Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- you will use the copy only for the purposes of research or private study
- you will recognise the author's right to be identified as the author of the thesis and due acknowledgement will be made to the author where appropriate
- you will obtain the author's permission before publishing any material from the thesis.
Remodelling circadian rhythm in *Drosophila melanogaster*:

To investigate the role of a new clock component

Clockwork Orange (CWO)

A thesis
submitted in partial fulfilment
of the requirements for the Degree of
Doctor of Philosophy
In Computational Systems Modelling

at
Lincoln University
by
Jeevabharathi Ranganathan

Lincoln University
2015
Abstract of a thesis submitted in partial fulfilment of the requirements for the Degree of Doctor of Philosophy.

Remodelling circadian rhythm in *Drosophila melanogaster*:

To investigate the role of a new clock component

Clockwork Orange (CWO)

by

Jeevabharathi Ranganathan

The ability of almost all organisms to change their behaviour on a daily basis is one of the remarkable features of life on earth. This phenomenon which is called circadian rhythm is observed in diverse organisms such as algae, fruit flies and humans and is a response arising due to the rotation of the earth around the axis resulting in an internal time-keeping system. Changes in myriad of biochemical and physiological processes take place in order for an organism to adapt for changes in physical environment. The period of this process is close to 24 hours in duration, hence the name “circadian rhythms”, from Latin *circa diem* meaning about a day. In the fruit fly *Drosophila melanogaster*, due to the increase in knowledge of genetics and molecular biology the molecular components such as genes and proteins involved in circadian rhythm and their roles are well understood. Due to the oscillatory properties of clock components they are an ideal candidate for mathematical models and many such models have been developed in the past.

In this study, three new *Drosophila* circadian rhythm models were developed, each with three transcriptional regulatory feedback loops. Among which, two feedback loops (VRI/PDP1 and PER/TIM) are well known and have been included in earlier models. The main focus of this study is the newly discovered third feedback loops (CWO). The differences between the three models are defined by our conceptualization of three probable actions by which the newly discovered clock component CWO (Clockwork Orange) performs its dual role both as an activator and repressor of *per, tim, vri, pdp1* genes, and *cwo* genes. We included existing *in vitro* understanding of molecular components and extended it to include probable molecular roles of the newly discovered clock component CWO. We based our hypothesis on discovered *in vivo* dynamics and by analysing the CWO protein sequence using basic bioinformatics servers. Detailed modelling in the form of probability based transcription factor binding and unbinding processes are used. All three models are
expressed by a set of probability based mass action governed ordinary differential equations and the parameters were estimated using modelling tool COPASI. Due to the randomness and variation of different data sets generated for CWO activity by biologists, we made a choice to differ from a traditional approach in modelling, by not over-relaying on data generated from *in vitro* analysis. The reliance on wet-lab data was scaled down and we include them only to choose manageable mathematical inputs and validate a solved model. This approach gave us a relative degree of space to be innovative and permitted us to test different hypothesis at conceptual level in three models. We proceeded to solve the models and validate the oscillations by testing with mutations. Outputs of our simulations will help broaden the research arguments in the field of cricadian biology. In particular our models hypothetically answers the molecular role of CWO protein.

**Keywords:** COPASI, circadian clock, systems biology, *Drosophila melanogaster*, genetic regulatory networks, circadian rhythms, mathematical molecular modelling, protein, mRNA, kinetic modelling, mass action kinetics, deterministic modelling, ordinary differential equations, clockwork orange, oscillations.
Acknowledgements

Foremost of all, I would like to thank my parents for their deep faith, love, encouragement and moral support. I am deeply indebted to my supervisor Prof. Don Kulasiri for being my mentor, friend. His excellent teaching through his research and lab meetings which has guided me through all my PhD years. Moreover, Prof. Don taught me patience and integrity, which will stay with me for the rest of my research career. I also thank Dr. Rainer Hofmann, Prof. Sandhya Samarasinghe and Prof. Jim Morton for their inputs in my work and invaluable encouragements.

I would like to thank the whole C-fACS (Centre for Advanced Computational Solutions) team members, all of them contributed to my research life in many pleasant ways. In particular, I would like to thank Dr. Zhongkun Zhou for his friendship and interesting discussions from start of my PhD to end. Yao He for all his help in lab as well as during my travelling in China.

