levels are surprisingly high in \textit{clk}^{brk} which produces non-functional CLK, suggesting that CLK represses its own expression (Glossop, Lyons et al. 1999). Since CLK is known to be a transcription activator and there are no consensus E-boxes in or around the \textit{clk} promoter, it is unlikely that CLK directly represses its transcription. In addition, the high levels of \textit{clk} mRNA in the \textit{per}^{01}; \textit{clk}^{brk} double mutant indicate that a separate \textit{clk} activator is present (Glossop, Lyons et al. 1999). All these findings prompted the discovery of a CLK repressor VRILLE (VRI), which is a rhythmic expressed PAS domain factor and is activated by CLK/CYC. It was found that a consensus “VRI box” is in the \textit{clk} promoter and VRI over-expression represses \textit{clk} mRNA levels, suggesting that VRI directly represses \textit{clk} expression (Glossop, Lyons et al. 1999). Subsequently, a second PAR domain factor was identified by its homology with VRI, the PAR Domain Protein 1 (PDP1) (Cyran, Buchsbaum et al. 2003). Like the \textit{per} and \textit{tim} genes, E-boxes are also found in the promoters of \textit{vri} and \textit{pdp1} genes and CLK/CYC dimers have been shown to activate \textit{vri} and \textit{pdp1} expression \textit{in vitro} in an E-box-dependent manner (Cyran, Buchsbaum et al. 2003; Glossop, Houl et al. 2003). Both VRI and PDP1 belong to basic zipper TFs with highly conserved basic DNA binding domains, suggesting that they bind to the same set of target genes. Indeed, \textit{in vitro} experiments showed that PDP1 can bind to the VRI box consensus sequence (henceforth referred to as a V/P box), and compete with VRI to regulate the \textit{clk} mRNA (Cyran, Buchsbaum et al. 2003).
In summary, six genes have been identified as necessary for the circadian clock functions in *Drosophila*. These genes can be divided into two categories according to the molecular nature of their protein products. These proteins include (1) transcriptional activators: CLK, CYC and PDP1; (2) transcriptional repressors: PER, TIM and VRI. These components appear to be organised into a transcriptional regulatory network where the protein products of one or more clock genes indirectly regulate expression of their own genes.

### 3.3.2 Transcriptional feedback loops

The *Drosophila* circadian clock is composed of two interlocked feedback loops in gene expression, as shown in Figure 3-2 (Glossop, Lyons et al. 1999; Cyran, Buchsbaum et al. 2003; Hardin 2005).

The first loop, named the PER/TIM loop, starts with activation of the *per* and *tim* expression from mid day. Activation of the *per* and *tim* transcription is mediated by two TFs, CLK and CYC. CLK and CYC form dimers that target E-boxes in the *per* and *tim* promoters (Allada, 1998). After initial activation of the *per* and *tim* expression, there is a 4 h – 6 h delay between the peak concentrations of *per* and *tim* mRNAs and that of PER and TIM proteins (Zerr, Hall et al. 1990; Zeng, Qian et al. 1996). As a result, CLK/CYC can continue to activate transcription of the *per* and *tim* genes, while PER and TIM proteins accumulate in the cytoplasm. PER and TIM also form PER/TIM dimers while accumulating. In the middle of the night PER/TIM dimers are transported into the nucleus. After entering the nucleus, they can bind to CLK/CYC dimers effectively inhibiting CLK/CYC binding ability to E-boxes without disrupting the dimeric structure of CLK/CYC (Lee, Bae et al. 1998). This inhibition lasts until PER and TIM proteins are degraded. Then the expressions of *per* and *tim* are reactivated by CLK/CYC dimers in the following mid day.

The second loop, named the VRI/PDP1 loop, consists of a VRI-mediated negative feedback loop and a PDP1-mediated positive feedback loop. This loop starts with
Figure 3-2 Interactions in the two loop model of Cyran et al. (2003).
activation of *vri* and *pdp1* transcription by CLK/CYC during the late day and early night. VRI accumulates first in phase with its mRNA then PDP1 accumulates during the mid to late evening. VRI binds the V/P box in the *clk* regulatory elements to inhibit the *clk* transcription and PDP1 can compete with VRI for binding to the V/P box and actives the *clk* transcription (Cyran, Buchsbaum et al. 2003). The effects from the initial VRI-dependent repression in the early night and the subsequent PDP1-dependent activation in the middle to late night determine the rhythmic expression of *clk*. However, the newly produced CLK at the end of night and early morning is inactive temporarily due to high levels of PER/TIM dimers induced by the previously produced CLK. Once PER/TIM dimers are degraded, CLK/CYC reactivates the gene expression of *per*, *tim*, *vri* and *pdp1* and starts a new cycle.

In addition to regulation at the transcriptional level, many clock components in *Drosophila* are also regulated post-transcriptionally and post-translationally. For example, Doubletime (DBT) destabilises PER. Casein Kinase 2 (CK2) destabilises PER and also affects its nuclear localisation. Shaggy (SGG) phosphorylates TIM to promote nuclear localisation of PER/TIM dimers. Slimb (SLMB) targets phosphorylated PER for degradation (Hardin 2005). These processes might provide time delays between mRNAs and proteins. For example, a 4 h – 6 h delay between accumulation of *per* mRNA in the cytoplasm and PER in the nucleus results from the initial destabilisation of PER by DBT dependent phosphorylation, and possibly also CK2 dependent phosphorylation, followed by the stabilisation of PER by dimerisation with TIM before nuclear entry (Price, Blau et al. 1998).

### 3.4 Mathematical models of the circadian clock in *Drosophila*

A range of mathematical models for the circadian clocks in different organisms have been proposed in the literature, including *Arabidopsis thaliana*, *Neurospora*, *Drosophila* and mammals (Goldbeter 1995; Smolen, Baxter et al. 2001; Leloup and Goldbeter 2003; Locke, Southern et al. 2005). The common character of these models is that they contain at least one negative feedback loop and have a capability to produce sustained
The first circadian clock model of *Drosophila* was proposed by Goldbeter (Goldbeter 1995) which is based on the negative feedback exerted by PER on the transcription of the *per* gene. This simple model containing five variables described the multiple phosphorylation of PER. Numerical simulations showed that a single negative feedback of PER alone can produce limit-cycle oscillations for appropriate parameter values. However, the early model did not account for the effect of light on the circadian system because the light receptor TIM was not taken account in the model.

Later, Leloup et al. presented an extended model based on the auto-regulatory negative feedback exerted by a complex between PER and TIM proteins on the expression of *per* and *tim* genes (Leloup and Goldbeter 1998; Leloup and Goldbeter 2000). The model produced essentially the same result as the first model; in addition, it explicitly incorporated the effect of light on the TIM degradation rate. A closely related model incorporating the formation of a PER-TIM complex has been proposed for *Drosophila* circadian rhythms (Tyson, Hong et al. 1999). The difference from the model proposed by Leloup et al. is that the model contains an additional positive feedback loop based on stabilisation of PER upon dimerisation. The model proposed by Tyson et al. accounted for several properties of circadian rhythms, including temperature compensation and the *per*<sup>L</sup> mutant.

With the discovery of an additional regulatory loop related to the *clk* gene and CLK (Glossop, Lyons et al. 1999), more detailed mathematical models involving two interlocked *per-tim* and *clk* loops were created and examined (Smolen, Baxter et al. 2001; Ueda, Hagiwara et al. 2001). The simulations showed that the model can produce sustained oscillations of *clk* mRNA and CLK which were not explained in previous models, and the analysis revealed that the interlocked feedback model provided a possible explanation for the robust oscillation of *Drosophila* circadian rhythms.

More recently, while two additional proteins, VRI and PDP1, were identified to be involved into the regulation of *clk* gene, two new models that reproduced CLK
expression regulation by VRI and PDP1 were proposed (Smolen, Hardin et al. 2004; Ruoff, Christensen et al. 2005). The model proposed by Smolen et al. (Figure 3-3) contained feedback loops based on transcriptional regulation of per, clk, pdp1, and vri, in particular, pdp1 expression was modelled with time delay. The role of PER was described in detail in the model in which PER protein underwent a two-step phosphorylation in the cytoplasm and nucleus. Simulations suggested that vri and pdp1 feedback loops were not essential for oscillations, however, the negative feedback loop in which PER represses per expression was critical for producing oscillations. In addition, the model simulated a range of behaviour of the circadian clock in Drosophila, including null mutations of per, vri, pdp1 and clk, photic phase-response curves resembling experimental curves, and the entrainment to light-dark cycles. The incompleteness of Smolen’s model is that the regulation of TIM was not included.

The core of the model proposed by Ruoff et al. (2005) is that CLK is subjected to positive and negative regulations by PDP1 and VRI, whose transcriptions are activated by CLK (Figure 3-4). The model did not differentiate the per and tim gene expressions, instead, treating PER/TIM complex as a whole whose expression is activated by CLK. The results of simulations suggested that the positive feedback loop and negative feedback loop of pdp1 and vri were essential for the overall oscillations. The PER/TIM complex only played a role in amplification and stabilisation of the oscillations. This conclusion contradicts the one drawn by Smolen in which the PER feedback loop was found to be vital to produce oscillations for the circadian clock system. The other contribution of this model is the calculation of the phase resetting of temperature compensation and losses of temperature compensation in perS and perL mutants. However, the model showed poor entrainment under light/darkness cycles.

Although all the mathematical models reviewed above can produce oscillations of some mRNAs or proteins in the clock system, it is however an incomplete view of the regulatory networks of the circadian clock in Drosophila. No previous model has included the six TFs unveiled in the in vivo experimental work (PER, TIM, CLK, CYC, VRI and PDP1). We propose, in the next chapter, a new model incorporating the current knowledge of the clock that has been discussed in this chapter.
Figure 3-3 The model proposed by Smolen et al. (2004). (A) The model accounts for three feedback loops. In the per loop, PER interacts with CLK forming a negative feedback loop. In the vri loop, vri is activated by CLK, and VRI in turn represses clk. In the pdp1 loop, pdp1 is activated by CLK, and PDP1 in turn activates clk. (B) PER undergoes two cytosolic phosphorylations and then enters the nucleus where PER interacts with CLK, suppressing CLK’s activation of per. Nuclear PER undergoes further phosphorylations before degradation.
Figure 3-4 The model proposed by Ruoff et al. (2005), where dCLK denotes *Drosophila* CLK, the subscript letters “c” and “n” denote cytoplasm and nucleus, respectively, and *per/tim*, as well as PER/TIM, are treated as one component in the system. In the core of this model, the transcription factor CLK is subjected to positive and negative regulation by the proteins PDP1 and VRI, whose transcription is activated by CLK. CLK also activates the clock genes *per* and *tim* and the PER/TIM complex binds to CLK and, thus, reduces the activity of CLK.
Chapter 4: Development of a new circadian clock model

In this chapter, a mathematical model of the circadian clock is developed. In Section 1, a conceptual model is developed based on the previously described molecular basis and a number of assumptions. In Section 2, the modelling method for transcription processes for the current system is developed. In Section 3, the conceptual model is converted into a mathematical model as a set of non-linear ODEs.

4.1 Conceptual model

Based on the molecular basis of the circadian clock reviewed in the previous chapter, a conceptual model was developed as schematised in Figure 4-1. The core structure of the model is similar to the model proposed by Cyran et al. (2003), as schematised in Figure 3-2. The model contains two feedback loops, namely the per/tim loop and vri/pdp1 loop. These two loops are linked by the requirement of CLK/CYC as can be seen in the middle of Figure 4-1. In one loop, per and tim genes are activated by CLK/CYC; and their protein products, PER and TIM, form a dimmer, PER/TIM, to repress their own gene by forming a complex with CLK/CYC. In the other loop, vri and pdp1 genes are also activates by CLK/CYC; and the protein product of vri, VRI, represses the clk gene expression while the protein product of pdp1, PDP1, activated the clk gene expression. The development of the conceptual model was also relied upon a number of assumptions that were used for simplifying the model, and the rationale of the assumptions is as follows:

1. The separate nuclear and cytoplasmic compartments are ignored in the model; instead we assume that all the reactions take place over a whole cell. Although eukaryotic species have compartments separated by nuclear membranes and TFs have to be located into the nucleus in order to affect gene expression, some prokaryotes, which lack a nucleus or nuclear envelope, such as cyanobacteria, can also generate circadian rhythms. This demonstrates that it is possible for cells to
Figure 4-1 The schematic diagram of the model. The model shows the regulatory relationships among genes, mRNAs and proteins in the negative and positive transcriptional feedback loops. Transcription of per, tim, vri and pdp1 genes are activated by CLK/CYC dimers binding to E-boxes in their promoter regions. In one loop, per and tim mRNAs are translated to PER and TIM proteins which form PER/TIM dimers. PER/TIM binds to CLK/CYC to form PER/TIM/CLK/CYC complex. In another loop, vri and pdp1 mRNAs are translated to VRI and PDP1 proteins. They compete to bind the V/P box in the promoter in clk gene. Transcription of clk gene is repressed by VRI and activated by PDP1. clk mRNA is translated to CLK which forms CLK/CYC dimers with CYC. Proteins, mRNAs, dimers and complexes are degraded at certain kinetic rates. CYC is assumed to be constant, therefore, there is no degradation process for CYC. Variable names used in the model are indicated in parentheses. The number of the E-boxes and the V/P box in the promoters is also shown here.
maintain sustained circadian rhythms without compartmentalisation. A theoretical study by Kurosawa et al. (2002) also showed that a cell can generate a sustained oscillation in the absence of compartmentalisation with a single negative feedback model.

2. Phosphorylation of proteins is not considered. Although we are aware that phosphorylation is important to provide the time delay between mRNAs and proteins, as reviewed above, the focus of the current study is on the transcriptional regulation, and phosphorylation of proteins is not included at this stage for the sake of simplification.

3. Gene expression of per, tim, vri and pdpl is activated by binding of CLK/CYC dimers to E-boxes in their promoter regions. Analysis of the first 4 kb of sequence upstream of the start site of pdpl transcription revealed six E-boxes (Cyran, Buchsbaum et al. 2003). The vri promoter sequence was searched and four E-boxes were found (Blau and Young 1999). In the tim promoter, three functional E-boxes were discovered within about 150 bp (McDonald and Rosbash 2001). In addition, two TER boxes (11-bp Tim E-box-like repeats) that serve as additional binding sites for CLK/CYC dimers were also found in the tim promoter (McDonald and Rosbash 2001). Therefore, five binding sites are assumed, including E-boxes and E-box-like binding sites, in the tim promoter region. Five E-boxes were found in the per1 promoter in mammals (Yamaguchi, Mitsui et al. 2000), and the similar case is assumed in the current model for the per promoter.

4. PER/TIM dimers are assumed not to bind to CLK/CYC dimers if the latter are bound to promoters. In mammals, mCRY complexes bind to CLK/BMAL1 and repress transcription without removing CLK/BMAL1 from E-boxes (Etchegaray, Lee et al. 2003). However, in Drosophila, PER/TIM has not been shown to bind to CLK/CYC complexes which are bound to E-boxes (Yu, Zheng et al. 2006).

5. In vitro experiments showed that the concentration of CYC is always constitutive, with high levels in the cells (Glossop, Lyons et al. 1999). Therefore, it is assumed
that the concentration of CYC in the system is constant (1.00 nM is assumed) so that there is always enough CYC bound to CLK to form dimers.

### 4.2 Modelling transcription processes

As discussed in Chapter 3, there are six TFs in the network of the circadian clock in *Drosophila*. CLK and CYC form CLK/CYC which activates the transcription processes of per, tim, vri and pdp1 by directly binding the E-boxes in their promoters. PDP1 activates the transcription of clk, and VRI represses it by competing to bind to the V/P box in the *clk* promoter. In addition, PER/TIM inhibits the transcription processes of per, tim, vri and pdp1 by disabling their activators, CLK/CYC.

To describe the transcription processes, two methods are commonly used. The first one is that the transcription rate is assumed to be an increasing or decreasing function of activators or repressors, and a Hill function is used to describe the activation or repression term of mRNAs. This method has been used in most of the previous circadian clock models (Goldbeter 1995; Ruoff, Vinsjevik et al. 2001; Smolen, Hardin et al. 2004; Locke, Southern et al. 2005). The second one is that binding and unbinding of activators or repressors to promoter sites are modelled explicitly with forward and reverse reactions. This method has been used in a genetic circadian clock model and a mammalian clock model (Vilar, Kueh et al. 2002; Forger and Peskin 2003).

Here, we propose to use the explicit binding/unbinding and activation/repression processes, instead of using the conventional Hill function. First, the justification for developing such explicit description is given. Then a detailed derivation of our method is provided.
4.2.1 Motivation for using detailed modelling of the transcription processes

A common characteristic of the previous models for the circadian clock in *Drosophila* is that the activation and repression processes were described by the Hill function which implied switch-like behaviour of the transcriptional effects. With such transcriptional description, the models produced sustained oscillations in appropriate parameter regimes (Leloup and Goldbeter 1998; Smolen, Baxter et al. 2001; Ueda, Hagiwara et al. 2001). These models included the repression of *per* and *tim* by PER/TIM or the activation of *per* and *tim* by CLK (or CLK/CYC). With the assumption of fast binding/unbinding between TFs and promoters, the simultaneous expression of *per* and *tim* can be readily modelled by the Hill function. However, there are two inherent difficulties using the Hill function to simulate the transcription processes if we incorporate the expression of *vri* and *pdp1* into the model.

First, some experiments have shown that there are 3 – 4 h time lags between the rise of *vri* and *pdp1* expression although they are both activated by CLK/CYC (Cyran, Buchsbaum et al. 2003). The simulated concentration of *vri* and *pdp1* mRNAs would arise concurrently under the assumption of fast binding by the Hill function. In a previous model, which included *vri* and *pdp1* expression proposed by Smolen et al. (2004), the authors did not take account of mRNA in the model and used a delay term to simulate the several hours between VRI and PDP1 proteins. However, the different phases of *vri* and *pdp1* mRNAs may reflect the subtly different transcriptional activities of their promoters and the *vri* promoter could be stronger than the *pdp1*, as suggested in a recent research (Cyran, Buchsbaum et al. 2003). We propose that this subtle mechanism underlying the time lag could be easily created by the explicit description of binding and unbinding processes.

Second, it has been shown in *in vitro* experiments that PDP1 and VRI directly regulate *clk* transcription by competing to bind with the same site in the *clk* promoter (Cyran, Buchsbaum et al. 2003). The Hill function can be used to represent activation by PDP1 and repression by VRI, but cannot represent the details of the competition. For a clear
illustration, we take the following equation which describes the regulation of \( clk \) expression by PDP1 and VRI from the model proposed by Smolen et al. (2004),

\[
R_{clk} = V_{clk} \left( \frac{[PDP1]^2}{[PDP1]^2 + K_{PDC}^2} \right) \left( \frac{K_{VC}^2}{[VRI]^2 + K_{VC}^2} \right) + R_{bas},
\]

(4.1)

where \( R_{clk} \) = the effective transcription rate of \( clk \), \( V_{clk} \) = the maximal transcription rate of \( clk \), and \( R_{bas} \) = a small basal transcription rate in the absence of PDP1 and VRI. The activation of PDP1 is represented by a Hill function of PDP1, where \( K_{PDC} \) is the dissociation constant of PDP1 binding to the \( clk \) promoter. The inhibition of VRI is represented by a Hill function of VRI, where \( K_{VC} \) is the dissociation constant of VRI binding to the \( clk \) promoter. Even though this equation does have a competitive component as a certain decrease of PDP1 can be compensated by an increase of VRI, it does not show the competition between PDP1 and VRI because PDP1 and VRI can affect the \( clk \) expression simultaneously.

4.2.2 Modelling transcriptional activation by CLK/CYC

For now, we consider that the efficient transcription rate of a gene is the sum of the probabilities of its promoter in particular states multiplied by the transcription rates when the promoter is in that state. We first consider the transcription of \( per \), \( tim \), \( vri \) and \( pdp1 \) expression activated by CLK/CYC. As a promoter can be either activated or deactivated, we can calculate the probability that the promoter is activated or deactivated.

As have discussed in the previous chapter, there are a number of E-boxes in the promoter region of \( per \), \( tim \), \( vri \) and \( pdp1 \) genes. Here we make an important assumption that CLK/CYC dimers independently bind to an individual E-box in a promoter. In the functional analysis of E-boxes in the mouse \( mPer1 \) promoter, the levels of \( mPer1 \) transcriptional expression activated by CLK/BMAL1 were roughly proportional to the number of conserved E-boxes (Hida, Koike et al. 2000). The result suggests that there is
no or negligible cooperative interaction in the E-box binding activities of CLK/BMAL1. Since no information is available about cooperativity in the E-box binding activities by CLK/CYC in Drosophila, we treat it as the case in mPer1 promoter. For the same reason, we also assume that if CLK/CYC is bound to just one E-box for a given gene, the transcription of that gene is activated and the effect of binding additional E-boxes on transcription activation is additive (Hida, Koike et al. 2000).

Suppose there are $n$ E-boxes in a promoter where CLK/CYC dimers can bind. Since CLK/CYC dimers bind to individual E-boxes independently, we can consider each E-box separately. The binding and unbinding processes of CLK/CYC to an E-box can be formulated as

\[ B + T \xrightleftharpoons{b_{bt}}{u_{bt}} BT, \quad (4.2) \]

where B is an available binding site, i.e., an unbound E-box; T is the TF, CLK/CYC; and BT denotes CLK/CYC bound to the E-box; $b_{bt}$ is the rate of CLK/CYC binding to the E-box and $u_{bt}$ is the rate of CLK/CYC releasing from the E-box. We can get Eq. (4.3) using mass-action kinetics,

\[ \frac{d[BT]}{dt} = [B][T]b_{bt} - [BT]u_{bt}, \quad (4.3) \]

where $[B]$ is the concentration of unbound binding sites, $[T]$ is the concentration of the TFs and $[BT]$ is the concentration of bound binding sites. Suppose the volume of the cell is $V$. The number of B and BT in the cell are $[B]V$ and $[BT]V$. Since the total number of B and BT is $n$ then,

\[ \frac{d[BT]}{dt} = ((n/V) - [BT])[T]b_{bt} - [BT]u_{bt}. \quad (4.4) \]
Let $Pr_{bt}$ be the number of bound binding sites over the total number of binding sites. 

$$[BT] = \frac{\text{total number of binding sites}}{V} \times Pr_{bt}.$$ 

Since the total number of binding sites is $n$, Eq. (4.4) becomes

$$d\left(\frac{(n/V) Pr_{bt}}{Pr_{bt}}\right)/dt = \left(\frac{n}{V} - \frac{n}{V} Pr_{bt}\right)[T] b_{bt} - (n/V) Pr_{bt} u b_{bt},$$

which simplifies to

$$d Pr_{bt}/dt = (1 - Pr_{bt})[T] b_{bt} - Pr_{bt} u b_{bt}. \quad (4.6)$$

Now we can calculate probabilities for CLK/CYC binding to the whole promoter in a gene. Assuming that CLK/CYC can bind independently to any of $n$ binding sites (E-boxes) and if one or more E-boxes are bound, transcription of that gene is activated at a rate $tc_{av}$, otherwise at a deactivated rate $tc_{dynam}$. The probability of none of the E-boxes being bound is $(1 - Pr_{bt})^n$. The rate of transcription would then be

$$tc_{av} (1 - (1 - Pr_{bt})^n) + tc_{dynam} (1 - Pr_{bt})^n. \quad (4.7)$$

### 4.2.3 Modelling competition of PDP1 activation and VRI repression

Here we consider the $clk$ expression activated by PDP1 and repressed by VRI. We first calculate probabilities of VRI and PDP1 binding to the V/P box in the $clk$ promoter. Assuming there is only one binding site B, i.e. the V/P box, in the $clk$ promoter and an activator PDP1, denoted by A, and a repressor VRI, denoted by R, competing to bind that site. We write the reactions as below,

$$B + A \xrightarrow{h_{ba}} BA, \quad (4.8)$$
\[ B + R \xrightarrow{b_{ba}} BR , \] (4.9)

where \( b_{ba} \) is the rate of PDP1 binding to the V/P box, and \( ub_{ba} \) is the rate of PDP1 releasing from the V/P box; \( b_{br} \) is the rate of VRI binding to the V/P box, and \( ub_{br} \) is the rate of VRI releasing from the V/P box. We can get Eqs. (4.10) and (4.11) by using the mass action rate law:

\[ \frac{d}{dt} \left[ \frac{[BA]}{[B][A]} \right] = b_{ba} - [BA] ub_{ba} , \] (4.10)

\[ \frac{d}{dt} \left[ \frac{[BR]}{[B][R]} \right] = b_{br} - [BR] ub_{br} . \] (4.11)

Suppose the volume of the cell is \( V \). The number of B, BA and BR in the cell are \([B]V\), \([BA]V\) and \([BR]V\). As the total number of B, BA and BR is one, we eliminate B from the above two equations and get

\[ \frac{d}{dt} \left[ \frac{[BA]}{1 - [BA] - [BR]} \right] = \frac{b_{ba} - [BA] ub_{ba}}{1 - [BA] - [BR]} , \] (4.12)

\[ \frac{d}{dt} \left[ \frac{[BR]}{1 - [BR] - [BA]} \right] = \frac{b_{br} - [BR] ub_{br}}{1 - [BR] - [BA]} . \] (4.13)

Let \( Pr_{ba} \) and \( Pr_{br} \) be the probabilities of A bound to B and R bound to B. We can write Eqs. (4.12) and (4.13) in the form of probabilities,

\[ \frac{d}{dt} \left[ Pr_{ba} \right] = (1 - Pr_{ba} - Pr_{br}) [A] b_{ba} - Pr_{ba} ub_{ba} , \] (4.14)

\[ \frac{d}{dt} \left[ Pr_{br} \right] = (1 - Pr_{ba} - Pr_{br}) [R] b_{br} - Pr_{br} ub_{br} . \] (4.15)

Assuming that if a PDP1 is bound to the V/P box, transcription of \( clk \) genes occurs at a rate of \( tc_{pc} \); if a VRI is bound to the V/P box, transcription rate is \( tc_{vc} \) and if neither PDP1 nor VRI binds the V/P box, transcription occurs at a deactivated rate \( tc_{dypmt} \). The transcription rate of \( clk \) gene would then be
Knowing the components and their interconnections in the *Drosophila* circadian clock schematised in Figure 4-1, we can convert the conceptual model into a mathematical model. In a deterministic model if we specify the current concentrations of all the components as the state of the system, the time evolution of the system can be described by a system of kinetic equations by using the general principles of biochemical kinetics.

For the current system, the rate constants in the transcription processes are determined by the effective transcription rates as derived in Eqs. (4.7) and (4.16). The other reactions are governed by the mass action rate law, instead of the commonly used Michaelis-Menten kinetics in the previous model (Leloup, Gonze et al. 1999; Ueda, Hagiwara et al. 2001; Smolen, Hardin et al. 2004). The use of the mass action rate law keeps the model simple and the number of parameters low. Moreover, there is no justification whether Michaelis-Menten kinetics are the correct description for these processes as they have not been understood in detail yet. The study by Kurosawa et al. (2002) on a single negative feedback oscillator showed that by introducing Michaelis-Menten type kinetics within the model, its robustness may be enhanced. It will be shown later that simulated oscillations are even more robust to parameter variations in our model. Therefore, the mass action rate law is used in the current model.

The model described by a system of ODEs is outlined below. For clarity, these equations are grouped into four categories. Some of the variable names and rate constants are written in mixed normal and subscript fonts in the equations for better visualisation. The name of mRNAs is written in lower case with a subscript ‘m’ denoting mRNA. The name of proteins and complexes is written in upper case. Abbreviations used for variable names are: PDP for PDP1, CC for CLK/CYC dimer, PT for PER/TIM dimer and CCPT for CLK/CYC/PER/TIM complex. The biochemical meaning of the parameters is explained in Table 4-1.
1. Probabilities of TFs binding to a binding site (an E-box or a V/P box) in promoters:

The binding probabilities defined in the model are CLK/CYC binding to an E-box element in per promoter ($Pr_{per}$), in tim promoter ($Pr_{ct}$), in vri promoter ($Pr_{cv}$), and to pdp1 promoter ($Pr_{pdp}$); VRI binding to the V/P box in clk promoter ($Pr_{pc}$), and PDP1 binding to that in clk promoter ($Pr_{vc}$). Eqs. (4.17) - (4.20) were derived based on Eq. (4.6). Eqs. (4.21) and (4.22) were derived according to Eqs. (4.14) and (4.15), respectively.

\[
d(Pr_{per})/dt = (1 - Pr_{per}) \times bccper_p \times CC - Pr_{per} \times ubccper_p \\
\] (4.17)

\[
d(Pr_{ct})/dt = (1 - Pr_{ct}) \times bccctim_p \times CC - Pr_{ct} \times ubccctim_p \\
\] (4.18)

\[
d(Pr_{pdp})/dt = (1 - Pr_{pdp}) \times bccpdp_p \times CC - Pr_{pdp} \times ubccpdp_p \\
\] (4.19)

\[
d(Pr_{cv})/dt = (1 - Pr_{cv}) \times bccvri_p \times CC - Pr_{cv} \times ubccvri_p \\
\] (4.20)

\[
d(Pr_{pc})/dt = (1 - Pr_{pc} - Pr_{vc}) \times bpdpclk_p \times PDP - Pr_{pc} \times ubpdpclk_p \\
\] (4.21)

\[
d(Pr_{vc})/dt = (1 - Pr_{vc} - Pr_{pc}) \times bvriclek_p \times VRI - Pr_{vc} \times ubvriclek_p \\
\] (4.22)

2. Time evolution of mRNAs of per, tim, clk, vri and pdp1:

The first three terms on the right side of Eq. (4.23) were derived according to Eq. (4.16). The first two terms in that of Eqs. (4.24) - (4.27) were derived according to Eq. (4.7). All these terms describe transcriptional processes; the last term in these equations describes degradation processes of mRNAs.
\[
d(\text{clk}_m)/dt = (\text{Pr}_c \times \text{tevriclek}_p + \text{Pr}_c \times \text{tcepdpckl}_p + (1 - \text{Pr}_c - \text{Pr}_p) \times \text{tceclk}_p) \\
\times \text{clk}_p - \text{dclkm} \times \text{clk}_m \\
(4.23)
\]

\[
d(\text{per}_m)/dt = ((1 - (1 - \text{Pr}_c))^{\text{app}} \times \text{tccceper}_p + (1 - \text{Pr}_c)^{\text{app}} \times \text{tcevpm}) \times \text{per}_p \\
- \text{dperm} \times \text{per}_m \\
(4.24)
\]

\[
d(\text{tim}_m)/dt = ((1 - (1 - \text{Pr}_c)^{\text{app}} \times \text{tccctim}_p + (1 - \text{Pr}_c)^{\text{app}} \times \text{tcevpm}) \times \text{tim}_p \\
- \text{dtimm} \times \text{tim}_m \\
(4.25)
\]

\[
d(\text{vri}_m)/dt = ((1 - (1 - \text{Pr}_c)^{\text{app}} \times \text{tcevri}_p + (1 - \text{Pr}_c)^{\text{app}} \times \text{tcevpm}) \times \text{vri}_p \\
- \text{dvrim} \times \text{vri}_m \\
(4.26)
\]

\[
d(\text{pdp}_m)/dt = ((1 - (1 - \text{Pr}_c)^{\text{app}} \times \text{tcccpdp}_p + (1 - \text{Pr}_c)^{\text{app}} \times \text{tcevpm}) \\
\times \text{pdp}_p - \text{dpdpm} \times \text{pdp}_m \\
(4.27)
\]

3. Time evolution of PER, TIM, CLK, VRI and PDP1 proteins:

The first term on the right side of Eqs. (4.28) - (4.32) expresses transcriptional processes, and the last term expresses degradation processes of proteins. The second term in Eqs. (4.28) - (4.30) denotes association of complexes, and the third term denotes dissociation of complexes.

\[
d(\text{PER})/dt = \text{tlper} \times \text{per}_m - \text{bpt} \times \text{PER} \times \text{TIM} + \text{ubpt} \times \text{PT} - \text{dper} \times \text{PER} \\
(4.28)
\]

\[
d(\text{TIM})/dt = \text{tltim} \times \text{tim}_m - \text{bpt} \times \text{PER} \times \text{TIM} + \text{ubpt} \times \text{PT} - \text{dtim} \times \text{TIM} \\
(4.29)
\]

\[
d(\text{CLK})/dt = \text{tclkl} \times \text{clk}_m - \text{bcl} \times \text{CLK} \times \text{CYC} + \text{ubcl} \times \text{CC} - \text{dcll} \times \text{CLK} \\
(4.30)
\]
\[ \frac{d(VR)}{dt} = tlvri \times vri_m - dvri \times VR \]  \hspace{1cm} (4.31)

\[ \frac{d(PDP)}{dt} = tlpdp \times pdp_m - dpdp \times PDP \]  \hspace{1cm} (4.32)

4. PER/TIM, CLK/CYC and PER/TIM/CLK/CYC complexes:

The first and second terms in Eqs. (4.33) - (4.35) describe the association and dissociation of PT, CC and CCPT complexes, respectively, and the last term describes the degradation processes of these complexes. The third and fourth terms in Eqs. (4.33) and (4.34) denote association and dissociation of CCPT complex.

\[ \frac{d(PT)}{dt} = bpt \times PER \times TIM - ubpt \times PT - bccpt \times PT \times CC \\
+ ubccpt \times CCPT - dpt \times PT \]  \hspace{1cm} (4.33)

\[ \frac{d(CC)}{dt} = bcc \times CLK \times CYC - ubcc \times CC - bccpt \times PT \times CC \\
+ ubccpt \times CCPT - dcc \times CC \]  \hspace{1cm} (4.34)

\[ \frac{d(CCPT)}{dt} = bccpt \times PT \times CC - ubccpt \times CCPT - dccpt \times CCPT \]  \hspace{1cm} (4.35)

In summary, the circadian clock model is described by a system of 19 ODEs with 47 parameters.
Table 4-1 Biochemical meaning of the parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Biochemical meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>bccpdpp</td>
<td>binding rate of CLK/CYC to an E-box in pdp1 promoter</td>
</tr>
<tr>
<td>bccperp</td>
<td>binding rate of CLK/CYC to an E-box in per promoter</td>
</tr>
<tr>
<td>bccctimp</td>
<td>binding rate of CLK/CYC to an E-box in tim promoter</td>
</tr>
<tr>
<td>bccvrip</td>
<td>binding rate of CLK/CYC to an E-box in vri promoter</td>
</tr>
<tr>
<td>bpdpcikp</td>
<td>binding rate of PDP1 to the V/P box in clk promoter</td>
</tr>
<tr>
<td>bvrclkp</td>
<td>binding rate of VRI to the V/P box in clk promoter</td>
</tr>
<tr>
<td>ubccpdpp</td>
<td>unbinding rate of CLK/CYC to an E-box in pdp1 promoter</td>
</tr>
<tr>
<td>ubccperp</td>
<td>unbinding rate of CLK/CYC to an E-box in per promoter</td>
</tr>
<tr>
<td>ubccctimp</td>
<td>unbinding rate of CLK/CYC to an E-box in tim promoter</td>
</tr>
<tr>
<td>ubccvrip</td>
<td>unbinding rate of CLK/CYC to an E-box in vri promoter</td>
</tr>
<tr>
<td>ubpdpcikp</td>
<td>unbinding rate of PDP1 to the V/P box in clk promoter</td>
</tr>
<tr>
<td>ubvrclkp</td>
<td>unbinding rate of VRI to the V/P box in clk promoter</td>
</tr>
<tr>
<td>bcc</td>
<td>association rate of CLK/CYC dimer</td>
</tr>
<tr>
<td>bpt</td>
<td>association rate of PER/TIM dimer</td>
</tr>
<tr>
<td>bccpt</td>
<td>association rate of CLK/CYC/PER/TIM complex</td>
</tr>
<tr>
<td>ubcc</td>
<td>dissociation rate of CLK/CYC dimer</td>
</tr>
<tr>
<td>ubpt</td>
<td>dissociation rate of PER/TIM dimer</td>
</tr>
<tr>
<td>ubccpt</td>
<td>dissociation rate of CLK/CYC/PER/TIM complex</td>
</tr>
<tr>
<td>tcccdpdpp</td>
<td>transcription rate of CLK/CYC-activated pdp1 gene</td>
</tr>
<tr>
<td>tccperp</td>
<td>transcription rate of CLK/CYC-activated per gene</td>
</tr>
<tr>
<td>tccctimp</td>
<td>transcription rate of CLK/CYC-activated tim gene</td>
</tr>
<tr>
<td>tccvrip</td>
<td>transcription rate of CLK/CYC-activated vri gene</td>
</tr>
<tr>
<td>tcpdpcikp</td>
<td>transcription rate of PDP1-activated clk gene</td>
</tr>
<tr>
<td>tcvrclkp</td>
<td>transcription rate of VRI-repressed clk gene</td>
</tr>
<tr>
<td>tccclk</td>
<td>transcription rate of clk gene binding neither PDP1 nor VRI</td>
</tr>
<tr>
<td>tcdvpmt</td>
<td>transcription rate of deactivated per, tim, vri or pdp1 gene</td>
</tr>
<tr>
<td>tlcik</td>
<td>translation rate of clk mRNA</td>
</tr>
<tr>
<td>ttpdp</td>
<td>translation rate of pdp1 mRNA</td>
</tr>
<tr>
<td>ttperr</td>
<td>translation rate of per mRNA</td>
</tr>
<tr>
<td>tltim</td>
<td>translation rate of tim mRNA</td>
</tr>
<tr>
<td>tlvri</td>
<td>translation rate of vri mRNA</td>
</tr>
<tr>
<td>dclkm</td>
<td>degradation rate of clk mRNA</td>
</tr>
<tr>
<td>dpdpdm</td>
<td>degradation rate of pdp1 mRNA</td>
</tr>
<tr>
<td>dperrm</td>
<td>degradation rate of per mRNA</td>
</tr>
<tr>
<td>dtimmm</td>
<td>degradation rate of tim mRNA</td>
</tr>
<tr>
<td>dvrim</td>
<td>degradation rate of vri mRNA</td>
</tr>
<tr>
<td>dclk</td>
<td>degradation rate of CLK protein</td>
</tr>
<tr>
<td>dpdp</td>
<td>degradation rate of PDP1 protein</td>
</tr>
<tr>
<td>dper</td>
<td>degradation rate of PER protein</td>
</tr>
<tr>
<td>dtim</td>
<td>degradation rate of TIM protein</td>
</tr>
<tr>
<td>dvri</td>
<td>degradation rate of VRI protein</td>
</tr>
<tr>
<td>dpt</td>
<td>degradation rate of PER/TIM dimer</td>
</tr>
<tr>
<td>dcc</td>
<td>degradation rate of CLK/CYC dimer</td>
</tr>
<tr>
<td>dccpt</td>
<td>degradation rate of CLK/CYC/PER/TIM complex</td>
</tr>
<tr>
<td>npt</td>
<td>number of E-boxes in per or tim promoter</td>
</tr>
<tr>
<td>nvri</td>
<td>number of E-boxes in vri promoter</td>
</tr>
<tr>
<td>npdp</td>
<td>number of E-boxes in pdp1 promoter</td>
</tr>
</tbody>
</table>
Chapter 5: Computational implementation of the model and parameter estimation

This chapter describes the implementation of the proposed mathematical model into a computer solvable system. In Section 1, a SBML file is derived from the mathematical model using CellDesigner. In Section 2, parameters are estimated using COPASI and, finally, in Section 3 the initial conditions required by simulations are given.

5.1 Conversion to SBML by CellDesigner

After defining the system of ODEs for the biochemical reaction network model, these ODEs should be converted to a computer understandable form in order to solve it once the parameter values for the rate constants are given. Commonly, a model is translated into a general purpose simulation programming language, such as Mathematica or Matlab, which is widely used in the systems biology community. There are also many software tools available for systems biologists to solve many specific problems, such as parameter estimation, bifurcation analysis and sensitivity analysis. Since each tool was designed for some specific tasks, it is important to reuse models between different model-building tools for information integration and resource utilisation. The most common format shared by software tools to store and exchange data in the systems biology community is the Extensible Markup Language (XML) since such a format is compatible with nearly all computational operating systems. Among several XML-based Markup languages, Systems Biology Markup Language (SBML) is the most prominent one (Hucka, Finney et al. 2003; Hucka, Finney et al. 2004).

In a SBML model, the focus is systems of biochemical reactions. Models consisting of biochemical species are linked by reactions to form a biochemical network. However, SBML does not interoperate the reactions into a set of differential equations or other specific representation of the network, such as a discrete stochastic representation. Only a software package reading SBML model interprets it into its own internal representation. The advantage of using SBML to represent a biochemical network
model is that we can fully utilise the capability of each software package without transformation of the format. Up to Feb, 2007, there are over 110 software packages supporting SBML with each having specific capabilities (http://sbml.org, accessed on 01, Feb, 2007).

In this study, a SBML creation tool, CellDesigner, was used to build a SBML model for the circadian clock network (Funahashi 2003). The current version is 4.0 which can be downloaded free from http://www.celldesigner.org/. The following procedures were performed to create the SBML model.

1. Create compartment

   All species in the system have to be located in a single or multiple compartments in a SBML model. No nucleus or cytoplasm compartment was assumed for the current model, therefore a compartment, named “WholeCell”, was created.

2. Define species

   In SBML, a species is defined as one type of molecules in a system. As shown previously in Figure 4-1, 19 species were input into the model. In addition, six variables which define the probabilities of TFs binding to a binding site (E-box or V/P box), as in Eqs. (4.17) – (4.22), were also treated as species. Finally, the product of degradation processes was defined as a species, named “EmptySet”. In total, 26 species were identified in the SBML model.

3. Create reactions and kinetic laws

   Although SBML allows for both irreversible and reversible reactions, in CellDesigner, reversible reactions have to be broken into forward and backward irreversible reactions. Therefore, the ODEs defined by Eqs. (4.17) – (4.35) were broken into individual chemical reaction equations as in Appendix C, where the kinetic law for each reaction was also given.

4. Specify the initial values of species and parameters
To simulate the time evolutions of the species, the initial condition of the model and parameter values need to be specified. For the current model, no values are available from the in vitro experiments and so they have to be obtained by some estimation techniques, which are described in the following section.

5.2 Parameter estimation

In typical parameter estimation problems, a model contains a set of unknown parameters. The goal of parameter estimation is to find parameter values which give the model the best goodness of fit with given measured data. This is sometimes also called the inverse problem. For kinetic modelling in biological networks, particularly in GRNs, numerical values of many kinetic parameters are lacking. Parameter estimation is, therefore, highly important.

For the current model, although some information is available about the relationship between the transcription, translation and degradation of mRNAs and proteins rates in the circadian clock in plants (Shu and Hong-Hui 2004), no quantities of the rates have been examined. Even if quantitative information was available, its usefulness would be limited since these quantities are inevitably influenced by experimental factors, such as experimental setting, cell types and states of cells. Because all the 44 kinetic parameters in the current model are unknown and these parameters are involved in non-linear ODEs, it makes the parameter estimation a very difficult task. The following procedures were performed to find a set of parameters which reproduced the target experimental dataset.

First, a set of parameters was estimated as initial parameters which can produce roughly 24 h oscillations. To find a large number of parameters for the solution of a set of non-linear ODEs, a common practice among modellers is to use trial-and-error (Zwolak, Tyson et al. 2005). We, therefore, set the model to produce roughly 24 h oscillations of mRNAs and proteins by trial-and-error guesstimation. This required much time and patience. Although, initially, a number of numerical optimisation algorithms had been tried to find out the first approximation to parameter values, it failed to do so because of
Next, the initial parameters were fine tuned by comparing the simulation output to an experimental dataset. The experimental dataset (time courses of mRNA and protein concentrations) was produced according to some criteria. The criteria for finding precise parameters were that the model should produce sustained circadian oscillations of mRNAs and proteins, correctly measure phase relationships between gene expression and proteins, and have appropriate time delays between mRNAs and proteins in the condition of constant darkness.

Because concentrations of ‘clock’ mRNAs and proteins in the cell are not known and only relative concentration abundance was measured, it was assumed that around 1000 protein molecules and 100 mRNA molecules are in a cell, according to a previous theoretical model (Vilar, Kueh et al. 2002). The number of molecules corresponds to protein concentrations of 3 – 4 nM and mRNA concentrations of 0.3 – 0.4 nM according to the following calculation: A radius of a lateral neuron in *Drosophila* is about 5 – 6 µm (Ewer, Frisch et al. 1992), and so 5 µm is taken in our model. The volume of the cell is

\[ V = \frac{4}{3} \pi r^3 = 5.23 \times 10^{-13} L. \]

Therefore, the number of molecules that corresponds to 1 nM is

\[ 1 \text{nM} = (5.23 \times 10^{13} L)(10^9 \text{ mole} / L)(6 \times 10^{23} \text{ molecules} / \text{mole}) \approx 314 \text{ molecules}. \]  

To fine tune the parameters according to the experimental dataset, there are many existing tools. One such tool, called COPASI, was used in this study (Hoops, Sahle et al. 2006). As COPASI reads a SBML file, the SBML model made by CellDesigner can be readily reused for COPASI. Here, a brief explanation is given about how the goodness of fit is calculated and how the parameter estimation algorithms work in COPASI.
To express the goodness of fit of the model to a given experimental dataset, an objective function based on “a weighted sum of squares” (WSS) is used in COPASI (personal communication with Dr. Stefan Hoops, one of COPASI’s developers, in June, 2006). This sum is taken over all measured data points between model predictions and experimental observations. The weight is the inverse of the standard deviation, so that points with low variance are given greater value. The difference between the measured data and the computed dynamic profiles becomes a least squared error and is expressed in the form

\[
WSS = \sum_{i=1}^{n} \frac{(y_i^{ex} - y_i^{mod})^2}{\sigma(y_i^{ex})},
\]

(5.2)

where the sum is over all the experimental data, \(y_i^{ex}\) is the measured value or experimental value of the variables, \(y_i^{mod}\) is the simulated value of the variables. The smaller WSS, the better the fit.

To minimise WSS, COPASI uses a number of algorithms, including global and local algorithms, that were all tested in this study. Although none of these algorithms guaranteed global optimality, the confidence of finding “optimal” parameters can be increased. Generally, the Levenberg-Marquardt (LM) algorithm is the best currently known deterministic local optimisation algorithm for nonlinear least-squares problems (Zwolak, Tyson et al. 2005). It is particularly suitable in least squares problems based on the sum of squares functions (Mendes and Kell 1998). In this study, it was also found that several algorithms converged to the same solution and the LM algorithm always converged quickly to the solution. Because a detailed description of the LM algorithm is beyond the scope of the thesis the interested reader is referred to Kelley (1999) and Nocedal and Wright (1999) for a more comprehensive treatment.

The parameter estimation process is summarised in Figure 5-1. The parameters obtained in the solution are shown in Table 5-1. Time is in units of hour. Concentrations are referenced to the total cell volume, and they are in units of nM.
Experimental data

Model (ODEs)

Optimisation (e.g. LM)

\[ y_{i}^{\text{ex}} \]

\[ y_{i}^{\text{mod}} \]

Goodness of fit of the model (e.g., WSS)

Criterion

Result

Adjust model parameters

Unsatisfied

Figure 5-1 An overview of the parameter estimation process.
Table 5-1 Parameters of the model: The units of binding rates and association rates are nM$^{-1}$h$^{-1}$ and the units of the other parameters are h$^{-1}$.