Special thanks to all my friends, family, and well-wishers in Christchurch city, who during painful times made me have the ability to laugh at it. In particular I would like to thank Dr. Lingam Balakrishnan and his family for their special friendship. I would like to thank my sister Nandhini Ranganathan for her encouragement, scientific temper and being a patient listener.

I am indebted to my father Ranganathan Muthunaidu, and mother Kanchana Ranganathan for their constant love, dedication and encouragement throughout my study and life. And lastly a very special thanks to the love of my life Jocelyn for her never ending love, understanding and inspiration that kept me going all these years and tough times. Finally, I sincerely thank my loving parents again and always, to whom I dedicate this thesis to.
Table of Contents

Abstract ............................................................................................................................................. 1
Acknowledgements ......................................................................................................................... iii
Table of Contents ............................................................................................................................ iv
List of Tables ....................................................................................................................................... vii
List of Figures ....................................................................................................................................... viii
Abbreviations ..................................................................................................................................... x

Chapter 1. Introduction ....................................................................................................................... 1
1.1 Circadian rhythm ....................................................................................................................... 1
1.2 Research motivation ............................................................................................................... 3
1.3 Objectives of the research ..................................................................................................... 4
1.4 Overview of chapters ............................................................................................................. 5

Chapter 2. Background ...................................................................................................................... 6
2.1 Central dogma of molecular biology ....................................................................................... 6
2.2 Gene expression ..................................................................................................................... 8
2.3 Regulation of gene expression ............................................................................................... 8
  2.3.1 Regulation of transcription ............................................................................................. 9
  2.3.2 Cis-regulatory regions ................................................................................................... 9
  2.3.3 Enhancers .................................................................................................................... 9
2.4 Transcriptional factors ......................................................................................................... 11
  2.4.1 Transcriptional activators and repressors ................................................................. 11
  2.4.2 Cooperativity .............................................................................................................. 12
  2.4.3 Hill equation .............................................................................................................. 14
2.5 Translation ........................................................................................................................... 15
  2.5.1 Three step process ...................................................................................................... 16
  2.5.2 Achieving critical mass .............................................................................................. 16

Chapter 3. Concepts in biosystems modelling .............................................................................. 17
3.1 Concept of expression rates ................................................................................................. 17
3.2 Enzyme-Substrate interactions ......................................................................................... 18
  3.2.1 Lock-key model ......................................................................................................... 18
  3.2.2 Induced-fit model ...................................................................................................... 18
  3.2.3 Multiple substrate enzyme reactions ....................................................................... 20
3.3 Kinetics of expression rates ............................................................................................... 22
  3.3.1 Michaelis-Menten kinetics ...................................................................................... 23
  3.3.2 Mass action kinetics ................................................................................................. 29
3.4 Genetic regulatory networks ............................................................................................... 31
  3.4.1 Modelling genetic regulatory networks .................................................................. 32
  3.4.2 Noise in GRNs .......................................................................................................... 33
  3.4.3 Multi-scale issues in modelling biochemical systems ........................................... 35
  3.4.4 Randomness of biochemical reactions .................................................................. 36

Chapter 4. Circadian clocks ............................................................................................................ 38
4.1 Importance of clock ............................................................................................................. 38
Chapter 5. Review of circadian models in *Drosophila melanogaster* ........................................... 65
5.1 First clock component PER ................................................................. 65
5.2 Circadian models .................................................................................... 65
5.2.1 Goldbeter Model ............................................................................. 65
5.3 Discovery of PER/TIM feedback loop.................................................... 66
5.3.1 Leuloup and Goldbeter model ......................................................... 66
5.3.2 Tyson model .................................................................................. 68
5.4 Discovery of CLK/CYC .......................................................................... 68
5.4.1 Smolen model ................................................................................ 70
5.5 Discovery of VRI/PDP1 feedback loop .................................................. 70
5.6 Later models with two feedback loops .................................................. 70
5.6.1 Ruoff model .................................................................................. 71
5.6.2 Xie and Kulasiri model ................................................................. 74
5.6.3 Highlights of Xie and Kulasiri model ............................................... 74