<table>
<thead>
<tr>
<th>Index</th>
<th>Parameter</th>
<th>Value</th>
<th>Index</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>bccpdpp</td>
<td>0.062</td>
<td>25</td>
<td>tccldkp</td>
<td>1.42</td>
</tr>
<tr>
<td>2</td>
<td>bccperp</td>
<td>0.069</td>
<td>26</td>
<td>tcdvpt</td>
<td>0.053</td>
</tr>
<tr>
<td>3</td>
<td>bccctimp</td>
<td>0.069</td>
<td>27</td>
<td>tclkl</td>
<td>35</td>
</tr>
<tr>
<td>4</td>
<td>bccvrip</td>
<td>0.1</td>
<td>28</td>
<td>tlpdp</td>
<td>1.87</td>
</tr>
<tr>
<td>5</td>
<td>bpdpcclkp</td>
<td>1.155</td>
<td>29</td>
<td>tlper</td>
<td>36</td>
</tr>
<tr>
<td>6</td>
<td>bvriclekp</td>
<td>1.858</td>
<td>30</td>
<td>tltim</td>
<td>36</td>
</tr>
<tr>
<td>7</td>
<td>ubccpdpp</td>
<td>0.145</td>
<td>31</td>
<td>tvri</td>
<td>14.68</td>
</tr>
<tr>
<td>8</td>
<td>ubccperp</td>
<td>0.262</td>
<td>32</td>
<td>dclkm</td>
<td>0.643</td>
</tr>
<tr>
<td>9</td>
<td>ubccctimp</td>
<td>0.262</td>
<td>33</td>
<td>dpdpm</td>
<td>0.06</td>
</tr>
<tr>
<td>10</td>
<td>ubccvrip</td>
<td>0.276</td>
<td>34</td>
<td>dperm</td>
<td>0.053</td>
</tr>
<tr>
<td>11</td>
<td>ubpdpclkp</td>
<td>0.952</td>
<td>35</td>
<td>dtimm</td>
<td>0.053</td>
</tr>
<tr>
<td>12</td>
<td>ubvrieclkp</td>
<td>1.043</td>
<td>36</td>
<td>dvri</td>
<td>0.07</td>
</tr>
<tr>
<td>13</td>
<td>bcc</td>
<td>2.349</td>
<td>37</td>
<td>dclk</td>
<td>0.2</td>
</tr>
<tr>
<td>14</td>
<td>bpt</td>
<td>1.1</td>
<td>38</td>
<td>dpdp</td>
<td>0.156</td>
</tr>
<tr>
<td>15</td>
<td>bccpt</td>
<td>51</td>
<td>39</td>
<td>dper</td>
<td>0.62</td>
</tr>
<tr>
<td>16</td>
<td>ubcc</td>
<td>0.89</td>
<td>40</td>
<td>dtim</td>
<td>0.62</td>
</tr>
<tr>
<td>17</td>
<td>ubpt</td>
<td>2.93</td>
<td>41</td>
<td>dvri</td>
<td>1.226</td>
</tr>
<tr>
<td>18</td>
<td>ubccpt</td>
<td>7.89</td>
<td>42</td>
<td>dpt</td>
<td>0.279</td>
</tr>
<tr>
<td>19</td>
<td>teccpdpp</td>
<td>9.831</td>
<td>43</td>
<td>dce</td>
<td>0.184</td>
</tr>
<tr>
<td>20</td>
<td>.teccperp</td>
<td>11</td>
<td>44</td>
<td>decept</td>
<td>15.122</td>
</tr>
<tr>
<td>21</td>
<td>teccctimp</td>
<td>11</td>
<td>45</td>
<td>npt</td>
<td>5</td>
</tr>
<tr>
<td>22</td>
<td>teccvrip</td>
<td>16.86</td>
<td>46</td>
<td>nvr</td>
<td>4</td>
</tr>
<tr>
<td>23</td>
<td>tcpdpcclkp</td>
<td>125.54</td>
<td>47</td>
<td>npdp</td>
<td>6</td>
</tr>
<tr>
<td>24</td>
<td>tcvrieclkp</td>
<td>0.028</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.3 Initial conditions

Because the system governed by Eqs.(4.17) to (4.35) can maintain sustained, periodic oscillations using the set of parameters in Table 5-1 regardless of initial conditions, the initial conditions have no influence on the final state of the system. However, to eliminate the transient dynamics from the initial state to the stable oscillation state, the initial conditions used are listed in Table 5-2. The concentrations for each gene are constant. Additionally, as explained previously, the concentration of CYC was also assumed to be constant.
Table 5-2 Initial conditions. Abbreviations: CC – CLK/CYC, PT – PER/TIM, CCPT – CLK/CYC/PER/TIM. Constant values in the system are denoted by *.

<table>
<thead>
<tr>
<th>Species</th>
<th>Concentration (nM)</th>
<th>Species</th>
<th>Concentration (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>0.5566</td>
<td>clkp*</td>
<td>0.003185</td>
</tr>
<tr>
<td>CCPT</td>
<td>0.4982</td>
<td>pdpp*</td>
<td>0.003185</td>
</tr>
<tr>
<td>CLK</td>
<td>3.6628</td>
<td>perp*</td>
<td>0.003185</td>
</tr>
<tr>
<td>clkm</td>
<td>0.2583</td>
<td>timp*</td>
<td>0.003185</td>
</tr>
<tr>
<td>PDP</td>
<td>4.1953</td>
<td>vrip*</td>
<td>0.003185</td>
</tr>
<tr>
<td>pdpm</td>
<td>0.3175</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PER</td>
<td>2.7527</td>
<td></td>
<td></td>
</tr>
<tr>
<td>perm</td>
<td>0.2395</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PT</td>
<td>0.4014</td>
<td>prcpdp</td>
<td>0.08</td>
</tr>
<tr>
<td>TIM</td>
<td>2.7527</td>
<td>preper</td>
<td>0.0431</td>
</tr>
<tr>
<td>timm</td>
<td>0.2395</td>
<td>prct</td>
<td>0.043</td>
</tr>
<tr>
<td>VRI</td>
<td>3.175</td>
<td>prcv</td>
<td>0.0585</td>
</tr>
<tr>
<td>vrim</td>
<td>0.2571</td>
<td>prpc</td>
<td>0.426</td>
</tr>
<tr>
<td>CYC*</td>
<td>1.00</td>
<td>prvc</td>
<td>0.489</td>
</tr>
</tbody>
</table>
Chapter 6: Simulation results and discussion of the circadian clock

In this chapter, a range of \textit{in silico} experiments is tested, including simulations of the circadian rhythms under the condition of constant darkness, the robustness of the system, and the response of the system to light and to a number of mutants. All the \textit{in silico} experimental data were thoroughly and rigorously compared with the \textit{in vitro} experimental data. In addition, a possible function of VRI and PDP1 feedback loops is proposed by theoretically studying this model via sensitivity analysis. In Section 2, a discussion of the results and a comparison with some previous models are given.

6.1 Simulations results

6.1.1 Circadian oscillations in constant darkness

For simulations under the condition of constant darkness, the parameters did not change over the course of time. The numerical solution of the model showed sustained oscillations with 24 h period in the concentrations of \textit{per}, \textit{tim}, \textit{vri}, \textit{pdp1} and \textit{clk} mRNAs and their corresponding proteins using the standard parameter set, as given in Table 5-1.

Oscillations in mRNA concentrations from the simulation are plotted in Figure 6-1A. The oscillations of \textit{per} and \textit{tim} mRNAs were in phase and their levels peaked at CT12\textsuperscript{†}. The oscillation of \textit{clk} mRNA was in anti-phase with \textit{per} and \textit{tim} mRNAs; it peaked at CT3 and subsequently bottoms in CT13.5. These results are consistent with observations that \textit{per} and \textit{tim} mRNA levels oscillate in phase with one another and they reach peak levels early in the evening at circadian time (CT12 – CT16) (Hardin, Hall et al. 1990); \textit{clk} mRNA levels oscillate in anti-phase to \textit{per} and \textit{tim} mRNA levels and \textit{clk}

\footnote{Circadian time (CT): A standardised 24-hour notation of the phase in a circadian cycle that represents an estimation of the organism’s subjective time. CT 0 indicates the beginning of a subjective day, and CT 12 is the beginning of a subjective night.}
mRNA levels peak from late at night to early in the morning (CT23 – CT4) (Bae, Lee et al. 1998); the simulated concentration of vri mRNA peaked at CT11.5 and that of pdp1 mRNA reached maximum at CT13.5 with a 2 h delay. This agrees with experimental data that vri mRNA oscillates in anti-phase with clk mRNA, and pdp1 mRNA oscillates with a similar phase to vri mRNA after several hours delay (Cyran, Buchsbaum et al. 2003; Glossop, Houl et al. 2003).

Figure 6-1B illustrates the oscillations in concentrations of the proteins. On the one hand, the peaks of PER and TIM concentrations were at CT15, the peak of VRI concentration was at CT12 and that of PDP1 was at CT18. On the other hand, the concentration of CLK peaked at CT4.5 and bottomed at CT14.5. In vitro experimental data showed that protein levels of PER and TIM are at their highest in the middle of the night with four to six hours delay from their mRNA peaks (Zeng, Qian et al. 1996), and a lag of 3 h – 6 h exists between the rise of VRI and that of PDP1 (Cyran, Buchsbaum et al. 2003). The concentration of CLK peaked at CT4.5 and bottomed at CT14.5. The phase of the maximum and minimum from the simulated results are all in good agreement with the experimental observations.

6.1.2 Robustness to parameter variations

The circadian clock is known to have the ability to regulate the phase relationships of different physiological processes in a daily cycle. Normally it should maintain circadian rhythms with a period close to 24 hours regardless of parameter variations. It has been reported that there was only 0.1 hours variation from the mean value of 24.3 hours for WT (wild type) flies (Levine, Funes et al. 2002). In another report, 0.06 hours variation from the mean value was found under the temperature of 29°C, and 0.1 and 0.2 hours variations are under 20°C and 25°C, respectively (Bao, Rihel et al. 2001). It should be noted that under 20°C and 25°C, there were non-negligible percentages of flies appearing arrhythmic (4/15 and 3/20, respectively). Therefore, in a model of the circadian clock, modest parameter variations should only result in minor period changes given that arrhythmic flies are not considered.
Figure 6-1 Sustained oscillations for the concentrations of the mRNAs and the proteins: (A) Oscillations for the mRNAs and (B) oscillations for the proteins. The time scale of clk in (A) and PDP1 in (B) has been enlarged to allow better visualisation.
As there are 44 parameters to be tested (not including three parameters for the number of the E-boxes in promoters, which are actually related to the model structure), it is not possible to explore the behaviour of the system in its full dimensional parameter spaces. To investigate the behaviour of the system to parameter variations, we followed the methods given in previous models (Lema, Golombek et al. 2000; Leloup and Goldbeter 2003; Smolen, Hardin et al. 2004). One parameter was changed at a time while keeping the others at their standard values. Perturbations were simulated by increasing or decreasing 20% from the standard value for each individual parameter.

Oscillations were preserved in all the simulations. From Figure 6-2 we can see that the periods vary less than 0.8 hours from the control values of 24 hours with 20% perturbation to each parameter. The largest period increase (+0.75 h) was caused by the increases in binding rate of PDP1 to the clk promoter. And the largest two period decreases were very close (-0.8 h), which were caused by decreases in binding rates of CLK/CYC to pdp1 and per promoters. In a direct comparison with a previous model using Michaelis-Menten kinetics (Smolen, Hardin et al. 2004), where three parameter with 20% changes yielded periods differing > 3 hours from 24 hours, this model showed lower period variations to the same parameter disparities. However, we note this comparison of 20% changes in parameter values with what was done in the model of Smolen may not be easy as implied. This is because the parameters in Smolen’s model are from Michaelis-Menten kinetics which correspond to rational functions of the rate constants used here. To directly compare the robustness property of two different model structures as the current and Smolen’s model, we point out that a model using Michaelis-Menten kinetics needs to be done for the future work. This issue will be addressed in the final chapter.

### 6.1.3 Response of the circadian clock to light

On the one hand, the circadian clocks are robust to parameter variations, on the other hand, a fundamental characteristic of the circadian clocks is that they are also entrained (phase-adjusted) by Zeitgeber (Zeitgeber means “time giver”, it gives environmental time cue). This entrain-ability gives the circadian systems a proper phase in synchrony with the outside world. Although both ambient light and temperature cycles on a daily
Figure 6-2 Period variations of the circadian oscillations in respect to parameter variations, one parameter was increased or decreased by 20% each time while the other parameters were kept at the basal values. The most sensitive parameters are indicated. Parameter names corresponding to the parameter index are denoted in Table 5-1. The number of the E-boxes or V/P box is related to the structure of the model, and therefore these there parameters are not included in this figure.
basis, light is often thought to be the predominant Zeitgeber. Here we test the entrainability of our circadian clock model in response to light.

Entrainment by light is generally considered to occur through changing particular parameters in the circadian clocks. In *Drosophila*, it has been shown experimentally that light enhances degradation of TIM and, consequently, degradation of TIM in the light alters the level of other clock components and thus resets the phase of an oscillator (Zeng, Qian et al. 1996). In terms of modelling, increase in TIM degradation rate has been used to model light response and entrainment to light dark cycles in some previous models (Leloup and Goldbeter 1998; Tyson, Hong et al. 1999; Leloup and Goldbeter 2003). As TIM stabilises PER in the cytoplasm, the indirect effect of light is to regulate the localisation of PER and, in turn, to decrease the PER level in the nucleus. Therefore, change in degradation rate of PER has also been used in some models (Schepers, Klinkenberg et al. 1999; Lema, Golombek et al. 2000; Smolen, Hardin et al. 2004). Indeed experimental findings have shown that *tim* \(^{01}\) mutant induced an absence of TIM leads to a substantial lowering of PER abundance (Vosshall, Price et al. 1994; Price, Dembinska et al. 1995), an effect that happens to be similar to the result of exposing flies to constant light (Zerr, Hall et al. 1990; Price, Dembinska et al. 1995). Because we did not include the detailed translocation mechanisms of PER and TIM into the nucleus, as well as the associated Sgg-dependent TIM phosphorylation and CK2-dependent PER phosphorylation processes in the model (Shafer, Rosbash et al. 2002), we simulated the effect of light by increasing the degradation rates of both TIM and PER. Consequently, a new degradation rate, \(k_{\text{light}}\), replaced the degradation rates of TIM and PER in the condition of darkness, indicated as ‘dtim’ and ‘dper’ in Table 5-1.

To model entrainment to light-dark cycles, we used a higher value of \(k_{\text{light}}\) (> 0.62, where 0.62 is the value of the degradation rates of PER and TIM in darkness) in the light phase at Zeitgeber time (ZT) 0 – ZT12 \(^{\ddagger}\), and restored its original value (0.62) during the dark phase ZT12 – ZT24. The value of \(k_{\text{light}}\) during the light phase was arbitrarily chosen. Figure 6-3A shows that oscillations in all proteins were maintained during entrainment by light dark cycles with a \(k_{\text{light}}\) value of 0.8. The phase and

\(^{\ddagger}\) \(\ddagger\) Zeitgeber time (ZT): A standardised 24-hour notation of the phase in an entrained circadian cycle in which ZT 0 indicates the beginning of day, or the light phase, and ZT 12 is the beginning of night, or the dark phase. For comparison, see circadian time on the page 72.
anti-phase relationship between mRNAs and proteins were also maintained under the condition of constant darkness. Simulations have shown that the phase changes were dependent upon the magnitude of k_{light} during light exposure. For the ease of comparison of phase changes, CLK concentrations were plotted using different k_{light} values in Figure 6-3B, which shows the phases were delayed for several hours depending on the different k_{light} values chosen.

As shown experimentally (Figure 6-4A), disappearance of the rhythmicity in flies in constant light (LL) can also be simulated by holding k_{light} at a high constant value (Qiu and Hardin 1996). It was found that the oscillations were damped in LL when the k_{light} value was close to or more than five. The damped protein oscillations were plotted using a k_{light} value of five in Figure 6-4B.

Next, we investigated the oscillatory behaviour of the clock model under influence of a light pulse. *In vitro* experiments, typically, a few minutes of stimuli are delivered when observing phase shifts induced by light pulses (Hall 2003). For example, one minute and 15 minutes of light pulses were applied in the experiments carried out by Hall et al. (1987) and Matsumoto et al. (1994), respectively. However, the duration of magnitude of the biochemical changes caused by the light pulses was different from the actual duration of light pulses applied. In a previous computational model, Leloup et al. (1999) proposed that even if the pulses were short, the consequent parameter changes were much longer because light could turn on a gene and induce the synthesis of an enzyme that may take effects for hours. In *in vitro* experiments, it has been shown that the concentration of PER and TIM remained at trough levels for about several hours following the premature disappearance of these two proteins by light pulses in the late night (Sidote, Majercak et al. 1998). In some previous models, 1 – 4 h durations were used to simulate the effects of a light pulse (Leloup, Gonze et al. 1999; Smolen, Hardin et al. 2004). In our model, phase responses were simulated by applying a 2 h duration light pulse to the system at different time points during the free-running conditions of constant darkness.

The phase shifts were determined from the difference in the maximum values of a specified protein between the free-running system and the perturbed system. Because all
Figure 6-3 (A). Entrainment by light dark cycles. $k_{\text{light}}$ is increased (0.8) during the light phase and remains at the original value (0.62) during the dark phase. Simulation was done with ZT0 lights on, ZT12 lights off. (B). The phases of oscillations after entrainment depended on the different values of $k_{\text{light}}$. We plotted the 6th cycle after the cycles were stable to eliminate the transient effect of light.
Figure 6-4 (A) *In vitro* experiments showed that rhythmicity of *per* mRNA disappeared under the condition of constant light for three days, replotted from Qiu and Hardin (1996). (B) Computational simulations showed that rhythmicity disappeared under constant light condition when $k_{\text{light}} > 5$. A $k_{\text{light}}$ value of 5 was used to produce this figure.
the proteins oscillate with a same period, the choice of protein should not make any difference to the phase shifts. The phase shifts were measured after the transient effect of the light pulse was over. This procedure was applied 24 times by increasing one hour on the time point of application of light pulse each time. The phase response curve (PRC) was determined by plotting the phase shifts as a function of the circadian time at which perturbation was applied. We defined CT0 – CT12 as subjective day and CT12 – CT24 as subjective night.

Similar to the simulations of entrainment by light dark cycles, the effect of the light pulse was simulated by replacing degradation rates of PER and TIM by $k_{\text{light}}$. Simulations showed that the magnitude of the phase shifts varied depending on the value of $k_{\text{light}}$. The bigger the value, the more significant the phase shifts obtained. The best fit with the experimental PRC was obtained by using $k_{\text{light}}=1.3$, as plotted in Figure 6-5. Like the PRC plot in Smolen et al. (2004), the mean value of the PRC obtained by Konopka et al. (1991) was also plotted for comparison.

From Figure 6-5, it is shown that the simulated data showed a consistent 5 h time lag from the experimental data. To clearly compare the actual values of the theoretical PRC from our model with the PRC obtained from the in vitro experiment, we shifted the simulated PRC by advancing it by 5 h. Now we saw a good agreement between the shifted PRC and the experimental PRC. Light pulses delayed the phase of the circadian rhythms during early subjective night and advanced the phase during late subjective night. The shifted PRC showed a dead zone (the area where the phase shift is zero) at the middle of subjective day (CT5 – 9) whereas (CT6 – 10) was shown in the experimental PRC. The crossovers from the advance shifts to the delay shifts were at CT18.5 for the shifted data and at CT18 for the experimental data. The reason for a 5 h lag between the theoretical (simulated) PRC and the experimental PRC could be that we did not include phosphorylation of PER and TIM or the separate nuclear and cytoplasmic compartments in the model. However, phosphorylation and nuclear entry of PER and TIM provided an important time delay between cytoplasmic PER and TIM and nuclear PER/TIM. This time delay also implied a time lag between the effects of light and the repression of CLK/CYC by PER/TIM which is unpresented in the current model. We, however, need to point out that it is only one hypothesis that the phase difference between the simulated and experimental PRCs was introduced by the non-
Figure 6-5 Phase response curve (PRC) obtained by using $k_{\text{light}} = 1.3$. The x-axis represents the time of onset of each light pulse, and on the y-axis positive values represent phase advances and the negative values represent phase delays. The means of experimental values for phase shifts from Konopka (1991) are denoted by diamonds. The simulated PRC is shown by the dashed curve and the shifted simulated PRC is shown by the solid curve. The shifted simulated PRC was obtained by advancing the simulated PRC by 5 h, and it is plotted here only for comparison purposes.
inclusion of phosphorylation and lack of separation of nucleus and cytoplasm. Although we did not carry out a model for competing the hypotheses stated because of the scope of this Ph.D. research, it would be necessary to build such a model in the future, and then experiments and simulations would help decide the most appreciate model.

6.1.4 Mutations

A number of mutations that influence circadian rhythms have been reported in Drosophila. Mutations can be readily simulated in the model by changing particular parameters according to the functionality of mutants while keeping the rest of the parameter sets as standard.

We first explored E-box mutations. As explained in Section 2, we used the number of E-boxes of five in per and tim genes, six in pdpl gene and four in vri gene. Here we reduced the number of E-boxes in one gene and kept the others unchanged to create a single E-box mutation. We also reduced the number of E-boxes in more than one gene simultaneously for multiple E-boxes mutations. In all the simulations, oscillations in concentrations of all the mRNAs and proteins were preserved with shorter periods and lower amplitudes. The periods of the oscillations and the amplitudes of the phases were reduced to different extents for different E-box mutations. Figure 6-6 shows mRNA oscillations in the case of only one copy of E-box existing in each gene. The phase and anti-phase relationship between mRNAs were maintained and the period of oscillations (22.5 hours) was close to WT, as shown in Figure 6-1. The notable difference between the E-box mutation and WT was that the transcription levels of all the genes were reduced. This is consistent with experimental observations that rhythmic per and tim transcription remained in E-box mutations, although the transcription level was reduced (McDonald and Rosbash 2001).

Next, we tested some arrhythmic mutants. Drosophila per$^{01}$, tim$^{01}$ and clk$^{brk}$ refer to null mutations in per, tim and clk genes which produce non-functional proteins. These mutations were simulated by setting the translation rates of their respective proteins to zero. Figure 6-7 illustrates that sustained oscillations are abolished in per$^{01}$, tim$^{01}$ and
Figure 6-6 mRNA oscillations in E-box mutation simulations, where only one copy of E-box remains in each type of gene.
The results are consistent with the reports that rhythmicity of \( \text{per}, \text{tim} \) and \( \text{clk} \) mRNAs was blocked in \( \text{per}^{01}, \text{tim}^{01} \) and \( \text{clk}^{brk} \) (Bae, Lee et al. 1998), and oscillations in \( \text{pdp1} \) and \( \text{vri} \) mRNA levels were also blocked by these mutations (Cyran, Buchsbaum et al. 2003). However, some simulated mRNA levels, particularly \( \text{per} \) and \( \text{tim} \), greatly differed from the experimental reports. Here, we make some qualitative comparison between the simulated and experimental mRNA and protein levels. No quantitative comparison have been made because only relative abundance of molecular levels, rather than the concentration values for the molecules have been measured \textit{in vitro}. The experiments have shown that in mutants lacking \( \text{PER} \) (\( \text{per}^{01} \)) and \( \text{TIM} \) (\( \text{tim}^{01} \)), \( \text{per} \) and \( \text{tim} \) mRNA levels are constitutive and low (So and Rosbash 1997); \( \text{vri} \) mRNA levels are at intermediate levels (Blau and Young 1999); \( \text{pdp1} \) mRNA levels are high (Cyran, Buchsbaum et al. 2003); and the levels of \( \text{clk} \) mRNA are low (Glossop, Lyons et al. 1999). The simulated results showed high levels of \( \text{per}, \text{tim}, \text{vri} \) and \( \text{pdp1} \) mRNAs and a low level of \( \text{clk} \) mRNA (Figure 6-7A). These results can be explained by the structure of the model: \( \text{per}^{01} \) and \( \text{tim}^{01} \) induced absences of \( \text{PER} \) and \( \text{TIM} \) leading to a loss of \( \text{PER/TIM} \) which, in turn, caused a very high level of \( \text{CLK} \). As the activation effects of \( \text{CLK/CYC} \), \( \text{per} \), \( \text{tim} \), \( \text{vri} \) and \( \text{pdp1} \) mRNAs were all higher than their peaks in \( \text{WT} \), consequently, high concentrations of \( \text{VRI} \) and \( \text{PDP1} \) were produced. Because we assumed that \( \text{VRI} \) has a stronger binding ability to \( \text{CLK} \) than \( \text{PDP1} \) (probabilities of \( \text{VRI} \) and \( \text{PDP1} \) binding to \( \text{CLK} \) are 0.65574 and 0.304181 in this condition from calculation), strong repression from \( \text{VRI} \) made a low level of \( \text{clk} \) mRNA. In \( \text{clk}^{brk} \) flies, experimental data have shown low levels of \( \text{per}, \text{tim}, \text{vri} \) and \( \text{pdp1} \) mRNAs (Allada, White et al. 1998; Cyran, Buchsbaum et al. 2003), and a high level of \( \text{clk} \) mRNA which is near the \( \text{WT} \) peak (Glossop, Lyons et al. 1999). The simulated data (Figure 6-7B) showed an agreement in terms of low levels of \( \text{per}, \text{tim}, \text{vri} \) and \( \text{pdp1} \) mRNAs, but a disparity with a low level of \( \text{clk} \) mRNA. The mechanism underlying these data from the model can be explained as follows. Because of the absence of activation effects from afunctional \( \text{CLK} \), low levels of \( \text{per}, \text{tim}, \text{vri} \) and \( \text{pdp1} \) mRNAs are produced. Consequently, low levels of \( \text{PER}, \text{TIM}, \text{VRI} \) and \( \text{PDP1} \) followed. A small amount of \( \text{clk} \) mRNA was present because the repression effect from \( \text{VRI} \) was higher than the activation effect from \( \text{PDP1} \) under the assumptions in this model.

In \textit{in vitro} experiments, besides arrhythmic mutants in \textit{Drosophila}, a number of short and long mutants have also been observed. For example, \( \text{per}^L \) mutants lengthen the free-
Figure 6-7 Simulation of arrhythmic mutants. Parameter values are as in Table 5-1, except for $t_{per}=0$ for $per^{01}$, $t_{tim}=0$ for $tim^{01}$ and $t_{clk}=0$ for $clk^{Jrk}$. 
running periods to 29 h and \( \text{per}^S \) mutants shorten the free-running periods to 19 h (Konopka and Benzer 1971). It was suggested in a previous theoretical study that \( Drosophila \)'s \( \text{per}^S \) and \( \text{per}^L \) mutants can be modelled computationally by altering the stability of the PER protein or PER-protein interactions (Ruoff and Rensing 1996). In our model, \( \text{per}^S \) mutants were simulated by setting an enhanced rate of degradation of the PER/TIM dimer (dpt) according to the results from (Curtin, Huang et al. 1995). Similar to the simulations carried out by (Ruoff, Christensen et al. 2005), \( \text{per}^L \) mutants were represented by increases in the PER/TIM stability although this has not been experimentally confirmed. Figure 6-8 shows the PER plots of \( \text{per}^S \) and \( \text{per}^L \) mutants with a period of 19 h and 29 h where the degradation rate of PER/TIM was set to 0.9 and 0.08, respectively. In \textit{in vitro} experiments, it has been shown that nuclear entry of PER is delayed in the three \( \text{per}^L \) types compared with that in WT flies (Curtin, Huang et al. 1995; Lee, Parikh et al. 1996) and a larger proportion of \( \text{PER}^S \) is phosphorylated at an earlier time in the morning than PER in \( \text{per}^S \) mutants (Edery, Zwiebel et al. 1994).

However, as the current model does not include phosphorylation of PER and separation of nucleus and cytoplasm, we intend to simulate these experimental findings in a complete model in future.

### 6.1.5 Possible function of VRI and PDP1 feedback loops

The previous mutant analysis revealed that the expression of \( \text{per} \) and \( \text{tim} \) is critical for maintaining the oscillations of all the components in the circadian clock. A question arising here is the nature of functionality of the newly found VRI and PDP1 feedback loops. Previous research on other organisms, such as \textit{Neurospora}, suggested that the interlocked feedback loops may contribute to the robustness of the circadian clock (Cheng, Yang et al. 2001). Do the VRI and PDP1 feedback loops play a role in increasing the robustness of the circadian clock in \textit{Drosophila}? To find out the answer, we compared the robustness of the system, with and without the two feedback loops, toward parameter changes. To get a quantitative measure of the robustness, sensitivity analysis was performed.

We first removed the VRI negative feedback loop and/or the PDP1 feedback loop in the model by fixing their expression. It was found that the rhythmicity of the mRNAs and
Figure 6-8 Simulation of short and long mutants. Parameter values are as in Table 5-1, except for dpt=0.9 for per$^S$ and dpt=0.08 for per$^L$. 
the proteins was preserved in all the cases except the one being fixed. We plotted the
time evolution of proteins in Figure 6-9. For comparison, we also plotted the protein
concentrations when PER/TIM feedback loops were removed in Figure 6-9. It clearly
showed the disappearance of oscillations of all proteins with a fixed per and tim gene
expression.

Next, we calculated the ratios of the periodic sensitivity of the models without the VRI
and/or PDP1 feedback loops compared to the model with complete feedback loops. The
periodic sensitivity was calculated according to Eq. (6.1), and its derivation is given in
Appendix D.

\[
S(\tau; p_j) = \frac{p_j}{\tau(p_j)} \frac{d\tau(t, p_j)}{dp_j}, \tag{6.1}
\]

where \(p_j\) is the parameter with parameter index \(j\), \(\tau(p_j)\) is the period of the system for
parameter \(j\). Five percent perturbation in parameters was applied. It is clear that the
smaller the \(S\) value, the greater the robustness of the system to perturbations for this
parameter.

The simulation results are plotted in Figure 6-10, which show five parameters whose
sensitivity increases more than 10 times whereas the sensitivity of the other parameters
only yielded small changes when removing the VRI or/and PDP1 feedback loops. All
five parameters were related to the \(clk\) gene, mRNA or CLK protein. In particular, the
sensitivities of the transcription and translation rates of CLK increased more than fifty
times when removing the VRI feedback loop. Because of the critical roles that CLK
plays in regulating all the rhythmically expressed genes, and even some non-
rhythmically expressed genes, we propose a possible function of the VRI and PDP1
feedback loops is to decrease the sensitivity of CLK to parameter variations and
therefore increase the robustness of the circadian clock.
Figure 6-9 Time evolutions of proteins. The feedback loop was removed by making its corresponding gene expression constant.
Figure 6-10 Parametric sensitivity results. The parameter indexes are as indicated in Table 5-1.
6.2 Discussion of the circadian clock model

In Chapters three to six, a model has been presented for the circadian clock in *Drosophila* incorporating the key clock component genes identified so far. This model has unique properties compared with most of the previous models: (1) The model incorporates the transcriptional regulation of *per*, *tim*, *vri*, *pdp1* and *clk* genes. (2) The conventional Hill function to describe the regulation of gene expression was not assumed in the model, and that paves the way to study transcriptional regulation in the circadian clock at a more detailed level. (3) First-order reactions were used to describe translation and degradation processes, leading to a simpler model that is easier to analyse.

Using a set of parameters, the model produced autonomous sustained oscillations under conditions corresponding to constant darkness. The simulated results showed correct phases for all the components in the system, correct phase and anti-phase relationships of mRNAs and proteins, as well as appropriate lags between mRNAs and proteins. The model also accounted for the disappearance of the oscillations in the condition of constant light.

Robustness is an important characteristic of the circadian clock, which should produce close to 24 hours periodic oscillations regardless of modest variations in parameters under certain conditions. We measured variations of period by increasing and decreasing each parameter in turn by 20%. The oscillatory patterns remained in all the cases with the largest period of variation being around 0.8 hours for 20% parameter perturbations. Parameter sensitivity analysis suggested that several most sensitive parameters were binding rate of PDP1 to *clk* promoter, and binding rates of CLK/CYC to *pdp1* and *per* promoters. These are all positive elements (transcriptional activators) in the network.

It is also essential that the circadian clock should have the ability to reset phases in response to Zeitgeber, where light is the most important. We simulated the effect of light by increasing the degradation rates of TIM and PER. Simulations have shown the
entrainment of the system by light dark cycles and the induction of phase shifts by light pulses. In the entrainment by light dark cycles, the phase relationship in mRNAs and proteins were well maintained with a period of 24 hours and the phase of oscillations was delayed depending on the particular degradation rates we chose. We also constructed a phase-response curve to represent the phase shifts induced by temporal promotion of TIM and PER degradation. When normalising the simulated PRC by advancing it for 5 hours, the agreement between the normalised and experimental PRCs appeared very good. Both data showed a dead zone in the middle of subjective day, a phase delay during early subjective night, and a phase advance during late subjective night. The time lag between the simulated and the experimental data suggested that some unrepresented mechanisms in the model, such as phosphorylation and nuclear entry of TIM and PER, are important to providing a time delay in response to light.

We also carried out a number of tests for simulating mutations. Mathematical mutants were simulated by setting an appropriate parameter value according to the functionality of mutants. The simulated short and long mutants, perS and perL, resembled their phenotypes where 19 h and 29 h of period were found, respectively. In arrhythmic mutants, oscillations of all the mRNAs and proteins were blocked in per01, tim01 and clkbrk, as shown in the experiments. However, some mRNA levels significantly differed from the experimental data. Particularly, simulated data have shown high levels of per and tim mRNA in per01 and tim01 and low level of clk mRNA in clkbrk, which were opposite to that in experiments. This deficiency obviously came from the structure of the model, as discussed previously. In the model we assumed that the per, tim, vri, and pdp1 promoters were all strongly activated by CLK/CYC. The low levels of per and tim mRNAs in per01 and tim01 cannot be explained with this model because the loss of PER/TIM directly resulted in a high level of CLK and, consequently, high levels of per and tim mRNAs. Furthermore, although the assumption of strong binding ability of VRI to CLK gave a reasonable low level of clk mRNA in per01 and tim01, this assumption, nevertheless, makes a low level of clk mRNA in clkbrk, which was again different from the experimental observations in which a high level of clk mRNA was found. The deficiency of the model could indicate the possibility of an unknown part in the gene-protein network.
To find out the possible function of VRI and PDP1 feedback loops, we fixed the expression of VRI and PDP1 and found that the oscillations of PER, TIM and CLK remained, whereas the oscillations of VRI, PDP1 and CLK disappeared when PER/TIM loops were removed. Through the sensitivity analysis, it suggested that the VRI and PDP1 feedback loops decreased the sensitivity of CLK to the parameter changes and therefore increased the robustness of the circadian clock.

An important property of the model, which distinguished itself from the previous models, was the way that the regulation of transcription processes is modelled. In previous models, transcriptional regulation was modelled by the Hill function without explicit description of binding and unbinding processes of TFs to E-boxes elements in promoters (Allada, White et al. 1998; Leloup and Goldbeter 1998; Glossop, Lyons et al. 1999; Ueda, Hagiwara et al. 2001; Smolen, Hardin et al. 2004). The Hill cooperativity coefficient may correspond to the number of binding sites of genes (Hill 1910; Segel 1993). Different models used different Hill cooperativity coefficients to make sustained oscillations. The exact value of the minimum cooperativity coefficients depended on the choice of the model structure and model parameters. In most of the previous models, a Hill coefficient of more than one was used to describe the activation of per expression by CLK or repression of per expression by PER to create oscillations, whereas in some models it was found that limit cycle oscillations were preserved with a Hill coefficient of one if other parameters were properly chosen (Leloup and Goldbeter 1998; Tyson, Hong et al. 1999; Kurosawa, Mochizuki et al. 2002). Using the explicit description of TFs binding to promoters and activating or repressing gene expressions, our model can readily take account of different binding sites and cooperativity. The simulation has shown that even with one E-box in per, tim, vri and pdp1 promoters, oscillations were preserved with reduced transcription levels in agreement with in vitro experiments that only one copy of E-box did not abolish rhythmic per and tim transcription, although the transcription levels were reduced.

Finally, we shall make some comparisons with two previous models as the core mechanisms of these models are similar (Smolen, Hardin et al. 2004; Ruoff, Christensen et al. 2005). All the models contained two interlocked transcription and translation feedback loops, where on the one hand PER repressed its own gene expression by
binding to its activator CLK and, on the other hand, VRI and PDP1 regulated the gene expression of *clk*.

In the model proposed by Ruoff et al. (2005), the core mechanism is that VRI and PDP1 regulate *clk* expression with negative and positive feedback loops and CLK, the product of *clk* expression, activates *vri*, *pdp1* and *per/tim* (two genes were combined) expression. The simulation showed that VRI and PDP1 feedback loops generated sustained oscillations even in the absence of PER/TIM. Therefore, the authors concluded that positive and negative feedback loops of VRI and PDP1 were essential for the overall oscillations, whereas PER/TIM played a role in the amplification and stabilization of the oscillations. This result was in contrast to the findings from the model proposed by Smolen et al. (2004) in which PER feedback loop was found crucial for the oscillations. Ruoff et al. (2005) stated that the discrepancy of the findings may be because Smolen’s model used differential equations with delay terms where delay terms alone can generate oscillation. In our model, *per*\(^{01}\) and *tim*\(^{01}\) mutants suggested that PER and TIM were required for the oscillations of all the mRNAs and proteins and removal of VRI and/or PDP1 feedback loops did not remove rhythmicity of *per*, *tim* and *clk* expression. Our findings confirmed the roles of PER/TIM and VRI/PDP1 feedback loops made by Smolen et al. (2004) with a different model representation and a set of parameters.

The main difference in terms of model representation between Smolen’s and the present models is that different assumptions are used to capture the essence of various interactions. Smolen’s model used the Hill function and Michaelis-Menten rate expression describing transcriptional activation and phosphorylation processes, and discrete time delay terms were included in the equations to describe the time lags between proteins. Our model took account of binding and unbinding processes of TFs to promoters but ignored the nuclear entry of proteins and phosphorylation of PER. However, the simulated results of the two models were very similar regarding oscillations in constant darkness, photic entrainment of oscillations, the PRC and null mutations of *per* and *clk*. Nevertheless, different predictions were made by the two models. For example, E-box mutations were readily simulated in our model whereas some short and long period mutants were observed in Smolen’s model. This confirmed the statement made by Murray (2002) that different mathematical models might be able
to create similar behaviours, and they are mainly distinguished by the different predictions they suggest and how close they are to the real biology. As both models were simplified to some extent from the real network, we expect that a more sophisticated model should be developed in future as more data emerge from experiments.
Chapter 7: Modelling noise in GRN: Biology and models of viral infection

From now, another important issue in systems biology, namely noise, is investigated. As we have discussed in Chapter one, ideally, the circadian clock system could be used as a biological model where the effects of noise are explored. However, a large number of reactions and parameters make the analysis of noise difficult. Instead, we will use another genetic network which has a simpler structure and contains the simplest, yet most well-researched biological organism, the virus. Viruses are important to study molecular and cellular biology because an understanding of viral genomes and virus replication provides basic information concerning cellular processes in general (Lodish 2003). In this chapter, the biological background of viruses and viral infection processes in host cells is given first. Then the existing mathematical models of viral infection are briefly reviewed. Finally, the model proposed by Srivastava (2002) is described. This will be used as a base model for developing our models.

7.1 Biological background

One of the most distinguishing characteristics of viruses is that they are obligate intracellular parasites. Although viruses carry their own genome, they fully depend on the host cellular machineries for their replication and spread. The life cycle of a virus is, therefore, intertwined with the host life cycle at all stages of cellular biological processes.

7.1.1 Virus structure

A single virus particle is called a virion. A typical virion is composed of a nucleic acid genome, which is the genetic material of the virus, surrounded and protected by virus-encoded proteins, called the capsid. The nucleic acid can be DNA or RNA and may be single or double stranded, linear or circular, fragmented or continuous, but is always the
same for any given type of virus. The capsid is a complex structure but is highly ordered, being made up of many identical subunits of viral protein. Some viruses have an extra layer of protection, namely an envelope, surrounding the capsid.

7.1.2 Viral infection in host cells

The growth of viruses in their host cells is a complex and highly organised process. Since many viruses make few or no enzymes because of their simple structure, they are dependent on host cell enzymes to produce more virus particles. For example, most viruses lack a polymerase, which is needed to copy genomes and ribosomes. Although various patterns of replication are applied to different types of viruses, a common character shared by all the viruses is that they must replicate in living cells. They first must induce either profound or subtle changes in the cell so that viral genes in the genome are replicated and viral proteins are expressed. Viruses then use portions of the cellular machinery to produce progeny viruses. The basic pattern of replication is as follows:

1. Entry of virus into the host cell

Viruses enter cells in a variety of ways according to the nature of the viruses but, generally, contain two steps. A virion must first specifically interact with the host cell surface before the viral genome is introduced into the cell. Then the virion penetrates into a cell. In particular, enveloped viruses penetrate through fusion of the viral envelope with the plasma membrane of the cell. Non-enveloped viruses may cross the plasma membrane directly or may be taken up into endosomes. They then cross (or destroy) the endosomal membrane.

2. Uncoating

The purpose of this is to make viral nucleic acids available to permit transcription. The nucleic acids have to be sufficiently uncoated so that virus replication can begin. When the nucleic acids are uncoated, infectious virions cannot be recovered from the cell.
3. Synthesis of viral nucleic acid and protein

Once inside the cell, the virus utilises the cellular machinery to replicate. The replication cycle produces functional RNAs and proteins as well as genomic RNAs or DNAs and structural proteins. The actual number of infectious viruses produced in an infected cell is called the burst size, and this number can range from fewer than 10 to over 10000, depending on the types of cell infected, nature of the virus, and many other factors.

Different types of viruses have different mechanisms for replicating their genomes. DNA viruses generally replicate their DNA in the nucleus of the host cell, while RNA viruses generally replicate their RNA in the cytoplasm. Newly synthesised viral DNA or RNA becomes the genome of progeny viruses, but it is also used to make mRNAs that code for viral proteins.

4. Virion Assembly

At the time that viral genomes are replicated, viral capsid proteins must be present to form viral structures. New viral coat proteins assemble into capsids including viral genomes. This brings together the newly formed nucleic acids and structural proteins to form the virus.

5. Release

The virion is released and repeats the process of it infecting new cells.

According to the viral infection processes described above, viruses can have one of two different effects on their cellular hosts once they enter into the hosts. On the one hand, they may result in abortive infections when a virus mistakenly infects a cell that does not permit viral replication. On the other hand, the virus may replicate its genetic material until the host cell is so full that it bursts. The released viruses then invade other host cells.
7.2 Mathematical modelling of viral infections

Numerous mathematical models have been proposed in the literature to understand virus-mediated diseases and to develop antiviral strategies. Traditionally, the focus of theoretical models for describing the interaction between viruses and hosts is on population dynamics. Research of virus-host interactions at the population level is a very mature field and is closely linked to the understanding of the immune system and immune responses. In this type of model, the cell infection rate is normally taken to be proportional to the concentration of uninfected cells and virions and the virion production rate is proportional to the concentration of infected cells (May 2000; Wodarz and Nowak 2002; Orive, Stearns et al. 2005).

With the increasing understanding of virus and hosts in molecular biology over recent decades, kinetic models at a molecular level have been developed for virus infection cycles in bacteria, insect and mammalian cells. These models mainly take the form of a system of deterministic differential equations.

A model from Dee et al. (1995) described a Semliki Forest virus infection in baby hamster kidney (BHK-21) cells which only dealt with the initial steps of the virus infection cycle including virus attachment and viral fusion in the endosome. From the model, the authors found that the infection was most likely controlled at the level of viral uncoating and increasing the attachment rate could lead to new virion production.

A more detailed kinetic model developed by Endy et al. (1997) examined the growth of the bacteriophage T7, a virus that infects Escherichia coli bacterium. The model described individual steps of infection in detail, including the entry of the T7 genome into its host, expression and replication of T7 gene, translation of mRNA, and progeny phage assembly. The simulation of predicted concentrations for each component of the phage progeny show good agreement with experimental data. Later, Endy et al. (2000) used this model to study the effect on viral growth of drugs targeting different virus functions. Simulations showed that drugs targeting components of negative feedback loops were effective against mutant viruses that attenuated the drug-target interaction.
Furthermore, an extended model based on the previous T7 model was developed to account for the effects of host physiology on phage development (You, Suthers et al. 2002). It was found that phage growth was most sensitive to the host translation machinery, specifically, the level and elongation rate of the ribosomes. Another detailed deterministic model for intracellular processes of influenza virus replication in animal cells was developed by Siorenko and Reichl (2004). The model accounted for the individual steps of the process such as attachment, internalisation, genome replication and translation, and progeny virion assembly. Simulation results showed that an important factor limiting the growth rate of progeny viruses and their release was the total amount of matrix proteins in the nucleus while other newly synthesised viral proteins and viral RNAs accumulated.

In stochastic modelling of viral infections, the dynamics of virus and host cells may show different behaviour under the same initial conditions. The first stochastic model which displayed a switch-like behaviour of the virus was proposed by Arkin et al. (1998). The model used a statistical thermodynamic model to describe promoter regulation and stochastic chemical kinetics to formulate gene expression of phage lambda in its host cells. It successfully simulated the lysis-lysogeny decision of the phage lambda and predicted the fraction of infected cells at different phage/cell ratios that were consistent with experimental observations where conventional deterministic kinetics cannot be used to produce probabilistic outcomes.

All the previous models used either a deterministic or stochastic approach. Later a simple mathematical model for a generic case of viral infection was developed by Srivastava et al. (2002) that used both modelling approaches. A comparative study suggested that different modelling approaches provided different transient kinetics and different steady state levels of viral components, particularly when the infection of the cell was initiated by a low number of virus particles.
7.3 Description of an intra-cellular viral infection model

In this research, the models we intend to develop are based on the model proposed by Srivastava et al. (2002). This model has a simple structure to describe all the key steps in the intracellular viral infection processes. Specifically, the model accounts for transcription of viral genome, translation of the resulting mRNAs and degradation, protein synthesis and degradation and, eventually, production of phage progeny (Figure 7-1).

We first make some important assumptions that greatly simplify the cellular process in the model.

- Many individual reaction steps are combined into a single step. For example, the initialisation, elongation and termination of transcription processes are combined into a single process from DNA to mRNA.
- Many possible sources of interactions with host processes are neglected. For example, the same viral RNA that serves as a message for protein synthesis is also treated as a template for RNA replication.
- A cell cannot be re-infected once it is infected.
- No consideration is given to the division of the cells, so the size of cells does not change during viral infection processes.
- No consideration is given to different cellular compartments such as cytoplasm or nucleus.

Based on the assumptions and the biological basis, the compacted reactions in Table 7-1 are presented for the intracellular viral infection processes, where G denotes genome, R denotes mRNA, P denotes protein, and V denotes progeny virus.
Figure 7-1 The scheme of the viral replication cycle.
Table 7-1 Reaction steps and descriptions. The units of $k_2$ are molecules$^{-1}$day$^{-1}$ and the units of other parameters are day$^{-1}$.

<table>
<thead>
<tr>
<th>Reaction step</th>
<th>Description</th>
<th>Parameter value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G \xrightarrow{k_1} R$</td>
<td>Production of viral mRNA</td>
<td>$k_1=0.025$</td>
</tr>
<tr>
<td>$R \xrightarrow{k_2} \emptyset$</td>
<td>Degradation of mRNA</td>
<td>$k_2=0.25$</td>
</tr>
<tr>
<td>$R \xrightarrow{k_3} G + R$</td>
<td>Production of viral genome</td>
<td>$k_3=1$</td>
</tr>
<tr>
<td>$G + P \xrightarrow{k_4} V$</td>
<td>Virus assembly</td>
<td>$k_4=7.5 \times 10^{-6}$</td>
</tr>
<tr>
<td>$R \xrightarrow{k_5} R + P$</td>
<td>Production of viral structure protein</td>
<td>$k_5=1000$</td>
</tr>
<tr>
<td>$P \xrightarrow{k_6} \emptyset$</td>
<td>Degradation of protein</td>
<td>$k_6=1.99$</td>
</tr>
</tbody>
</table>
7.4 Deterministic model

If the time evolution of the system is considered deterministic and all the reactions are considered to be elementary in the sense that their rate is proportional to the number of the corresponding reactants, we can apply the mass action rate law to derive a set of differential equations that govern the components of the system:

\[ \frac{d[R]}{dt} = k_1 [G] - k_2 [R], \quad (7.1) \]

\[ \frac{d[G]}{dt} = k_3 [R] - k_4 [G][P], \quad (7.2) \]

\[ \frac{d[P]}{dt} = k_5 [R] - k_6 [P] - k_4 [G][P], \quad (7.3) \]

\[ \frac{d[V]}{dt} = k_4 [G][P], \quad (7.4) \]

where \([G], [P], [R]\) and \([V]\) are the concentrations of \(G, P, R\) and \(V\). The values of the rate constants, as indicated in Table 7-1, are determined by setting the steady-state values of \(R, G\) and \(P\) to 20, 200, and 10000 molecules, respectively, in the deterministic solution. Initial conditions represent the state of different multiplicity of infection (MOI) where the system starts from, where an initial infection of one molecule of \(G\) represents the low MOI and an initial infection of one molecule of \(G\) and five molecules of \(R\) represents the high MOI.

Numerical examination of the above equations revealed that the components in the system evolved deterministically, as illustrated in Figure 7-2. The system of differential equations was solved using the “ode45” function in Matlab. Linear stability analysis of the system was previously examined by Srivastava et al. (2002) who pointed out that these equations allow two steady-states. The trivial solution occurred when \([R]=[G]=[P]=0\) which also represented that the viral infection was extinct. Deterministically, this unique state was possible either right from the beginning if we
Figure 7-2 Time evolution of the components of the system described by ODEs. Simulations are for low MOI. (A) plots of genome (G) and protein (P). (B) plots of mRNA (R) and virus (V).
started with initial concentrations \([R_0]= [G_0]= [P_0]=0\), or after the decaying of all \([R]\), \([G]\) and \([P]\) to zero. The other steady state occurred when \([R]=20, [G]=200,\) and \([G]=10000\). Furthermore, it was revealed that the first steady state is unstable due to the presence of a positive eigenvalue, and the second one is stable since all of the eigenvalues are negative. For the details of the linear stability analysis of this system, we refer to Srivastava et al. (2002). Since the system started with some molecules being present and, therefore, was perturbation from the unstable steady state, the deterministic solutions always showed that the concentrations of \(G, R,\) and \(P\) grew until they reached the stable steady state.

### 7.5 Stochastic simulations via the Gillespie algorithm

The time evolution of the viral components from the deterministic solution represents the behaviour of the system without any noise. If we model the state of the biochemical system as discrete, using the Gillespie algorithm, rather than continuous, intrinsic noise may have effects on the dynamics of the system (Gillespie 1976). How the Gillespie algorithm works for this particular system will be presented in detail in the next chapter.

Because of the presence of intrinsic noise, the infection of each cell by the virus is different. In stochastic simulations via the Gillespie algorithm, each sample realisation (path) represents the behaviour of one cell. Figure 7-3 plots the sample realisations of mRNA generated from the Gillespie algorithm representing the dynamics of viral infection. In general, the individual sample realisation can be categorised into two distinct infections, called the successful infection and the abortive infection, respectively. The successful infection is defined as the viral components accumulate and eventually fluctuate around their stable steady states (Figure 7-3A). The abortive infection is defined as all the viral components reach to zero so no further viral infection is expected (Figure 7-3B). The time evolutions of \(P\) and \(G\) were similar to that of \(R\), either increasing until fluctuating around their respective steady states or dropping to zero so that the infection terminated.
Figure 7-4A shows the frequency distribution of mRNA at 200 days post-infection over 1000 realisations. Together with the frequency distribution of genome and protein (data not shown), it was found that about 26% of cells had abortive infections. Figure 7-4B shows the average behaviour of mRNA in 1000 cells and the average behaviour of mRNA in the filtered cells, which included only the cells with a successful infection. The average number of the unfiltered cells converged to 15 which was significantly lower than that of the deterministic solution due to the inclusion of the abortive infections. However, even in the filtered cells the stochastic average of the successful infections resulted in a plateau that was still slightly lower than the deterministically predicted stable model. Furthermore, the transient dynamics were also consistently lower. This illustrates that the effects of intrinsic noise not only existed during the initial stage of the infection, where there was a low number of components, but also during the whole stochastic process.
Figure 7-3 Time evolution of mRNA in the stochastic simulations solved by the Gillespie algorithm. The rate constants used are the same as their deterministic counterparts. Low MOI is used as the initial condition. (A) A sample realisation of successful infection. (B) A sample realisation of abortive infection.
Figure 7-4 Stochastic simulation results of 1000 realisations. (A) The mRNA frequency distribution at 200 days post-infection. The x-axis indicates the percentage of mRNA present in the cells. The y-axis indicates the number of mRNAs at 200 days. (B) The average time evolution of 1000 stochastic realisations (dashed line) and the average time evolution of filtered cells (dotted line). The filtered cells are the cells which have only successful infections. For comparison, the time evolution from the deterministic solution is also plotted (solid line).
Chapter 8: Investigation of intrinsic and extrinsic noises in the viral infection model

Genetically identical cells growing under the same conditions can still vary greatly in their internal mRNA and protein concentrations. This variation is generally not a direct measurement of intrinsic noise, instead it is a combination of the effects of intrinsic and extrinsic noises (Swain, Elowitz et al. 2002). Therefore, investigation of intrinsic and extrinsic noise effects is an essential part in understanding complex behaviour of genetic networks. However, all the models describing viral infection processes in the literature so far have focused on the stochastic effects of the intrinsic noise. No model has been proposed for accounting for its sources. Furthermore, no modelling efforts have been put into the significance of the extrinsic noise for viral infection systems. The aim of this chapter is to elucidate the effects of both types of noises in an intracellular viral infection model using mathematical models. Unlike using the conventional Gillespie algorithm, chemical Langevin equations (CLEs) are used to find out the sources of the intrinsic noise. Next, a set of stochastic differential equations (SDEs) is developed to explore the effects of irregular but continuous perturbations on the system parameters. Finally, numerical experiments are carried out, and simulation results, as well as the biological significance, are discussed.

8.1 Noise measurements

To compare the level of noise of a target species in the system, $\eta^2$ is used; that is the variance divided by the square of the mean value. The ratio $\eta^2$ (or, alternatively $\eta$) is typically referred to as the coefficient of variance, or the noise (Kærn, Elston et al. 2005). It is the most direct and unambiguous measure of noise and gives an indication of the spread of population with respect to the mean, and it has been used in previous research (Swain, Elowitz et al. 2002; Paulsson 2004; Dublanche, Michalodimitrakis et al. 2006).
Let $P(t)$ be the number of a measured species at time $t$, then $\eta^2(t)$ can be defined by

$$\eta^2(t) = \frac{\sigma^2(t)}{\mu^2(t)} = \frac{\langle P(t)^2 \rangle - \langle P(t) \rangle^2}{\langle P(t) \rangle^2},$$

where $\sigma^2(t)$ denotes the variance of $P(t)$, $\mu(t)$ denotes the mean of $P(t)$ and $\langle \rangle$ denotes the expected value. The higher the $\eta^2$ is, the noisier the system.

In the literature, a different measurement, the Fano factor, is also frequently used (Orrell and Bolouri 2004). Fano factor is defined by the variance divided by the mean values ($\frac{\sigma^2(t)}{\mu(t)}$) it is used primarily to reveal trends that would be obscured by the characteristic $1/\sqrt{\mu(t)}$ scaling of noise arising from finite-number effects (Kærn, Elston et al. 2005). However, the system size was assumed to be unchanged in this study, therefore only $\eta^2$ is used as the noise measurement.

### 8.2 Intrinsic noise

In this section, we present the internal stochastic properties of the model. We study the origins of the intrinsic noise that have not been considered in previous research. In order to focus on the intrinsic noise, we assume that the fluctuations caused by extrinsic sources are not taken into account at this stage.

#### 8.2.1 Method

To account for the intrinsic noise, we can describe the reaction system as a birth-death stochastic process governed by a chemical master equation (Eq. (2.13)), which tells the time evolution of the probability of having a given number of molecules of each species.
For the ease of writing the CME for this particular viral infection system, it suffices to recall that

\[
\frac{\partial P(x,t)}{\partial t} = \sum_{j=1}^{M} a_j (x - v_j) P(x - v_j, t) - a_j P(x,t).
\]  

(8.2)

where \(x\) is the molecular state of the system, \(a_j\) is the probability density of reaction \(R_j\), and \(v_j\) is the state-change vector.

To characterise \(x\) of the current system at time \(t\), as described in Table 7-1, we use

\[
x_i \equiv \begin{cases} 
  r & \text{for } i=1 \\
  g & \text{for } i=2 \\
  p & \text{for } i=3 \\
  v & \text{for } i=4
\end{cases}
\]  

(8.3)

where \(r\), \(g\), \(p\) and \(v\) are the number of mRNA, genome, protein and virus in the system, respectively.

The function \(a_j\) for this system is

\[
a_j = \begin{cases} 
  k_1 x_2 & \text{for } j=1 \\
  k_2 x_1 & \text{for } j=2 \\
  k_3 x_1 & \text{for } j=3 \\
  k_4 x_2 x_3 & \text{for } j=4 \\
  k_5 x_1 & \text{for } j=5 \\
  k_6 x_3 & \text{for } j=6
\end{cases}.
\]  

(8.4)

It should be noted that the deterministic rate constants \(k_j\) are used as the stochastic rate constants \(c_j\) in Eq. (8.4), as the cellular volume is assumed to be unity in this model so that the concentrations and numbers of molecules are basically equivalent.
The function $v$ for the systems is
\[
v = \begin{bmatrix}
1 & -1 & 0 & 0 \\
-1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & -1 & -1 & 1 \\
0 & 0 & 1 & 0 \\
0 & 0 & -1 & 0 \\
\end{bmatrix}.
\]

(8.5)

The $(i, j)$ element in the above matrix represents the change in the $j$th chemical species when the $i$th reaction occurs.

Then the CME for the current system gives
\[
\frac{\partial}{\partial t} P(x, t) = k_1 (x_2 + 1) P(x_1 - 1, x_2 + 1, x_3, x_4) + k_2 (x_1 + 1) P(x_1 + 1, x_2, x_3, x_4) \\
+ k_3 (x_1 + 1) P(x_1 + 1, x_2 - 1, x_3, x_4) \\
+ k_4 (x_1 + 1) (x_3 + 1) P(x_1, x_2 + 1, x_3 + 1, x_4 - 1) \\
+ k_5 (x_1 + 1) P(x_1, x_2, x_3 - 1, x_4) + k_6 (x_3 + 1) P(x_1, x_2, x_3 + 1, x_4) \\
- (k_1 x_2 + (k_2 + k_3 + k_4) x_1 + k_4 x_2 x_3 + k_6 x_3) P(x_1, x_2, x_3, x_4) \\

\]

(8.6)

Because the analytic solution of this differential equation does not exist, we can generate sample paths of the process using a well-known Monte Carlo simulation, the Gillespie algorithm (Gillespie 1977) that provides a direct way to study the effect of internal noise. In this algorithm, the time $t + \tau$ at which the next reaction will occur is randomly chosen with $\tau$ exponentially distributed with parameter $a_0$, which is $\sum_{j=1}^{M} a_j$, where $M$ is the number of the reactions. For the current system,
\[
a_0 = k_1 x_2 + k_2 x_1 + k_3 x_1 + k_4 x_2 x_3 + k_5 x_1 + k_6 x_3.
\]

(8.7)
Once $\tau$ has been determined, the reaction $R_j$ that will occur at time $t+\tau$ is randomly chosen with probability $a_j / a_0$ according to Eq. (2.21). Then the system is updated by

$$x(t+\tau) = x(t) + v_j.$$ (8.8)

This simulation method is considered to be exact and construct “unbiased realisations” because it accounts for the stochastic nature of every reaction event (Gillespie 1977). Such realisations are fully consistent with the CME. However, it suffers greatly in the computational cost. Furthermore, such direct stochastic simulation methods cannot afford us a clear perspective on the origin and magnitude of the internal noise in the system (Rao, Wolf et al. 2002).

We propose to use CLE here. This method was originally designed to improve the computational efficiency of the Gillespie algorithm. It is assumed that if a macro-infinitesimal time scale exists in a system, its dynamics described by the Gillespie algorithm with discrete time steps can be well approximated by CLE with continuous time steps (Gillespie 2000). The advantages of CLE was that it not only speeds up the simulations significantly, it also gave a general methodology for showing how the internal noise involved in chemical reactions is related to the parameter values and the state variables that evolve with time. Note that the CLE only accounts for the intrinsic noise of the system although it approximates the intrinsic stochasticity of the system by an explicit noise term.

According to the general form of CLE (Eq. (2.29)), the explicit form of the CLEs for the system reads

$$dx_1 = (k_1 x_2 - k_2 x_1) dt + a_1 \sqrt{k_1 x_2} dW_1 - a_2 \sqrt{k_2 x_1} dW_2,$$ (8.9)

$$dx_2 = (k_3 x_1 - k_1 x_2 - k_4 x_2 x_3) dt + a_2 \sqrt{k_3 x_1} dW_3,$$

$$- a_1 \sqrt{k_1 x_2} dW_1 - a_4 \sqrt{k_4 x_2 x_3} dW_4.$$ (8.10)
\[
dx_3 = (k_5 x_1 - k_6 p - k_4 x_2 x_3) dt + a_5 \sqrt{k_5 x_1} dW_5,
\]
where \( W_i (i=1,\ldots,6) \) is a standard Wiener process with mean 0 and variance \( t \) at time \( t \) and the \( W_s \)s are independent of each other. The coefficient \( a_i (i=1,\ldots,6) \) denotes the intrinsic noise coefficient which only takes binary value of 1 or 0 to include or exclude the intrinsic noise contribution from the corresponding reaction. The values of all the parameters were the same as the deterministic counterparts, as shown in Table 4-1. Clearly the deterministic equations can be obtained by removing the terms involving \( W_i \).

To solve the SDEs above, the numerical algorithm that we used is the Euler method, also called Euler-Maruyama method (Gard 1988). In addition to this common method, there are numerous higher order numerical schemes that require the sampling of multi-dimensional Itô integrals. For an overview of SDE integrators, we refer to a classic book on numerical methods for SDEs (Kloeden and Platen 1992). We will verify in the following section that the Euler method is sufficient to produce reliable results by showing that the difference between the numerical solutions of the SDEs and the exact Gillespie method is negligible. Therefore, we limit ourselves to the Euler method. In all the simulations, a small \( \Delta t \) of 0.0001 was used. The numerical algorithm was implemented using Matlab 7.0 and all the simulation were carried out on a personal computer (Pentium D 3.0 GHz and 1.0 GB RAM) running the Microsoft Windows XP operating system. All the results were calculated from 1000 realisations.

8.2.2 Simulations and results

We first verified the legitimacy of using CLE to investigate the intrinsic noise. For simplification, we chose mRNA as a representative of viral components and only measured some statistics of mRNA. Figure 8-1A shows the average behaviour of mRNA over 1000 realisations simulated by the Gillespie algorithm and CLE where the
system initially consisted of only one molecule of mRNA (low MOI). It shows that the average number of mRNA using CLE was consistently and considerably higher than that of the Gillespie algorithm once it rose initially. The reason can be explained from Figure 8-1B which shows the frequency distribution of mRNA in a long-time limit. Here we assume that the system reaches the steady-state in its long-time limit at 200 days post-infection according to the deterministic solution. Figure 8-1B clearly illustrates that CLE did not display an extinction of mRNA whereas extinctions appeared in around 26% of cells with the Gillespie algorithm. In order to cause the viral extinction, the three variables, R, G and P, must become zero simultaneously. In the Gillespie algorithm, the number of molecules was likely to drop from one to zero because of the discrete nature of the system state when there was only one mRNA molecule present initially. This did not occur for the CLE simulations because the continuous approximation of CLE allowed the number of molecules to be in fractions. Even when the number of molecules was close to zero, such as 0.1 or 0.01, it still had a chance to pick up again. We therefore conclude that the low MOI does not fulfill the legitimation of CLE.

Next we compared the stochastic simulation results for the high MOI case, where five molecules of mRNA and one genome were present initially. Figure 8-1C shows that the average population of mRNA during 200 days was nearly identical in both methods. Figure 8-1D shows that the frequency distribution of mRNA at time 200 days was also in good agreement in two cases. Furthermore, we compared the mean value and noise of all the viral components at 200 days post-infection in Table 8-1. The results produced by the discrete and continuum simulations agreed very well. In addition, the computation time of running 1000 realisations for CLE was significantly lower than that for the Gillespie, which were 0.97 and 27.56 hours, respectively. We therefore started with the initial state of the high MOI in all the following computations so that CLE remained reliable.

We now look for the sources of the intrinsic noise. As CLE is represented by adding white noise terms which describe the perturbation from the intrinsic noise into the ordinary differential equations, we can readily determine the contribution of the intrinsic noise from each reaction \( i \) by setting the corresponding intrinsic noise coefficient \( a_i \) to be one and setting \( a_i \) in the other reactions to be zero.
Figure 8-1 Stochastic simulations of mRNA by the Gillespie algorithm and CLE for the low MOI (the first row) and the high MOI (the second row). 1000 realisations were obtained for each of the figure. (A, C) Average behaviour of the realisations during 200 days post-infection. (B, D) Frequency distribution at 200 days post-infection.
Table 8-1 Mean and intrinsic noise of all the viral components at 200 days post-infection. The values were obtained by the stochastic simulations via the Gillespie and CLE, respectively.

<table>
<thead>
<tr>
<th>Component</th>
<th>Mean CLE</th>
<th>Mean Gillespie</th>
<th>Intrinsic noise CLE</th>
<th>Intrinsic noise Gillespie</th>
</tr>
</thead>
<tbody>
<tr>
<td>mRNA</td>
<td>19.9</td>
<td>19.9</td>
<td>0.05813</td>
<td>0.05774</td>
</tr>
<tr>
<td>Genome</td>
<td>197.1</td>
<td>197.5</td>
<td>0.00694</td>
<td>0.00706</td>
</tr>
<tr>
<td>Protein</td>
<td>9980.8</td>
<td>9944.3</td>
<td>0.05189</td>
<td>0.05167</td>
</tr>
<tr>
<td>Virus</td>
<td>2017</td>
<td>2003.7</td>
<td>0.01940</td>
<td>0.01885</td>
</tr>
</tbody>
</table>
Table 8-2 illustrates the mean and the intrinsic noise of mRNA at 200 days when accounting for the intrinsic noise from each reaction separately. The sum of the noise from each reaction was consistent with the total noise obtained from Table 8-1. The results show that Reaction 2 contributed half of the total noise (50.2%) followed by Reaction 1 which made up 42.3% of the noise. Reactions 3 and 4 supplied around 4% and 3% of the total noise, respectively. Finally, Reactions 5 and 6 provided negligible noise to the whole system. To visualise the spreads of mRNA, we plotted the frequency distribution of mRNA in Figure 8-2. It clearly illustrated that Reactions 1 and 2 played significant roles in contributing to the intrinsic noise.

### 8.3 Extrinsic noise

In this section, how extrinsic noises affect the production of viral components is demonstrated. Several elegant experiments have proven that the extrinsic noise contributes more than the intrinsic in gene expression networks in both prokaryotes and eukaryotes (Elowitz, Levine et al. 2002; Raser and O'Shea 2004). Perturbations of biochemical parameters, hence, must be considered (Becskei and Serrano 2000). The perturbations occur during transient changes of the biochemical processes, such as promoter binding in transcription and ribosome binding in translation (Kierzek, Zaim et al. 2001). For the current system, the biological processes are simplified by combining a complex sequence of reactions into single reactions. For example, the transcription process of a gene actually represents, in reality, the initiation, elongation and termination of transcription. It is therefore natural to assume that the rate constants in our model were more likely to be subject to the extrinsic noise, such as the variability of the number of RNAPs, ribosomes and degradosomes and environmental changes.

#### 8.3.1 Method

To clearly separate the extrinsic noise from the intrinsic noise, the impact of parametric variability is investigated in the macroscopic limit governed by the deterministic rate
Table 8-2 Mean and intrinsic noise of mRNA at 200 days post-infection as well as the contributions of noise from each reaction. Reaction i represents that \( a_i \) was set to be 1 in order to account for the intrinsic noise contributed from that reaction and the intrinsic noises from the other reactions were silenced.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Mean</th>
<th>Intrinsic noise</th>
<th>Contribution of noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reaction 1</td>
<td>20</td>
<td>0.02468</td>
<td>42.335%</td>
</tr>
<tr>
<td>Reaction 2</td>
<td>19.8</td>
<td>0.02931</td>
<td>50.276%</td>
</tr>
<tr>
<td>Reaction 3</td>
<td>19.9</td>
<td>0.00249</td>
<td>4.271%</td>
</tr>
<tr>
<td>Reaction 4</td>
<td>19.9</td>
<td>0.00181</td>
<td>3.113%</td>
</tr>
<tr>
<td>Reaction 5</td>
<td>19.9</td>
<td>1.44 e^{-06}</td>
<td>0.002%</td>
</tr>
<tr>
<td>Reaction 6</td>
<td>19.9</td>
<td>1.44 e^{-06}</td>
<td>0.002%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>0.05830</strong></td>
<td></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>
Figure 8-2 Frequency distribution of mRNA at 200 days post-infection. Each figure accounts for the intrinsic variable of each reaction.
equations. In this limit, we replace the original rate constants $k_n$ by the noisy terms of $k_n + c_n \xi_n$, where $\xi_n$ is a Gaussian process which follows the statistics,

$$\langle \xi_n(t) \rangle = 0, \quad \langle \xi_n(t) \xi_n(t') \rangle = \delta(t-t'),$$  \hspace{1cm} (8.13)$$

where $\delta$ is the Dirac delta function. For this model, the explicit form of the SDEs is

$$dx_1 / dt = k_1 x_2 - k_2 x_1 + c_1 x_2 \xi_1 - c_2 x_1 \xi_2,$$  \hspace{1cm} (8.14)$$

$$dx_2 / dt = k_3 x_1 - k_4 x_2 x_3 + c_3 x_1 \xi_3 - c_4 x_2 x_3 \xi_4,$$  \hspace{1cm} (8.15)$$

$$dx_3 / dt = k_5 x_1 - k_6 x_3 - k_4 x_2 x_3 + c_5 x_1 \xi_5 - c_6 x_3 \xi_6 - c_4 x_2 x_3 \xi_4,$$  \hspace{1cm} (8.16)$$

$$dx_4 / dt = k_4 x_2 x_3 - c_4 x_2 x_3 \xi_4,$$  \hspace{1cm} (8.17)$$

where $x_n$ is defined in Eq. (8.3), $c_n$ is the extrinsic noise coefficient which allows us to take into account the intensity of random fluctuations. Here it is assumed that there is no correlation among noises, though they might be contributed by the same sources of environmental fluctuations. For numerical treatment, Eqs. (8.14) to (8.17) can be transformed into the form using the Wiener process. Then the current model gives

$$dx_1 = (k_1 x_2 - k_2 x_1) dt + c_1 x_2 dW_1 - c_2 x_1 dW_2,$$  \hspace{1cm} (8.18)$$

$$dx_2 = (k_3 x_1 - k_4 x_2 x_3) dt + c_3 x_1 dW_3 - c_4 x_2 x_3 dW_4,$$  \hspace{1cm} (8.19)$$

$$dx_3 = (k_5 x_1 - k_6 x_3 - k_4 x_2 x_3) dt + c_5 x_1 dW_5 - c_6 x_3 dW_6 - c_4 x_2 x_3 dW_4,$$  \hspace{1cm} (8.20)$$

$$dx_4 = (k_4 x_2 x_3) dt + c_4 x_2 x_3 dW_4.$$  \hspace{1cm} (8.21)$$
Put in a matrix form, we have

\[
\begin{bmatrix}
dx_1 \\
dx_2 \\
dx_3 \\
dx_4
\end{bmatrix} = \begin{bmatrix}
k_1 x_2 - k_2 x_1 \\
k_3 x_1 - k_1 x_2 - k_4 x_2 x_3 \\
k_5 x_1 - k_6 x_3 - k_4 x_2 x_3 \\
k_4 x_2 x_3
\end{bmatrix} dt
\]

\[
+ \begin{bmatrix}
c_1 x_2 & -c_2 x_1 & 0 & 0 & 0 & 0 \\
-c_1 x_2 & c_3 x_1 & -c_4 x_2 x_3 & 0 & 0 \\
0 & 0 & 0 & -c_4 x_2 x_3 & c_5 x_1 & c_6 x_3 \\
0 & 0 & 0 & c_4 x_2 x_3 & 0 & 0
\end{bmatrix} \begin{bmatrix}
dx_1 \\
dx_2 \\
dx_3 \\
dx_4 \\
dx_5 \\
dx_6
\end{bmatrix}
\]

(8.22)

where the second matrix in the right hand side can be further factored as

\[
\begin{bmatrix}
c_1 & -c_2 & 0 & 0 & 0 & 0 \\
-c_1 & 0 & c_3 & -c_4 & 0 & 0 \\
0 & 0 & 0 & -c_4 & c_5 & c_6 \\
0 & 0 & 0 & c_4 & 0 & 0
\end{bmatrix} \begin{bmatrix}
x_2 & 0 & 0 & 0 & 0 & 0 \\
0 & x_1 & 0 & 0 & 0 & 0 \\
0 & 0 & x_1 & 0 & 0 & 0 \\
0 & 0 & 0 & x_1 & 0 & 0 \\
0 & 0 & 0 & 0 & x_1 & 0 \\
0 & 0 & 0 & 0 & 0 & x_3
\end{bmatrix}
\]

(8.23)

We now can simply use a system of stochastic differential equations form to denote Eqs. (8.18) to (8.21) as

\[
dX = A dt + C B dW ,
\]

(8.24)

where \( X = \begin{bmatrix} r \\ g \\ p \\ v \end{bmatrix} \), \( A = \begin{bmatrix} k_1 x_2 - k_2 x_1 \\
k_3 x_1 - k_1 x_2 - k_4 x_2 x_3 \\
k_5 x_1 - k_6 x_3 - k_4 x_2 x_3 \\
k_4 x_2 x_3 \end{bmatrix} \),

\[
C = \begin{bmatrix} c_1 & -c_2 & 0 & 0 & 0 & 0 \\
-c_1 & 0 & c_3 & -c_4 & 0 & 0 \\
0 & 0 & 0 & -c_4 & c_5 & c_6 \\
0 & 0 & 0 & c_4 & 0 & 0
\end{bmatrix} , \quad B = \begin{bmatrix} x_2 & 0 & 0 & 0 & 0 & 0 \\
0 & x_1 & 0 & 0 & 0 & 0 \\
0 & 0 & x_1 & 0 & 0 & 0 \\
0 & 0 & 0 & x_2 x_3 & 0 & 0 \\
0 & 0 & 0 & 0 & x_1 & 0 \\
0 & 0 & 0 & 0 & 0 & x_3
\end{bmatrix} , \quad \text{and} \ W = \begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ W_4 \\ W_5 \\ W_6 \end{bmatrix} .
Again, as an analytic solution for the above stochastic differential equation does not exist, we use the Euler’s method to solve it,

\[ X(t) = X(0) + \int_0^t A \, dt + \int_0^t (C \, B) \, dW. \]

(8.25)

For comparison purposes, we make the initial conditions the same as those used for the intrinsic noise study, that is \( X(0) = [5, 1, 0, 0] \). In all the following simulations, \( \Delta t = 0.0001 \) was used in the numerical solution. All the mean and \( \eta^2 \) values were calculated from 1000 realisations.

### 8.3.2 Simulations and results

To elucidate the role played by each rate parameter perturbed by the extrinsic noise, each parameter was investigated separately. Experiments have shown that the intensity of the perturbations differs between cells in populations (Ko, Nakauchi et al. 1990). As we did not \( \text{a priori} \) know the magnitude of the noise, we applied the perturbation into the parameter \( k_i \), with values of \( c_i \) ranging of 0.1, 0.01, 0.001 and 0.0001, respectively.

Figure 8-3 shows the average behaviour of mRNA population and five sample realisations under different intensity levels of random fluctuations applied in the different parameters. It can be seen that most of the infections were successful in that mRNA converged to a steady-state of 20 molecules. The exceptions were that the average number of mRNA dropped below one when the noise coefficient (NC) was more than 0.001 in \( k_4 \) and reached to 40 when NC reached to 0.1 in \( k_1 \). The same noise level apparently had different effects on different parameters. The most insensitive parameter was \( k_5 \) for which the extrinsic noise seemed to have no impact on the system. For \( k_3 \) and \( k_6 \), the time evolution of mRNA only started fluctuating slightly when NC was as high as 0.1. Similar small fluctuations can be also seen when noise with a NC of 0.01 was applied in \( k_2 \). The system was more sensitive when the same level of noise was applied in \( k_1 \) rather than \( k_2 \). The most sensitive and interesting parameter was \( k_4 \) in which the noise with as low as 0.0001 of NC caused significant fluctuations on mRNA
levels. However, as NC reached to 0.001, the average level of mRNA decreased to a steady-state of around one molecule. With a NC value of 0.01, the average level of mRNA was only about 0.1 molecules and the variation of mRNA became larger. Moreover, an unrealistic average value of $10^{10}$ appeared with NC=0.1. The reason is that occasionally a huge fluctuation of mRNA, which might peak to more than $10^{10}$, led to abnormal behaviour of an average mRNA.

To investigate the quantitative relationship between the mean and $\eta^2$ of mRNA at a long-time limit and the magnitude of the external perturbation, we generated a number of simulations with NC values ranging from $10^{-1}$ to $10^{-6}$, in which 10 time points were used in every $10^{-n}$ to $10^{-n-1}$ range, where $n = 1$ to $5$, with even intervals. The mean values of mRNA at 200 days against the NC values were plotted in Figure 8-4A. It showed that the mean value of steady-state viral mRNA kept around 20 molecules when a perturbation was applied in any parameter except $k_4$ and $k_l$. While the NC > $10^{-5}$ on $k_d$, the logarithm of mean mRNA approximately had a linear decrease as the logarithm of NC increased. Another exception was that when NC > 0.02 on $k_1$, the logarithm of mean mRNA linearly increased as the logarithm of NC enlarged.

A similar plot is in Figure 8-4B to illustrate $\eta^2$ of mRNA at 200 days against the noise strength. It shows that the extrinsic noise also linearly increased as the NC increased logarithmically for all the parameters except $k_d$. Among all the parameters, the system was extremely vulnerable to the perturbation on $k_d$ which caused the value of $\eta^2$ to fluctuate at 1 when NC is more than 0.0001. Below this point, the logarithm of noise of mRNA was linearly proportional to the logarithm of NC. In contrast, the system was exceptionally robust with a noisy $k_5$. The $\eta^2$ was only $10^{-11}$ even when $k_5$ was perturbed with a high level of white noise (NC=0.1).
Figure 8-3 Time evolution of mRNA with different parameters perturbed under different noise levels. Black solid line denotes the average number of mRNA, and grey dashed line denotes the sample realisations. Five sample realisations are plotted. NC denotes the coefficient of noise ($c_i$).
Figure 8-4 The mean value and the extrinsic noise of mRNA at time 200 days over 1000 realisations against the coefficient of noise ($c_i$). The x-axis and y-axis in both figures are logarithmically scaled. Note in the figure A, the plots of reaction 2, 3, 5 and 6 are virtually identical.
8.4 Discussion

In this chapter, the stochastic properties of a viral infection model perturbed by the intrinsic and extrinsic noise have been investigated. The study of noise effects has two main purposes: (1) to analyse the sources of intrinsic noise in the model, and (2) to estimate the extent to which a given noise might influence the robustness of the system and the possible noise-induced phenomena.

The two types of the noises were studied separately in stochastic frameworks based on the deterministic model. In the first framework, the intrinsic noise was studied by means of a special type of SDE, the CLE, and in the second, the extrinsic noise was studied by a general form of SDE. Because of the non-linear nature of the models, they were solved by numerical schemes. The variation of the number of mRNA was taken as a measurement of the noise effects. Note that although we have only studied the viral mRNA in this research, a systematic and quantitative study of the other viral components can be carried out in the same way.

In the absence of an analytic solution of the master equation in the stochastic model, we compared the intrinsic noise properties with an exact numerical solution, the Gillespie algorithm, and by an approximation approach, CLE. The results from the Gillespie algorithm suggested that the probability of viral extinction was significantly influenced by the initial viral component density. With a starting point of low MOI, the infections were abortive in 26% of the cells. When the initial viral density was in high MOI, all the infections were successful. Then, the same initial conditions were used in the simulations obtained by the CLE approach. The average time evolution and frequency distribution of the mRNA population compared with the one from the Gillespie algorithm suggested that the CLE approach was not valid for simulations starting from a low MOI. This was because the CLE model can take the number of molecules with fractional values, with the result that it is much less likely that the number of species in the system goes down to zero. However, the results from the two approaches were in good agreement in terms of the average number and noise level of all the viral components with high viral density. Therefore, the use of CLE is validated for modelling for the high MOI system.
The advantage of CLE in accounting for the intrinsic noise is not only for speeding up the computational time considerably but also, and more importantly, the intrinsic stochasticity introduced by Gaussian white noise sources is directly related to each biochemical reaction. Therefore, we can distinguish different sources of intrinsic stochasticity in a system. From the numerical simulations, it was found that the intrinsic noise was a linear sum of the noise in each of the intrinsic variables \( \eta_{total}^2 = \sum_{i=1}^{n} \eta_i^2 \), where \( n \) is the number of reactions in the model. The intrinsic noise mainly arose from the degradation process of mRNA, which accounted for half of the total noise. The transcription process was another main source of intrinsic noise. In contrast, the translation and degradation of protein processes made marginal contributions. The findings agreed with previous work which proposed that the transcription process contributed to noise more than the translation process (Swain, Elowitz et al. 2002). This was also consistent with an in vitro experiment showing that a strong fluctuation in LacZ expression came from the transcription instead of the translation (Kierzek, Zaim et al. 2001).

The simulations studying extrinsic noise revealed that the system was most susceptible to the noise in \( k_4 \). Biologically, the parameter \( k_4 \) controls the rate of virus assembly process, where viral genome are packaged into structure proteins. The “susceptivity” could be explained by the absolute value of \( k_4 \) which was extremely low compared to other rate constants. The low rate constant made this part of the process more vulnerable to external fluctuations. Furthermore, the function of \( k_4 \) was nonlinear and multiplicative in the model whereas that of other rate constants was linear. This might suggest that the non-linearity function amplifies noise effects.

In the measurement of the mean of mRNA at a long-time limit, the system was insensitive to fluctuations in most of the rates \( (k_{2,3,5,6}) \). However, it is important to note that starting from a very low level of noise \( (NC=10^{-5}) \) in the viral assembly process, the average mRNA at a long-time limit decreased logarithmically with the increase in the logarithm of NC, and eventually the average mRNA converged to zero. This transition from the amplified to inhibited mRNA could have significant implications,
allowing identification of the most vulnerable process in the viral replication cycle to external perturbations from drugs, for example.

In the measurement of the extrinsic noise of mRNA at a long-time limit, the results showed consistently that the fluctuation of mRNA grew logarithmically with the increasing noise intensities. The system was exceptionally sensitive to the perturbed viral assembly process. Small deviations in the viral assembly rate led to large fluctuations in the production of mRNA. In contrast, the system was less susceptible to the noisy rate constant in the translation process. The other processes perturbed by the noise, which have less impact on the system, were degradation of protein, genome replication, degradation of mRNA, and transcription processes, respectively. The measurements of the susceptibility of the system to external noises could guide us in controlling experimental conditions when designing a synthetic network of viral replication.

However, an obvious limitation of this research should be noted here: the sources and outcomes of intrinsic and extrinsic noises were explicitly separated in this study. It is, however, often the case in practice that both types of fluctuations are simultaneously present in biological systems. The fluctuating species measured normally result from the combination of the effects of intrinsic and extrinsic noise (Swain, Elowitz et al. 2002) although separate measurements are now possible in in vitro experiments (Elowitz, Levine et al. 2002; Ozbudak, Thattai et al. 2002). To account for the simultaneous presence of both noises in a theoretical framework, we can introduce extrinsic noise through variations in an appropriate external parameter in the master equation and the resulting equation would contain both types of noises (Doraiswamy and Kulkarni 1987). However, because the transition probabilities in the master equation should be positive, this leads to the restriction that the extrinsic noise shall have bounded realisations and cannot be white noise. It is obviously a serious drawback as a large number of studies involving extrinsic noise have assumed it to be white noise, which is also the case in this research.
8.5 Concluding remarks

Numerous theoretical and experimental studies have shown that random noise plays a crucial role in the dynamic behaviour of GRNs. To aid in the analysis of the effects of noise in systems, we have proposed the use of SDEs as modelling frameworks to represent random noise. The frameworks can be readily applied to large scale models with both inherent and parametric perturbations.

A simple generic model of intracellular viral infection was chosen as an example of a GRN. Our theoretical frameworks provide a useful tool for understanding how viral infections propagate under the influence of noise. The analysis of the intrinsic noise helps us understand its sources and determines the relative importance of each biochemical process in the total noise. The analysis of the extrinsic noise allows us to explore how the target components of a virus may be influenced by the external forces, such as the fluctuating rate in a target process by drugs. The results give us insight into dominant reactions that drive the system’s dynamics, and therefore allow us to determine the most responsible pathway for controlling viral production. Eventually, it is hoped that the proposed models could provide experimental biologists with a more fundamental understanding of viral diseases and lead to better strategies for designing and interpreting *in vitro* experiments.
Chapter 9: Conclusions and future outlook

The overall goal of the thesis was to integrate our knowledge of mathematics and biology. In this thesis two mathematical models have been developed for two specific GRNs. Deterministic and stochastic approaches have been used for the two systems, respectively, depending on the detail level of the systems and the particular research questions we asked. The simulation results have shown that they have been successful in gaining insights into the biological systems. We now give an overview of what we have achieved and contributions of these achievements, future directions that can follow on from the current step and, finally, a conclusion.

9.1 General overview

The first focus of the work was to develop a deterministic model incorporating the current knowledge of the transcriptional regulatory networks in the circadian clock for Drosophila and then to analyse the model in detail. The purposes of the model were (1) to verify the current knowledge about the biology of the circadian clock system, (2) to increase our confidence in understanding the system correctly if the \textit{in silico} data are in good agreement with in vitro data, and (3) to guide the quest for missing parts if disagreements were found.

The reconstruction of the model was achieved with first the identification of the biological network of the circadian clock (Chapter 3). The biological processes in the regulatory network were identified using available experimental data from the literature, primarily, data from mutagenesis screens. After a careful reinterpretation of the literature, the conceptual model was built based on a number of assumptions (Chapter 4). Using the mass action rate law, a mathematical model consisting of 19 ordinary differential equations was derived (Chapter 4). The mathematical model was then converted into a computer solvable model. The kinetic model introduced 47 parameters which were all unknown experimentally. Therefore, they had to be determined by fitting
specific experimental observations. For the present, we used trial and error to get initial parameters then fine-tuned them by local parameter estimation algorithms (Chapter 5).

Finally, the resulting model was extensively tested and compared with the experimental data (Chapter 6). We summarised the findings of the model into the following three aspects. (1) The model simulated sustained circadian oscillations in mRNA and protein concentrations in constant darkness in agreement with experimental observations. It also simulated entrainment by light-dark cycles, disappearance of the rhythmicity in constant light and the shape of phase response curves resembling that of the experimental results. As the robustness was a vital characteristic for the biological system, the model was tested and found to be quite robust over a wide range of parameter variations. All these findings demonstrated the ability of the model to predict correctly experimental outcomes. (2) By extensively testing the parameter sensitivities of different model structures, we proposed that the function of VRI and PDP1 feedback loops is to decrease the sensitivities of CLK to parameter variations and thus increase the robustness of the whole system. This demonstrated that another important function that mathematical models serve is to make hypotheses that could be tested by the detailed experiments. (3) Through the mutant analysis, it was found that a deficiency existed between the simulated mRNA levels and experimental observations in \textit{per}^01, \textit{tim}^01 and \textit{clk}^{jrk} mutants. This suggested that unknown feedback loops might exist that control the regulation of \textit{per} and \textit{tim} genes.

The second part of the thesis investigated an intracellular viral infection model. It emphasised another key concept in systems biology — stochastic properties of the genetic networks. Specifically, three questions were asked: (1) what are the origins and consequences of the intrinsic noise? (2) How do the viral components in the individual cells change under the different noise levels? (3) How can we gain more insights into the system from the deterministic and stochastic simulations?

Before the development of stochastic models, we discussed the biology of viral infection and a deterministic model for a general viral infection process (Chapter 7). Based on this existing model, we derived CLEs to calculate the intrinsic noise level (Chapter 8). Originally, the CLE was proposed for reducing the computational cost of
stochastic simulations, but instead we used it as a tool to find out the origins of the intrinsic noise, which cannot be gained from Monte Carlo simulations. We compared the mean values of time evolution and frequency distribution of viral components at a long-time limit. It was shown that the simulation from the CLE agreed well with the Gillespie algorithm in the case of high MOI. By accounting for the intrinsic variable from each reaction, the contributions of the noise from each biochemical reaction were acquired. It was found that the degradation of mRNA and transcription processes accounted for considerable proportions of total intrinsic noise. In contrast, the translation process took marginal part of the noise. Next, we characterised the behaviour of viral components under the different levels of extrinsic noises (Chapter 8). The extrinsic noise had a significant impact on the process of virus assembly. With a noisy rate constant above certain noise intensity, the cell even inhibited the viral growth. The biological significance of this phenomenon is that if the virus assembly process is targeted by drugs, the viral infection is more likely to be aborted.

9.2 Contributions

The contributions of this thesis span two topics through modelling and analysis of the two systems.

- To advance the understanding of the circadian clock by building and validating models incorporating current knowledge.
- To develop a model incorporating the detailed transcription processes.
- To use CLE to explore the contribution of intrinsic noise from different biochemical reactions in a system.
- To understand the effects of the irregularity of parameters in a viral infection model and thus provide a theoretical basis of new anti-virus strategy.
9.3 Future directions

In the new arena emerging from molecular biology and mathematics, our work merely touches the surface of modelling the complex regulatory networks and dynamics in systems biology. There are several directions in which to extend and improve the models presented in the thesis.

The circadian clock model

- The current model required the assumption that gene regulation is primarily accomplished through transcription regulation. Although the reconstructed model has a good predictive performance for some phenotypes, there are still a large number of phenotypes to be explained. A more complete model could include more detailed post-transcriptional and post-translational regulations, such as phosphorylation of PER by DBT and CK2 (Sathyanarayanan, Zheng et al. 2004) and phosphorylation of TIM by SGG (Martinek, Inonog et al. 2001). Also, inclusion of separate compartments is necessary in an extended model, as discussed in Chapter 5.

- The differential equations are assumed to follow the mass action rate law. The assumption has been used because it can simplify the equations greatly. However, it is generally considered that for reactions of the model that are catalysed – which is always the case in biochemical reactions – it is more justifiable to think of them as Michaelian. The most striking problem with using the mass action law is that mass action kinetics is unbounded and it tends to infinity when the concentrations of substrates also tend to infinity. A comparative study of the current system described by the mass action rate law and by Michaelis-Menten kinetics could direct the choice of the kinetic approach in future research.

- Parameter estimation is a bottleneck in the model development. The initial parameters are estimated by hand, still a common practice among the system biologists (Hynne, Dano et al. 2001; Forger and Peskin 2003). Without a search of...
the complete parameter space, no ‘optimal’ parameters can be guaranteed. So we can only claim that our model equations and this particular parameter set are sufficient to account for many properties of the system studied. More sophisticated global algorithms need to be developed to replace or supplement the hand crafting of models (Moles, Mendes et al. 2003).

- Nearly all the organisms which have developed circadian rhythms share a great similarity in their circadian clocks. Among them, *Drosophila* is one of the most well-researched species. A comparative study between the circadian clock in *Drosophila* and other species could facilitate the understanding of the structures of the circadian clocks for the different species. Indeed, there has already been a successful model developed in which gene *gigantea* was found to be a candidate in a new feedback loop in the *Arabidopsis* circadian clock by comparing it with the circadian clocks in *Drosophila* and mammals (Locke, Southern et al. 2005).

- The choice of modelling approach is normally dictated by the information available in the literature. Our choice of the ODE was based on the quantitative data of gene expression and protein levels available. However, as the parameters of the biochemical processes are all unknown, qualitative approaches, such as Boolean networks, could be used. Furthermore, a thorough study and implementation of mathematical models using the stochastic kinetics approach could be attempted to explore stochastic effects in terms of internal noise and external perturbations.

Viral infection model

- Identification of valid regions of CLE. For the current model, CLE appeared not to be valid in the system starting from low EOI. However, the assumptions underlying the approximation and exact conditions for the validity of CLE are unclear. Little has been done to determine when it is legitimate to use such an approach. A systematic investigation of its regions of validity should be identified for making fast decisions about using CLE approximations instead of comparing it with the Gillespie algorithm over a large number of realisations.
The generation of more accurate approximate simulation algorithms. The Gillespie algorithm accounts for every possible discrete reaction event and captures the intrinsic fluctuations. However, despite this exactness it becomes extremely inefficient when there is a mixture of small and large numbers of species or slow and fast reactions in a system. Recently, great efforts have been put into multi-scale stochastic modelling allowing adequate levels of description for different species and different reactions. Some researchers (Haseltine and Rawlings 2002; Rao and Arkin 2003), for example, have considered partitioning the system into two subsets depending on the basis of propensity function values: fast and slow reaction subsets. Then they constructed the CME which describes the joint probability density functions of the number and the occurrences of both subsets. Cao et al. (2005) have proposed a virtual fast system which is Markovian rather than the real fast system being Non-Markovian. The authors used the stationary (asymptotic) properties of the virtual fast system to make a stochastic algorithm to construct the slow-scale reactions, which the authors argued, is more reliable and transparent. Overall, multi-scale stochastic modelling is certainly a worthwhile exercise in the future when the scale of modelled systems becomes large.

The current study used numerical solution of CME to explore the intrinsic noise. Alternatively, we can also derive Fokker-Planck equations for the evolution of its various moments (Kampen 2001). These moments form an infinite chain, and generally only the first and second moments have to be used for the calculation of mean and noise level. A comparison study of the current system using Fokker-Planck equations can be carried out.

For extrinsic noise, the parameter $k_4$ amplifies the noise to a considerable extent. This parameter is, however, the slowest. This leads to a question whether this parameter is also the one for which the steady state level (e.g. of mRNA) is the most sensitive to. A metabolic control analysis of the steady state of the model would have been able to answer this question.
• To develop a stochastic framework incorporating both the intrinsic and the extrinsic noise for large scale biochemical systems, as discussed in Chapter 8.

9.4 Conclusion

The study of highly dynamic, interacting and complex biological systems is a challenging topic. Our understanding of the dynamics and functions of the underlying biological and biochemical processes has been hampered by the complexity of the system. However, modelling and simulation-based approaches have the potential to assist in understanding such processes. Kinetics modelling, an important type of mathematical modelling, plays a rigorous and reliable role in revealing the complexities of biological networks. In this thesis, we have shown how kinetic models can be built using mathematical knowledge based on biological knowledge and how the models can be used to analyse properties of gene regulation.

Although our investigation focuses on a small subset of specific problems, there is indeed a large array of challenging and exciting biological phenomena awaiting exploration. It is our hope that the success of the models presented in the thesis will encourage more biologists and mathematicians to look beyond the difficulties of interdisciplinary work and to investigate the benefits of integrative approaches to systems biology.
Reference


Qiu, J. and P. E. Hardin (1996). "per mRNA cycling is locked to lights-off under photoperiodic conditions that support circadian feedback loop function." Molecular and cellular biology 16(8): 4182-8.


Appendices

Appendix A. Mass action rate law and Michaelis-Menten kinetics

A.1 Mass action rate law

The mass action rate law describes the relationship between reaction rates and molecular component concentrations in typical elementary reactions where no intermediates are formed (Waage 1864). It reveals that the instantaneous reaction rate in a spatially homogeneous medium is directly proportional to the product of the effective concentrations of each participating molecule. For a simple reaction like

\[ aA + bB \xrightarrow{k} cC, \]  \hspace{1cm} (9.1)

where \( A \) and \( B \) denote the reacting compounds, \( C \) denotes the product compound, \( a, b \) and \( c \) represent the number of molecules, and \( k \) is temperature dependent rate constant of the reaction. The reaction rate reads according to the mass action rate law

\[ V = k[A]^a[B]^b, \]  \hspace{1cm} (9.2)

where \( V \) is the reaction rate and \([ \ )\] denotes concentrations. Once \( V \) is known, we can write a set of differential equations to describe the dynamics of concentrations of all the components in the system:

\[ \frac{d[A]}{dt} = -a \ V, \]  \hspace{1cm} (9.3)
\[
\frac{d[B]}{dt} = -b \, V, \quad (9.4)
\]
\[
\frac{d[C]}{dt} = c \, V, \quad (9.5)
\]

The powers of the reactant concentrations \(a\) and \(b\) are called the orders of the reaction in \(A\) and \(B\), respectively. The total order of the reaction is defined as the sum of \(a\) and \(b\).

### A.2 Michaelis-Menten kinetics

In contrast to the mass action rate law, which is only valid for the elementary reactions, Michaelis-Menten kinetics is a special reaction system, and is highly important in biochemical systems (Michaelis and Menten 1913). A basic enzyme reaction converts a substrate to a product only in the presence of a catalyst. Here we consider an enzyme reaction first proposed by Brown et al. (1902),

\[
S + E \xrightleftharpoons[k_2]{k_1} ES \xrightarrow{k_3} P + E, \quad (9.6)
\]

where \(S\) is a substrate, \(P\) is a product, \(E\) is a catalyst or an enzyme, and \(ES\) is an enzyme-substrate complex. According to the mass action rate law, the rate of production of \(S\), denoted by \([S+]\), can be formulated

\[
\frac{d[S_+]}{dt} = k_2 [ES]. \quad (9.7)
\]

The rate of elimination of \(S\), denoted by \([S]\), is

\[
\frac{d[S]}{dt} = k_1 [E] [S]. \quad (9.8)
\]
Therefore, the rate of concentration changes of $S$ can be obtained if we combine Eqs. (9.7) and (9.8),

$$\frac{d[S]}{dt} = \frac{d[S_\text{r}]}{dt} + \frac{d[S_\text{i}]}{dt} = k_2 [ES] - k_1 [S][E]. \quad (9.9)$$

Applying the same principle, similar expressions for the change of concentration of $E$ and $ES$ can be found,

$$\frac{d[E]}{dt} = (k_2 + k_3)[ES] - k_1[E][S], \quad (9.10)$$

$$\frac{d[ES]}{dt} = k_1[E][S] - (k_2 + k_3)[ES], \quad (9.11)$$

$$\frac{d[P]}{dt} = k_3[ES]. \quad (9.12)$$

Michaelis-Menten kinetics is based on the assumption that the concentration of $ES$ remains constant (quasi-steady state)

$$\frac{d[ES]}{dt} = 0. \quad (9.13)$$

Consequently,

$$[ES] = \frac{1}{K_M}[E][S], \text{ where } K_M = \frac{k_2 + k_3}{k_1}, \quad (9.14)$$

where $K_M$ is the Michaelis-Menten constant. If we define the total enzyme concentration $[E_T]$ as the sum of the concentrations of uncombined enzyme $E$ and complex $ES$, that is

$$[E_T] = [E] + [ES]. \quad (9.15)$$
Using Eq. (9.15), Eq. (9.14) can be written as

\[
[ES] = \frac{([E_r] - [ES])[S]}{K_m}.
\]

(9.16)

Eq. (9.16) can be rearranged as

\[
[ES] = [E_r] \frac{[S]}{[S] + K_m}.
\]

(9.17)

Together with Eq. (9.12) and (9.17), the rate \( V_p \equiv d[P]/dt \) for the production \( P \) can be written

\[
V_p = k_3 [ES] = k_3 [E_r] \frac{[S]}{[S] + K_m}, \text{ or }
\]

\[
V_p = V_{\text{max}} \frac{[S]}{[S] + K_M}, \text{ where } V_{\text{max}} = k_3 [E_r],
\]

(9.18)

where \( V_{\text{max}} \) is the maximum reaction speed corresponding to the saturated case in which all the product \( P \) has been converted to the complex \( ES \). \( K_M \) is the temperature dependent Michaelis-Menten constant, as defined in Eq. (9.14).
Appendix B. Stochastic processes

Before going into the heavy mathematical part of chemical stochastic simulation, it is necessary to introduce some background of the stochastic processes directly required for the work in this thesis. All of this material is widely known and there are a number of books that provide detailed probability theory and the stochastic processes theory (Ross 2003). Gillespie’s book (1992) is particularly well written for the introduction of Markov processes. Kampen’s book (2001) gives a comprehensive description of stochastic processes in the context of physics and chemistry.

B.1 Random variables

A variable is an entity that has a value that we can measure or “sample”. If a variable $X$ is determined by its sampling context, it is called a “sure variable”. If and only if there exists a function $P$ of a variable $x$ such that $P(x)dx$ equals, to first order in $\Delta x$, the probability of finding the value of $X$ in the interval $[x, x + \Delta x]$, $X$ is called a “random variable”. In symbols, $P(x)dx = \Pr\{X \in [x, x + dx]\}$. $P$ is the density function of $X$. A random variable is completely specified by its density function, and there are many different random variables $X$ according to their density function $P$.

The average of any function $h$ with respect to the random variable $X$ is denoted by $\langle h(X) \rangle$, and is defined by

$$\langle h(X) \rangle \equiv \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} h(x^{(i)})$$ (10.1)

where $x^{(i)}$ is the value assumed by $X$ in $N$ independent samplings.
The average $\langle X^n \rangle$ is called the $n^{th}$ moment of $X$, or sometimes the $n^{th}$ moment of $P$. The mean of $X$ is defined as the first moment of $X$ and the variance of $X$ is defined as

$$\text{var}\{X\} \equiv \langle (X - \langle X \rangle)^2 \rangle \equiv \langle X^2 \rangle - \langle X \rangle^2. \quad (10.2)$$

Suppose there is a time-evolving system $(t_0 < t_1 < t_2 < \ldots < t_n)$ whose possible state of the system at time $t$ can be defined by $X(t)$, and the value of $X$ at the initial time $t_0$ is fixed, $X(t_0) = x_0$. The conditional probability

$$P_{n-1}^{j+1}(x_n, t_n; \ldots; x_{j+1}, t_{j+1} \mid x_j, t_j; \ldots; x_1, t_1; x_0, t_0) \quad (10.3)$$

is defined as the joint density function for the states $X_n, \ldots, X_{j+1}$ at time $t_n, \ldots, t_{j+1}$ on condition that the states $X_j, \ldots, X_0$ have been passed at time $t_j, \ldots, t_0$.

### B.2 Markov processes

A Markov process is defined as a stochastic process satisfying the Markov property,

$$P_i^j(x_j, t_j \mid x_{j-1}, t_{j-1}; \ldots; x_1, t_1; x_0, t_0)$$

$$= P_i^{(1)}(x_j, t_j \mid x_{j-1}, t_{j-1}) \equiv P(x_j, t_j \mid x_{j-1}, t_{j-1}). \quad (10.4)$$

$P$ is called the transition probability. The Markov property indicates only the most recent conditioning matters, therefore, the transition probability from state $X_{j-1}$ at time $t_{j-1}$ to state $X_j$ at time $t_j$ is determined by the knowledge of state at time $t_{j-1}$. For future reference, the formal definition of the Markov transition probability in (10.4) is simplified as
The increment in $X$ from any time $t$ to any infinitesimally later time $t + dt$ depends only on $t$, $dt$, and value of $X$ at $t$, therefore, we can define the increment in $X$ as propagator of the process $X(t)$,

$$\Xi(dt; x, t) \equiv X(t + dt) - X(t), \text{ given } X(t) = x. \quad (10.6)$$

If a Markov process has a property that $\Xi(dt; x, t) \to 0$ as $dt \to 0$ for all $x$ and $t$, it is called “continuous”. Obviously, in continuous Markov processes $X(t)$ has to be a real number because it evolves in a continuous, gradual manner; A jump Markov process is one in which the propagator $\Xi(dt; x, t)$ is usually exactly zero but occasionally finitely different from zero. However, in the jump Markov processes $X(t)$ can be either a continuum of real numbers or a discrete subset of the real numbers, such as the integers.

From the Markov property (10.4), the Chapman-Kolmogorov equation can be directly derived. For continuous Markov processes, it can be written as

$$P(x_2, t_2 \mid x_1, t_1) dx_2 = \int_{-\infty}^{\infty} P(x_2, t_2 \mid x_1, t_1) P(x_1, t_1 \mid x_0, t_0) dx_1. \quad (10.7)$$

The discrete state version of the Chapman-Kolmogorov equation can be expressed as

$$P(x_2, t_2 \mid x_0, t_0) = \sum_{x_1 = -\infty}^{\infty} P(x_2, t_2 \mid x_1, t_1) P(x_1, t_1 \mid x_0, t_0). \quad (10.8)$$
For the jump Markov processes with discrete state, Eq. (10.8) can be written in the form

\[
P(x, t + dt | x_0, t_0) = \sum_{v=\infty}^{\infty} P(x, t + dt | x - v, t)P(x - v, t | x_0, t_0),
\]

where \(v\) is the state of change vector. If we define a probability function \(\omega(v | n, v)\) and a smooth function of \(t\), \(\alpha(n, t)dt\), then Eq. (10.9) can be converted into the form

\[
P(x, t + dt | x_0, t_0) = \sum_{v=\infty}^{\infty} [\alpha(x - v, t) dt \omega(v | x - v, t)] P(x - v, t | x_0, t_0)
+ [1 - \alpha(x, t) dt] P(x, t | x_0, t_0).
\]

where,

\[
\alpha(x, t)dt \equiv \text{probability, given } X(t) = x, \text{ that the process will jump away from states } x \text{ in the next infinitesimal time interval } [t + dt],
\]

and,

\[
\omega(v | x, t) \equiv \text{probability that the process, upon jumping away from state } x \text{ at time } t, \text{ will land in state } x + v.
\]

If we take the limit \(dt \to 0\), we get

\[
\lim_{dt \to 0} \frac{P(x, t + dt) - P(x, t)}{dt} = \frac{\partial P(x, t)}{\partial t}.
\]
and we can obtain the differential-difference equation for $P(x,t | x_0, t_0)$. This equation is called the master equation, which is characterised by the functions $\alpha$ and $\omega$. The master equation is given as follows,

$$
\frac{\partial}{\partial t} P(x,t | x_0, t_0) = \sum_{v=\infty}^{\infty} \left[ \alpha(x-v,t) \omega(v | x-v,t) \right] P(x-v,t | x_0, t_0) - \alpha(x,t) P(x,t | x_0, t_0).
$$

(10.14)

### B.4 Langevin equation

- **Univariate Langevin equation**

The definition of continuous Markov processes in Section B.2 implies that the propagator must have the analytical form:

$$
\Xi(dt; x, t) = A(x,t)dt + D^{1/2}(x,t)\mathcal{N}(t)(dt)^{1/2},
$$

(10.15)

where $A(x,t)$ and $D(x,t)$ are any smooth functions, with $D(x,t)$ non-negative, $\mathcal{N}(t)$ is a temporally uncorrelated unit normal random variable, $\mathcal{N}(0,1)$.

**Proofs of Eq. (10.15):**

In order to derive Eq. (10.15), the time interval $dt$ is divided into $n$ subintervals of equal length $dt/n$ at the starting points $t_i = t_{i-1} + i(dt/n)$, $(i = 0,...,n)$. For the self-consistency of the Markov processes, it must have the form

$$
\Xi(dt; X(t), t) = \sum_{i=1}^{n} \Xi(dt/n; X(t_{i-1}), t_{i-1})).
$$

(10.16)
Since in continuous Markov processes \( dt \) is so small as to be close enough to \( t \), it can replace Eq. (10.16), that \( t_{i-1} \to t \) and \( X(t_{i-1}) \to X(t) \equiv x \). We therefore conclude, by giving, at least to the lowest order in \( dt \), Eq. (10.16) can result in

\[
\Xi(dt;x,t) = \sum_{i=1}^{n} \Xi_i(dt/n;x,t), \tag{10.17}
\]

where \( \Xi_1, \ldots, \Xi_n \) are statistically independent of each other. According to the central limit theorem, the sum of these \( n \) statistically independent random variables is normally distributed if we take \( n \to \infty \). According to the normal sum theorem, the mean and the variance of the sum of the random variables is equal to the sum of their respective mean and variance, resulting in

\[
\langle \Xi(dt;x,t) \rangle = n \times \langle \Xi_i(dt/n;x,t) \rangle, \tag{10.18}
\]

\[
\text{var}\{\Xi(dt;x,t)\} = n \times \text{var}\{\Xi_i(dt/n;x,t)\}. \tag{10.19}
\]

This can easily prove that the mean and variance of \( \Xi(dt;x,t) \) must both be linear in \( dt \); i.e.,

\[
\langle \Xi(dt;x,t) \rangle = A(x,t)dt, \tag{10.20}
\]

\[
\text{var}\{\Xi(dt;x,t)\} = D(x,t)dt. \tag{10.21}
\]

By now, we have proved Eq. (10.15), if we insert it to Eq. (10.6), we obtain the standard form of the Langevin equation.

\[
X(t + dt) = X(t) + A(X(t),t)dt + D^{1/2}(X(t),t)\mathcal{N}(t)(dt)^{1/2}. \tag{10.22}
\]
The function $A(x,t)$ is called the drift function of the process and the function $D(x,t)$ is called the diffusion function. Once these two functions are known, Eq. (10.22) tells us the value of $X(t + dt)$ from $X(t)$.

Gaussian white noise is defined as $\xi(t) = \lim_{dt \to 0} N(0, 1/dt)$, then Eq. (10.22) can be brought into the white-noise form Langevin equation

$$\frac{dX(t)}{dt} = A(X(t), t) + D^{1/2}(X(t), t)\xi(t). \quad (10.23)$$

- **Multivariate Langevin equation**

Let us suppose a general $M$-variate continuous Markov process $X(t) \equiv [X_1(t), \ldots, X_M(t)]$ fulfilling the following conditions:

1. The increment $\Xi_i(dt; x, t)$ in each component $X_i$, given $X(t) = x$, depends only on $dt$, $x$, and $t$.

2. All probability density functions $\Xi_i$ are continuously differentiable with respect to their parameters, $dt$, $x$, and $t$.

3. $\Xi_i(dt; x, t) \to 0$ as $dt \to 0$.

4. The mean and variance of each random variable $\Xi_i$ has been well defined.

With the definition of a multivariate continuous Markov process, the increments $\Xi_i$ have the analytical forms by using a similar procedure as the proofs of Eq. (10.15). By replacing $\Xi_i$ with $X_i(t + dt) - X_i(t)$, we can get the standard form of multivariate Langevin equation,
\[ X_i(t + dt) = X_i(t) + A_i(x,t)dt + \sum_{j=1}^{M} b_{ij}(x,t) \mathcal{N}_j(t)(dt)^{1/2}, \ (j = 1,\ldots,M), \ (10.24) \]

where \( A_i \) and \( b_{ij} \) are any smooth functions, and \( \mathcal{N}_1(t),\ldots,\mathcal{N}_M(t) \) are \( M \) statistically independent, temporally uncorrelated, unit normal random variables.
Appendix C. Biochemical reactions of the circadian clock model

1. Binding processes of a transcription factor to promoters.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Kinetic law</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Pr_{\text{cper}} \rightarrow (1 - Pr_{\text{cper}}) \times bccper_p \times CC$</td>
<td></td>
</tr>
<tr>
<td>$Pr_{\text{ct}} \rightarrow (1 - Pr_{\text{ct}}) \times bctime_p \times CC$</td>
<td></td>
</tr>
<tr>
<td>$Pr_{\text{cv}} \rightarrow (1 - Pr_{\text{cv}}) \times bccvri_p \times CC$</td>
<td></td>
</tr>
<tr>
<td>$Pr_{\text{cpdp}} \rightarrow (1 - Pr_{\text{cpdp}}) \times bccpdp_p \times CC$</td>
<td></td>
</tr>
<tr>
<td>$Pr_{\text{vc}} \rightarrow (1 - Pr_{\text{vc}} - Pr_{\text{pc}}) \times bvriclk_p \times VRI$</td>
<td></td>
</tr>
<tr>
<td>$Pr_{\text{pc}} \rightarrow (1 - Pr_{\text{vc}} - Pr_{\text{pc}}) \times bpdpcclk_p \times PDP$</td>
<td></td>
</tr>
</tbody>
</table>

2. Unbinding processes of a transcription factor from promoters.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Kinetic law</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Pr_{\text{cper}} \rightarrow Pr_{\text{cper}} \times ubccper_p$</td>
<td></td>
</tr>
<tr>
<td>$Pr_{\text{ct}} \rightarrow Pr_{\text{ct}} \times ubctime_p$</td>
<td></td>
</tr>
<tr>
<td>$Pr_{\text{cv}} \rightarrow Pr_{\text{cv}} \times ubccvri_p$</td>
<td></td>
</tr>
<tr>
<td>$Pr_{\text{cpdp}} \rightarrow Pr_{\text{cpdp}} \times ubccpdp_p$</td>
<td></td>
</tr>
<tr>
<td>$Pr_{\text{vc}} \rightarrow Pr_{\text{vc}} \times bvriclk_p$</td>
<td></td>
</tr>
<tr>
<td>$Pr_{\text{pc}} \rightarrow Pr_{\text{pc}} \times bpdpcclk_p$</td>
<td></td>
</tr>
</tbody>
</table>
3. Transcription processes of the *per, tim, vri, pdp1, clk* genes into the corresponding mRNAs.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Kinetic law</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{per}_p \rightarrow \text{per}_p + \text{per}_m )</td>
<td>((1 - (1 - \text{Pr}_{\text{per}})^{\text{m}}) \times t\text{tecccpt}<em>p + (1 - \text{Pr}</em>{\text{per}})^{\text{m}} \times t\text{dvpmt}\times \text{per}_p )</td>
</tr>
<tr>
<td>( \text{tim}_p \rightarrow \text{tim}_p + \text{tim}_m )</td>
<td>((1 - (1 - \text{Pr}_{\text{tim}})^{\text{m}}) \times t\text{ccctim}<em>p + (1 - \text{Pr}</em>{\text{tim}})^{\text{m}} \times t\text{dvpmt}\times \text{tim}_p )</td>
</tr>
<tr>
<td>( \text{vri}_p \rightarrow \text{vri}_p + \text{vri}_m )</td>
<td>((1 - (1 - \text{Pr}_{\text{vri}})^{\text{m}}) \times t\text{cccvri}<em>p + (1 - \text{Pr}</em>{\text{vri}})^{\text{m}} \times t\text{dvpmt}\times \text{vri}_p )</td>
</tr>
<tr>
<td>( \text{pdp}_p \rightarrow \text{pdp}_p + \text{pdp}_m )</td>
<td>((1 - (1 - \text{Pr}_{\text{pdp}})^{\text{m}}) \times t\text{cccpdp}<em>p + (1 - \text{Pr}</em>{\text{pdp}})^{\text{m}} \times t\text{dvpmt}\times \text{pdp}_p )</td>
</tr>
<tr>
<td>( \text{clk}_p \rightarrow \text{clk}_p + \text{clk}_m )</td>
<td>((\text{Pr}<em>{\text{vri}} \times \text{tcvriclk}<em>p + \text{Pr}</em>{\text{pdp}} \times \text{tcpdpclklk}<em>p + (1 - \text{Pr}</em>{\text{vri}} - \text{Pr}</em>{\text{pdp}}) \times t\text{ccclk}_p) \times \text{clk}_p )</td>
</tr>
</tbody>
</table>

4. Translation processes of these mRNAs into the proteins.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Kinetic law</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{per}_m \rightarrow \text{per}_m + \text{PER} )</td>
<td>( t\text{lper} \times \text{per}_m )</td>
</tr>
<tr>
<td>( \text{tim}_m \rightarrow \text{tim}_m + \text{TIM} )</td>
<td>( t\text{ltim} \times \text{tim}_m )</td>
</tr>
<tr>
<td>( \text{vri}_m \rightarrow \text{vri}_m + \text{VRI} )</td>
<td>( t\text{lvri} \times \text{vri}_m )</td>
</tr>
<tr>
<td>( \text{pdp}_m \rightarrow \text{pdp}_m + \text{PDP} )</td>
<td>( t\text{lpdp} \times \text{pdp}_m )</td>
</tr>
<tr>
<td>( \text{clk}_m \rightarrow \text{clk}_m + \text{CLK} )</td>
<td>( t\text{lclk} \times \text{clk}_m )</td>
</tr>
</tbody>
</table>

5. Association processes of PER and TIM, CLK and CYC, and PERTIM and CLKCYC.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Kinetic law</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{PER} + \text{TIM} \rightarrow \text{PT} )</td>
<td>( b\text{pt} \times \text{PER} \times \text{TIM} )</td>
</tr>
<tr>
<td>( \text{CLK} + \text{CYC} \rightarrow \text{CC} )</td>
<td>( b\text{cc} \times \text{CLK} \times \text{CYC} )</td>
</tr>
<tr>
<td>( \text{PT} + \text{CC} \rightarrow \text{CCPT} )</td>
<td>( b\text{ccpt} \times \text{PT} \times \text{CC} )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Kinetic law</th>
</tr>
</thead>
<tbody>
<tr>
<td>$PT \rightarrow PER + TIM$</td>
<td>$ubpt \times PT$</td>
</tr>
<tr>
<td>$CC \rightarrow CLK + CYC$</td>
<td>$ubcc \times CC$</td>
</tr>
<tr>
<td>$CCPT \rightarrow CC + PT$</td>
<td>$ubccpt \times CCPT$</td>
</tr>
</tbody>
</table>

7. Degradation processes of mRNAs, proteins and complexes.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Kinetic law</th>
</tr>
</thead>
<tbody>
<tr>
<td>$per_m \rightarrow dperm \times per_m$</td>
<td></td>
</tr>
<tr>
<td>$tim_m \rightarrow dtimm \times per_m$</td>
<td></td>
</tr>
<tr>
<td>$vri_m \rightarrow dvrim \times vri_m$</td>
<td></td>
</tr>
<tr>
<td>$pdp_m \rightarrow dpdpm \times pdp_m$</td>
<td></td>
</tr>
<tr>
<td>$clk_m \rightarrow dclkm \times clk_m$</td>
<td></td>
</tr>
<tr>
<td>$PER \rightarrow dper \times PER$</td>
<td></td>
</tr>
<tr>
<td>$TIM \rightarrow dper \times TIM$</td>
<td></td>
</tr>
<tr>
<td>$VRI \rightarrow dvri \times VRI$</td>
<td></td>
</tr>
<tr>
<td>$PDP \rightarrow dpdp \times PDP$</td>
<td></td>
</tr>
<tr>
<td>$CLK \rightarrow dclk \times CLK$</td>
<td></td>
</tr>
<tr>
<td>$PT \rightarrow dpt \times PT$</td>
<td></td>
</tr>
<tr>
<td>$CC \rightarrow dec \times CC$</td>
<td></td>
</tr>
<tr>
<td>$CCPT \rightarrow decpt \times CCPT$</td>
<td></td>
</tr>
</tbody>
</table>
Appendix D. Sensitivity analysis

Sensitivity analysis is a general purpose technique that is often used to analyse how sensitive a system is with respect to the change of a set of parameter values. Mathematical theories about sensitivity analysis can be found in Varma et al. (1999). In the dynamics of biological networks, the state variables mostly correspond to molecule concentrations and the parameter set consists of biochemical constants related to the system dynamics and initial conditions. The purposes of parameter sensitivity analysis in the biological networks are twofold: (1) It can be used to validate information about insensitivities and critical system parameters and provides a basis for system identification. (2) It allows observations to be made about intrinsic system properties like stability and robustness of the system behaviour with respect to parameter fluctuations in a systematic way (Fedorov and Hackl 1997).

For a model whose dynamics is described by a system of ODEs,

\[ \frac{dx}{dt} = f(x, p, t), \text{ with } x(t_0) = x_0, \]  

(11.1)

where \( x \) is the \( n_s \times 1 \) vector of state variables, \( t \) is the time where \( t \geq t_0 \), and \( p \) is the \( n_p \times 1 \) vector containing the parameters of interest for the system. Suppose the solution of the system is \( x = x(t, p) \). If we change the \( j \)th parameter in the parameter vector \( p \), from \( p_j \) to \( p_j + \Delta p_j \). Then, the corresponding solution for \( x \) becomes

\[ x = x(t, p_j + \Delta p_j). \]  

(11.2)

Since \( x \) is a continuous function of \( p_j \), the current solution can be expanded into a Taylor series as follows,
\[ x(t, p_j + \Delta p_j) = x(t, p_j) + \left( \frac{\partial x(t, p_j)}{\partial p_j} \right) \Delta p_j + \left( \frac{\partial^2 x(t, p_j + \theta \cdot \Delta p_j)}{\partial p_j^2} \right) \frac{\Delta p_j^2}{2}, \quad (11.3) \]

where \( 0 < \theta < 1 \). If \( \Delta p_j \) is sufficiently small, i.e., \( \Delta p_j \ll p_j \), the Taylor series can be truncated after the linear term, leading to

\[ \Delta x = x(t, p_j + \Delta p_j) - x(t, p_j) \approx \left( \frac{\partial x(t, p_j)}{\partial p_j} \right) \Delta p_j, \quad (11.4) \]

where \( \Delta x \) represents the variation of \( x \) due to the change of the input parameter \( p_j \), given by \( \Delta p_j \). If we consider an infinitesimal variation \( (\Delta p_j \to 0) \), the parameter sensitivity with respect to the system’s states along a specific trajectory \( s(x; p_j) \), which is the \( n_s \times n_p \) matrix of state sensitivity, is defined by

\[ s(x; p_j) = \left( \frac{dx(t, p_j)}{\partial p_j} \right) = \lim_{\Delta p_j \to 0} \frac{x(t, p_j + \Delta p_j) - x(t, p_j)}{\Delta p_j}. \quad (11.5) \]

This is also defined as the first-order local sensitivity, or simply local sensitivity of the dependent variable, \( x \), with respect to the input parameter, \( p_j \). Although higher-order local sensitivities can be defined in a similar fashion, we will limit the treatment to first-order local sensitivities, since most applications are based on linear sensitivity analysis (Varma, Morbidelli et al. 1999). Sometimes the local sensitivity is also called the absolute sensitivity. It is noticeable that the local parameter sensitivities are valid only in the neighbourhood of a specific parameter set. Thus, they provide information on the robustness of a particular model with a particular parameter.

Another quantity related to local sensitivity, commonly used in sensitivity analysis, is the normalised sensitivity, or relative sensitivity, defined as
\[
S(x; p_j) = \frac{p_j}{x} \cdot \frac{\partial x}{\partial p_j} = \frac{\partial}{\partial \ln x} \frac{\frac{\partial}{\partial x} x}{p_j} = \frac{p_j}{x} \cdot s(x; p_j). \quad (11.6)
\]

In most cases, the relative sensitivities are more meaningful. Based on this, the relative sensitivity matrix will be used in the thesis.

In oscillating systems, the primary interest of parameter sensitivity is generally period sensitivities which capture the change of period length upon changes in parameters. Suppose \( \tau(p) \) defines the period of the system for a given parameter \( p \). According to Eq. (11.5), the absolute period sensitivity to the parameter \( p_j \) is defined by

\[
s(\tau; p_j) = \frac{d\tau(t, p_j)}{\partial p_j} = \lim_{\Delta p_j \to 0} \frac{\tau(t, p_j + \Delta p_j) - \tau(t, p_j)}{\Delta p_j}, \quad (11.7)
\]

and the normalised sensitivity index is defined by

\[
S(\tau; p_j) = \frac{p_j}{\tau(p_j)} \cdot s(\tau; p_j). \quad (11.8)
\]

Eqs. (11.7) and (11.8) have been both used (Varma, Morbidelli et al. 1999; Stelling, Gilles et al. 2004) for analysing period sensitivity of oscillations in chemical and biochemical systems.
Appendix E. Programming

E.1 The deterministic circadian clock model

Main.m

% This is the main program for the deterministic circadian clock model;
% other files are needed to put in the same folder to run this script
%
% init_det.dat: initial values
% det_sol.m: ODE solver
% stats.m: define the states of the system
% parameters.m: define the rate constants
% reactions.m: define the reactions
% stoich.m: define the stoichiometry

clear;

% simulation time
Tfinal = 72;

% initial state of species in the system
load init_det.dat;
init = init_det;

timespan = [0, Tfinal];
[t, y] = ode15s(@det_sol, timespan, init);

% extraction of results
perm = y(:, 12);
timm = y(:, 13);
vrim = y(:, 14);
CLKm = y(:, 15);
pdpm = y(:, 16);
PER = y(:, 17);
TIM = y(:, 18);
VRI = y(:, 19);
PDP = y(:, 20);
CLK = y(:, 21);
PT = y(:, 22);
CC = y(:, 23);

% plots of mRNAs
figure
[ax, h1, h2] = plotyy(t, [perm, timm, vrim, pdpm], t, CLKm);
set(h1, 'color', 'blue')
set(h1, {'linestyle'}, {'--'; '--'; '--'; '--'})
axes(ax(1))
set(ax(1), 'ycolor', 'blue')
ylabel('Concentration of vri, pdp1, per and tim mRNAs (nM)')
set(h2, 'color', 'red')
set(h2, {'linestyle'}, {'-'})
axes(ax(2))
set(ax(2), 'ycolor', 'red')
ylabel('Concentration of clk mRNA (nM)')
legend([h1; h2], 'perm', 'timm', 'vrim', 'pdpm', 'CLKm')
title('mRNAs')
xlabel('Time(h)'),
% plots of proteins

figure

[ax, h1, h2] = plotyy(t, [PER, TIM, VRI, CLK], t, PDP);
set(h1, 'color', 'blue')
set(h1, {'linestyle'}, {'--'; '--'; '-'; '-.'})
axes(ax(1))
set(ax(1), 'ycolor', 'blue')
ylabel('Concentration of VRI, CLK, PER and TIM proteins (nM)')

set(h2, 'color', 'red')
set(h2, {'linestyle'}, {':'})
axes(ax(2))
ylabel('Concentration of PDP1 protein (nM)')

legend([h1; h2], 'PER', 'TIM', 'VRI', 'CLK', 'PDP')
title('Proteins'),
xlabel('Time(h)'),

init_det.dat

0.003185;  %clkp
0.003185;  %perp
0.003185;  %vrip
0.003185;  %pdpp
0.003185;  %tmp
0.0431;  %prcper
0.0585;  %prcv
0.08;  %prcpdp
0.489;  %prvc
0.426;  %prpc
0.043;  %prct
0.2395;  %perm
0.2395;  %tmp
0.2571;  %vrim
0.2583;  %clkm
0.3175;  %pdpm
2.7527;  %PER
2.7527;  %TIM
3.175;  %VRI
4.1953;  %PDP
3.6628;  %CLK
0.4014;  %PT
0.5566;  %CC
0.4982;  %CCPT
0 %time

Det_sol.m

function dydt = det_sol(t,y),

% Define states
states;

% Input parameters
parameters;

%Define reactions
reactions;

%stoichiometry
stoich;

%ODE
dydt = ydot;
states.m

%species

%--------- promoters--------
clkp = y(1);
perp = y(2);
vrip = y(3);
pdpp = y(4);
timp = y(5);

%--------- probability of binding promoters ----
prcper = y(6);
prcv = y(7);
prcpdp = y(8);
prvc = y(9);
prpc = y(10);
prct = y(11);

%--------mRNAs--------
perm = y(12);
timm = y(13);
vrim = y(14);
clkm = y(15);
pdpm = y(16);

%--------proteins--------
PER = y(17);
TIM = y(18);
VRI = y(19);
PDP = y(20);
CLK = y(21);

%--------complexes--------
PT = y(22);
CC = y(23);
CCPT = y(24);

Parameters.m

% rate constants. note CYC is a constant, so is treated as a parameter here
% there is a constraint that CYC*bcc=2.349. If using CYC = 100, bcc should
% be 0.02349 accordingly.
CYC = 1;

% parameters
dccpt = 15.122;
bccperp = 0.069;
ubccperp = 0.262;
bccpt = 51;
dperm = 0.053;
tlper = 36;
tccperp = 11;
tcdvpmt = 0.053;
dvri = 1.226;
tccclkp = 1.42;
bcc = 2.349;
dclk = 0.2;
ubcc = 0.89;
bpt = 1.1;
ubpt = 2.93;
dper = 0.62;
tccvrip = 16.86;
tccpdpp = 9.831;
dvrim = 0.07;
dpdp = 0.06;
ubccpt = 7.89;
tlvri = 14.68;
tlpdp = 1.87;
bcvrip = 0.1;
bcpdpp = 0.062;
ubcvrip = 0.276;
ubcpdpp = 0.145;
tcpxdcclkp = 125.54;
dclkm = 0.643;
bvriclkp = 1.858;
bdpcclkp = 1.155;
ubvriclkp = 1.043;
ubdpclkp = 0.952;
tcvriclkp = 0.028;
dpdp = 0.156;
slclk = 35;
cc = 0.184;
dprt = 0.279;
dtim = 0.62;
dtimm = 0.053;
tptimm = 36;
bccctimp = 0.069;
ubccctimp = 0.262;
tccctimp = 11;
npt = 5;
nvri = 4;
npdp = 6;

reactions.m

% define reactions

R = zeros(42,1);
R(1) = CC * PT * bccpt;
R(2) = CC * dcc;
R(3) = CCPT * dccpt;
R(4) = PT * dpt;
R(5) = clkm * dclkm;
R(6) = clkm * slclkm;
R(7) = perm * dperm;
R(8) = perm * tiper;
R(9) = (1 - power(1 - prcper, npt)) * tcccpdpp + power(1 - prcper, npt) * 
tccvrip + power(1 - prcv, nvri) * tccctimp) * vrip;
R(10) = ((1 - power(1 - prcv, nvri)) * tccvrip + power(1 - prcv, nvri) * 
tccctimp) * vrip;
R(11) = vrim * dvrim;
R(12) = vrim * tlvri;
R(13) = vri * dvri;
R(14) = (1 - power(1 - prcpdp, npdp)) * tcccpdpp + power(1 - prcpdp, npdp) * 
tccvrip + power(1 - prcv, nvri) * tccctimp) * vrip;
R(15) = pdp * dpdpm;
R(16) = pdp * tlpdp;
R(17) = PDP * dpdp;
R(18) = (prvc * tcvriclkp + prpc * tcpdpclkp + (1 - prvc - prpc) * tccclkp) * 
clkp;
R(19) = CLK * bcc * CYC;
R(20) = CLK * dclkm;
R(21) = CC * ubcc;
R(22) = PER * TIM * bpt;
R(23) = PT * ubpt;
R(24) = PER * dper;
R(25) = (1 - power(1 - prct, npt)) * tccctimp + power(1 - prct, npt) * 
tccvrip + power(1 - prcv, nvri) * tccctimp) * vrip;
R(26) = timm * dtimm;
R(27) = timm * tptimm;
R(28) = TIM * dtim;
R(29) = CCPT * ubccpt;
\[ R(30) = (1 - \text{prcper}) \times \text{bccperp} \times \text{CC}; \]
\[ R(31) = \text{prcper} \times \text{ubccperp}; \]
\[ R(32) = \text{ubccvrip} \times \text{prcv}; \]
\[ R(33) = (1 - \text{prcv}) \times \text{bccvrip} \times \text{CC}; \]
\[ R(34) = \text{ubccpdpp} \times \text{prcpdp}; \]
\[ R(35) = (1 - \text{prcpdp}) \times \text{bccpdpp} \times \text{CC}; \]
\[ R(36) = (1 - \text{prvc} - \text{prpc}) \times \text{bvriclekp} \times \text{VRI}; \]
\[ R(37) = \text{prvc} \times \text{ubvriclekp}; \]
\[ R(38) = (1 - \text{prvc} - \text{prpc}) \times \text{bpdpcikp} \times \text{PDP}; \]
\[ R(39) = \text{prpc} \times \text{ubpdpcikp}; \]
\[ R(40) = (1 - \text{prct}) \times \text{bcctimp} \times \text{CC}; \]
\[ R(41) = \text{prct} \times \text{ubcctimp}; \]
\[ R(42) = 1; \]

Stoich.m

```matlab
%stoichiometry
ydot = zeros(25,1);

ydot(23) = (-R(1)+R(19)-R(21)+R(29));
ydot(24) = (+R(1)+R(3)-R(29));
ydot(1) = (-R(18)+R(18));
ydot(2) = (-R(9)+R(9));
ydot(15) = (-R(5)+R(6)+R(6)+R(18));
ydot(12) = (-R(7)+R(8)+R(8)+R(9));
ydot(22) = (-R(1)+R(4)+R(22)-R(23)+R(29));
ydot(3) = (-R(10)+R(10));
ydot(14) = (+R(10)-R(11)-R(12)+R(12));
ydot(19) = (+R(12)-R(13));
ydot(4) = (-R(14)+R(14));
ydot(16) = (+R(14)-R(15)+R(16)+R(16));
ydot(20) = (+R(16)-R(17));
ydot(21) = (+R(6)-R(19)-R(20)+R(21));
ydot(17) = (+R(8)-R(22)+R(23)-R(24));
ydot(5) = (-R(25)+R(25));
ydot(13) = (+R(25)-R(26)+R(27)+R(27));
ydot(18) = (-R(22)+R(23)+R(27)-R(28));
ydot(6) = +R(30)-R(31);
ydot(7) = -R(32)+R(33);
ydot(8) = -R(34)+R(35);
ydot(9) = +R(36)-R(37);
ydot(10) = +R(38)-R(39);
ydot(11) = +R(40)-R(41);
ydot(25)=R(42);
```
E.2 The deterministic viral infection model by ODEs

virus_det.m

% virus_det.m is the main program and virus_ODE.m
% defines the ODEs
% Stochastic vs. deterministic modeling of intracellular viral kinetics.

close all;
clear all;

% 1) simulation time

t_0=0;
t_final=200;
tspan=[t_0, t_final];

% 2) initial number of species

R_int=1;
G_int=0;
P_int=0;
V_int=0;
y_init=[R_int G_int P_int V_int];

% 3) rate constants are defined in virus_ODE.m
% don't need change anything below

% simulation

[T,sol]=ode15s('virus_ODE',tspan,y_init);

% extraction of results

R=sol(:,1);
G=sol(:,2);
P=sol(:,3);
V=sol(:,4);

% solution

figure
hold on,
plot(T,R);
plot(T,G,'r');
plot(T,P,'g');
hold off,
title('Deterministic viral Kinetics'),
ylabel('concentrations')

figure
subplot(2,2,1);plot(T,R),legend('R')
subplot(2,2,2);plot(T,G,legend('G'),xlabel('Time(d)'),
subplot(2,2,3);plot(T,P,legend('P'))
subplot(2,2,4);plot(T,V,legend('V'),xlabel('Time(d)'))

virus_ODE.m
% virus_ODE.m defines the ODEs
function dy=virus_ODE(t,y)

dy=zeros(4,1); % initialisation of output

% parameters
k1=0.025;
k2=0.25;
k3=1;
k4=7.5e-6;
k5=1000;
k6=1.99;

% equations
dy(1)=k1*y(2)-k2*y(1);
dy(2)=k3*y(1)-k1*y(2)-k4*y(2)*y(3);
dy(3)=k5*y(1)-k6*y(3)-k4*y(2)*y(3);
dy(4)=k4*y(2)*y(3);

E.3 The viral infection model by the Gillespie algorithm

virus_ssa().m

% Stochastic simulation of intracellular viral infection model
% using Gillespie’s algorithm.
% Stochastic vs. deterministic modeling of intracellular viral kinetics.
% ************************************
% Model:
% Species=[R G P V]
% R:mRNA, G:DNA, P:protein and V:virus
% 6 Reactions:
% 1. G+P ---> V
% 2. G ---> R
% 3. R ---> 0
% (R)
% 4. R ---> G+R
% (R)
% 5. P ---> 0
% (R)
% 6. R ---> R+P
% Note: R is the catalyst in reaction (4) and (6).
% USAGE
% virus(runtimes)
% runtime: the number of realisations
% ************************************

function virus_ssa(runtimes)
tic;

DP=200; % number of data points stored in the output file
rand('state',10000);
% initial output file
outR=zeros(runtimes,DP);
outG=zeros(runtimes,DP);
outP=zeros(runtimes,DP);
outV=zeros(runtimes,DP);
% simulation
for runt=1:runtimes
runt
% user defined system parameters:
% 1) tstop is the time-span for the simulation.
% 2) nTotal is the maximum number of reaction steps to calculate.
% the simulation stops when either tstop or nTotal is reached.
tstop=200.01;
nTotal=1e8;

numspecies=4; % # of species
numreactions=6; % # of reactions
% initial numbers of molecules
numX=[1 0 0 0];
% rate constants
kc=[7.5e-6 0.025 0.25 1 1.99 1000];

% the first element in the output
outR(runt,1)=numX(1);
outG(runt,1)=numX(2);
outP(runt,1)=numX(3);
outV(runt,1)=numX(4);
% Stochiometry matrix
effects = [0  -1  -1  1;
           1  -1   0  0;
          -1   0   0  0;
           0   1   0  0;
           0   0  -1  0;
           0   0   1  0];
% initialization
t = 0; % initial time
tint = 1;
% the main simulation part
for i = 1 : nTotal    %until the max steps are reached (1 for)
% Calculate the redundancy function, h.
h = zeros(1, numreactions);
h(1) = numX(2)*numX(3);
h(2) = numX(2);
h(3) = numX(1);
h(4) = numX(1);
h(5) = numX(3);
h(6) = numX(1);
% calculate time increment tau
amu = h .* kc;
a = sum(amu);
r1=rand;
if (a > 0.0)
  tau = (1 / a) * log(1/r1);
else
  tau = tstop-t;
end

% find out the index of next reaction
r2=rand;
av = 0;

for jj = 1:numreactions % (1 for)
  av = av + amu(jj);
  if (av >= r2 * a) % (1 if)
    ir = jj;
    break;
  end % (1 if)
end % (1 for)

%update the system
t = t + tau;
numX = numX + effects(ir,:);

% keep the number of molecules of each species non-negative
for ii=1:numspecies
  if (numX(ii) < 0)
    numX(ii)=0;
  end
end

%stop the system if the tstop reaches
if t >= tstop
  break;
else
%save the results in the output, mT is the time steps to skip
  if t>tint
    outR(runt,tint+1)=numX(1);
    outG(runt,tint+1)=numX(2);
    outP(runt,tint+1)=numX(3);
    outV(runt,tint+1)=numX(4);
    tint=tint+1;
  end
end
end

% save output files
savefile = strcat('d:\ssa\outR');
save(savefile,'outR')
savefile = strcat('d:\ssa\outG');
save(savefile,'outG')
savefile = strcat('d:\ssa\outP');
save(savefile,'outP')
savefile = strcat('d:\ssa\outV');
save(savefile,'outV')
E.4 The viral infection model by CLE (intrinsic noise)

virus_cle().m

% Effects of intrinsic noise in viral infection model
% Modelled by chemical Langevin equation
% % dX(1) = (k1*X(2)-k2*X(1))*dt+c1*k1*sqrt(X(2))*dW(1)...% X(1)_0 = R_int
% % dX(2) = (k3*X(1)-k1*X(2)-k4*X(2)*X(3))*dt-c3*sqrt(k3*X(1))*dW(3)...% X(2)_0 =G_int
% % dX(3) = (k5*X(1)-k6*X(3)-k4*X(2)*X(3))*dt+c5*sqrt(k5*X(1))*dW(5)...% X(3)_0 =P_int
% % dx(4)=k4*X(2)*X(3)+c4*sqrt(k4*X(2)*X(3))dW(4);% X(4)_0 =V_int
% % X(1)is mRNA, X(2)is genome, X(3) is protein and X(4) is virus
% % USAGE
% []=intrinsic(par1,par2,par3,par4,par5,par6)% par1,...,par6 define intrinsic noise coefficient c1,...,c6.

function []=virus_cle(par1,par2,par3,par4,par5,par6)
clf
tic;
rnd('state',10000);

%system paramters
T = 200; % running time
N = 2000000; % time points to calculate
Delta = T/N; % delta t
DP=200; % data points to save
m=1000; % replications of realisations

% rate constants
k1=0.025;
k2=0.25;
k3=1;
k4=7.5e-6;
k5=1000;
k6=1.99;

% intrinsic noise coefficients
c1=par1;
c2=par2;
c3=par3;
c4=par4;
c5=par5;
c6=par6;

%initial number of species
R_int=5;
G_int=1;
P_int=0;
V_int=0;

% don't need to change anything below
% temp data
Xem1=zeros(1,N);
Xem2=zeros(1,N);
Xem3=zeros(1,N);
Xem4=zeros(1,N);

Xem1(1) = R_int;
Xem2(1) = G_int;
Xem3(1) = P_int;
Xem4(1) = V_int;

% initial index
jj=1;
mT=N/DP; % save every mT point

% output files
outR=zeros(m,DP+1);
outG=zeros(m,DP+1);
outP=zeros(m,DP+1);
outV=zeros(m,DP+1);

% simulation
for jjj=1:m
    jjj
    yout=zeros(4,DP);
    for j = 1:N
        Winc1 = sqrt(Delta)*randn;
        Winc2 = sqrt(Delta)*randn;
        Winc3 = sqrt(Delta)*randn;
        Winc4 = sqrt(Delta)*randn;
        Xem1(j+1) = ((Xem1(j) + (k1*Xem2(j)-k2*Xem1(j))*Delta + ... c1*sqrt(k1*Xem2(j))*Winc1-c2*sqrt(k2*Xem1(j))*Winc2));
        Xem2(j+1) = (Xem2(j) + (k3*Xem1(j)-k1*Xem2(j)-k4*Xem2(j)*Xem3(j))*Delta... - c3*sqrt(k3*Xem1(j))*Winc3-c1*sqrt(k1*Xem2(j))*Winc1- c4*sqrt(k4*Xem2(j)*Xem3(j))*Winc4);
        Xem3(j+1) = (Xem3(j) + (k5*Xem1(j)-k6*Xem3(j)-k4*Xem2(j)*Xem3(j))*Delta... + c5*sqrt(k5*Xem1(j))*Winc5-c6*sqrt(k6*Xem3(j))*Winc6- c4*sqrt(k4*Xem2(j)*Xem3(j))*Winc4);
        Xem4(j+1) = (Xem4(j) + k4*Xem2(j)*Xem3(j)*Delta + c4*sqrt(k4*Xem2(j)*Xem3(j))*Winc4);
        if (Xem1(j+1) < 0)
            Xem1(j+1)=0;
        end
        if (Xem2(j+1) < 0)
            Xem2(j+1)=0;
        end
        if (Xem3(j+1) < 0)
            Xem3(j+1)=0;
        end
        if (Xem4(j+1) < 0)
            Xem4(j+1)=0;
        end
    end
end
if mod(j,mT)==0
    yout(:,jj)=[Xem1(j+1) Xem2(j+1) Xem3(j+1) Xem4(j+1)];
    jj=jj+1;
end
end
outR(jjj,:)=[R_int yout(1,:)];
outG(jjj,:)=[G_int yout(2,:)];
outP(jjj,:)=[P_int yout(3,:)];
outV(jjj,:)=[V_int yout(4,:)];
jj=1; % reset jj counts
end
savefile = strcat('d:\cle\fR',num2str(c1),num2str(c2),num2str(c3),num2str(c4),num2str(c5),num2str(c6),'.mat');
save(savefile,'outR')
savefile = strcat('d:\cle\fG',num2str(c1),num2str(c2),num2str(c3),num2str(c4),num2str(c5),num2str(c6),'.mat');
save(savefile,'outG')
savefile = strcat('d:\cle\fP',num2str(c1),num2str(c2),num2str(c3),num2str(c4),num2str(c5),num2str(c6),'.mat');
save(savefile,'outP')
savefile = strcat('d:\cle\fV',num2str(c1),num2str(c2),num2str(c3),num2str(c4),num2str(c5),num2str(c6),'.mat');
save(savefile,'outV')

disp(['it takes ' num2str(toc) 's.']);

E.5 The viral infection model by SDEs (extrinsic noise)

virus_sde().m

% effects of extrinsic noise in viral infection model
% modelled by SDEs
% simulated by stochastic Euler method
% SDE is
% \[ dX(1) = (k1*X(2)-k2*X(1))*dt+c1*X(2)*dW(1)-c2*X(1)*dW(2), \]
% \[ X(1)_0 = R_int \]
% \[ dX(2) = (k3*X(1)-k1*X(2)-k4*X(2)*X(3))*dt-c3**X(1)*dW(3) \]
% \[ -c1*X(2)*dW(1)-c4*X(2)*X(3)*dW(4), \]
% \[ X(2)_0 = G_int \]
% \[ dX(3) = (k5*X(1)-k6*X(3)-k4*X(2)*X(3))*dt+c5**X(1)*dW(5) \]
% \[ -c6*X(3)*dW(6)-c4*X(2)*X(3)*dW(4), \]
% \[ X(3)_0 = P_int \]
% \[ dx(4)=k4*X(2)*X(3)+c4*X(2)*X(3)*dW(4); \]
% X(1) is mRNA, X(2) is genome, X(3) is protein and X(4) is virus
% USAGE
% []=extrinsic(flag,par1,par2,par3,par4,par5,par6)
% flag is used to name output files
% par1,...,par6 define extrinsic noise coefficients c1,...c6
function []=virus_sde(flag,par1,par2,par3,par4,par5,par6)
clf
tic;

rand('state',sum(clock));

% running parameters
T = 200;        % running time
N = 2000000;     % time points to calculate
Delta = T/N;    % delta t
DP=200;         % data points to save
m=1000;         % replications of realisations

% rate constants
k1=0.025;
k2=0.25;
k3=1;
k4=7.5e-6;
k5=1000;
k6=1.99;

% extrinsic noise coefficients
c1=par1;
c2=par2;
c3=par3;
c4=par4;
c5=par5;
c6=par6;

% initial number of species
R_int=5;
G_int=1;
P_int=0;
V_int=0;

% temp data
Xem1=zeros(1,N);
Xem2=zeros(1,N);
Xem3=zeros(1,N);
Xem4=zeros(1,N);

Xem1(1) = R_int;
Xem2(1) = G_int;
Xem3(1) = P_int;
Xem4(1) = V_int;

% initial index
jj=1;
mT=N/DP; % save every mT point
outR=zeros(m,DP+1); % outR is the final output of R
outG=zeros(m,DP+1);
outP=zeros(m,DP+1);
outV=zeros(m,DP+1);

for jjj=1:m
    jjj
    yout=zeros(4,DP);

    for j = 1:N
        Winc1 = sqrt(Delta)*randn;
        Winc2 = sqrt(Delta)*randn;
        Winc3 = sqrt(Delta)*randn;
        Winc4 = sqrt(Delta)*randn;
Winc5 = sqrt(Delta)*randn;
Winc6 = sqrt(Delta)*randn;

Xem1(j+1) = (((Xem1(j) + (k1*Xem2(j)-k2*Xem1(j)))*Delta + ... 
c1*Xem2(j)*Winc1-c2*Xem1(j)*Winc2));
Xem2(j+1) = ((Xem2(j) + (k3*Xem1(j)-k1*Xem2(j)-
k4*Xem2(j)*Xem3(j)))*Delta... 
- c3*Xem1(j)*Winc3-c1*Xem2(j)*Winc1-c4*Xem2(j)*Xem3(j)*Winc4);
Xem3(j+1) = (Xem3(j) + (k5*Xem1(j)-k6*Xem3(j)-
k4*Xem2(j)*Xem3(j)))*Delta... 
+ c5*Xem1(j)*Winc5-c6*Xem3(j)*Winc6-c4*Xem2(j)*Xem3(j)*Winc4);
Xem4(j+1) = (Xem4(j) + k4*Xem2(j)*Xem3(j)*Delta + 
c4*Xem2(j)*Xem3(j)*Winc4);

if (Xem1(j+1) < 0)
Xem1(j+1)=0;
end

if (Xem2(j+1) < 0)
Xem2(j+1)=0;
end

if (Xem3(j+1) < 0)
Xem3(j+1)=0;
end

if (Xem4(j+1) < 0)
Xem4(j+1)=0;
end

if mod(j,mT)==0
yout(:,jj)=[Xem1(j+1) Xem2(j+1) Xem3(j+1) Xem4(j+1)];
jj=jj+1;
end

outR(jjj,:)=R_int yout(1,:);
outG(jjj,:)=G_int yout(2,:);
outP(jjj,:)=P_int yout(3,:);
outV(jjj,:)=V_int yout(4,:);
jj=1; % reset jj counts
end

% save output files
savefile = strcat('D:\SDE\fRex','num2str(flag),'.mat');
save(savefile,'outR')
savefile = strcat('D:\SDE\fGex','num2str(flag),'.mat');
save(savefile,'outG')
savefile = strcat('D:\SDE\fPex','num2str(flag),'.mat');
save(savefile,'outP')
savefile = strcat('D:\SDE\fVex','num2str(flag),'.mat');
save(savefile,'outV')

disp(['it takes ' num2str(toc) 's.']);
clear;