Chapter 6. Developing three new *Drosophila* circadian rhythm models using CWO............... 77
6.1 Developing conceptual models ............................................................... 77
6.1.1 Conceptualising model A ............................................................... 77
6.1.2 Conceptualising model B ............................................................... 80
6.1.3 Conceptualising model C ............................................................... 83
6.2 Modelling transcriptional regulation..................................................... 88
6.2.1 The requirement for detailed modelling with explicit reactions ........ 88
6.3 Modelling transcriptional activation and repression .............................. 90
6.3.1 Kinetic equations explained .......................................................... 94
6.3.2 Variable names and parameters explained ..................................... 96
6.3.3 Binding probabilities explained ..................................................... 103
6.3.4 Models described by system of ODEs ............................................. 104
6.4 Implementing the models with parameter estimation ......................... 107
6.4.1 COPASI (Complex pathway simulator) .......................................... 108
6.4.2 Parameter estimation ................................................................. 109

Chapter 7. Simulation results and testing ............................................................................. 118
7.1 Simulation results of models A,B and C ............................................. 118
7.1.1 Oscillations of clock components in DD (constant darkness) .... 118
7.1.2 Robustness of the models to parameter variations ....................... 122
7.2 Light entrainment of the circadian clock ............................................ 127
7.2.1 Effect of light .............................................................................. 127
7.3 Testing mutations .............................................................................. 133
7.3.1 Testing null mutants ................................................................. 133
## List of Tables

| Table 1-1 | Molecular components of *Drosophila* circadian clock | 45 |
| Table 1-2 | COTH output data for PER homodimer | 54 |
| Table 1-3 | GPS2.0 predicted phosphorylation sites in CWO protein | 85 |
| Table 1-4 | Biological description of the kinetic parameters in model A | 97 |
| Table 1-5 | Biological description of the kinetic parameters in model B | 99 |
| Table 1-6 | Biological description of the kinetic parameters in model C | 101 |
| Table 1-7 | Estimated kinetic parameter values for model A | 114 |
| Table 1-8 | Estimated kinetic parameter values for model B | 115 |
| Table 1-9 | Estimated kinetic parameter values for model C | 116 |
| Table 2-1 | Initial conditions used for all three models | 117 |
| Table 2-2 | Oscillation peaks of clock components in DD (Constant darkness) | 123 |
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>A signalling pathway of human circadian rhythm entrained by light.</td>
</tr>
<tr>
<td>1-2</td>
<td>A flowchart depicting the central dogma of molecular biology.</td>
</tr>
<tr>
<td>1-3</td>
<td>Regulation of gene expression: (a) A model of a Eukaryotic promoter, and (b) Transcription initiation complex on a promoter showing cooperativity between various TFs.</td>
</tr>
<tr>
<td>1-4</td>
<td>A plot comparing positive cooperativity(red curve) to a hyperbolic response (blue curve).</td>
</tr>
<tr>
<td>2-1</td>
<td>Enzyme-substrate interactions: (a) Lock-key model and (b) Induced-fit model.</td>
</tr>
<tr>
<td>2-2</td>
<td>In the above figures, plot A shows a Lineweaver–Burk plot of multiple substrate enzyme reactions in sequential mechanist and plot B shows a Lineweaver–Burk plot of multiple enzyme substrate reactions in a ping-pong mechanism.</td>
</tr>
<tr>
<td>2-3</td>
<td>Comparing Kinetics: (a) Michaelis-Menten Kinetics and (b) Hill Kinetics.</td>
</tr>
<tr>
<td>2-4</td>
<td>A Plot of substrate concentration versus reaction rate.</td>
</tr>
<tr>
<td>3-1</td>
<td>Schematic diagram of Drosophila cerebral cells showing seven main groups of neurons.</td>
</tr>
<tr>
<td>3-2</td>
<td>Relative 24 h concentrations of circadian clock components in Drosophila melanogaster: (a) mRNA levels and (b) Protein levels.</td>
</tr>
<tr>
<td>3-3</td>
<td>Network diagram for phosphorylation interaction of PER.</td>
</tr>
<tr>
<td>3-4</td>
<td>COTH output result showing predicted secondary structure of PER PAS domains, and one of the output image is used.</td>
</tr>
<tr>
<td>3-5</td>
<td>Screening CRY protein sequence: A Part of the NetPhosK output image showing phosphorylation sites towards the N terminus of CRY protein, here C-indicates PKC, K(green)-indicates CKII, K(pink)-indicates CKI &amp; X-indicates GSK3.</td>
</tr>
<tr>
<td>3-6</td>
<td>Schematic diagram of circadian oscillation of clock components in Drosophila melanogaster and red blocked lines denote the CWO ambiguity.</td>
</tr>
<tr>
<td>4-1</td>
<td>Schematic diagram of Tyson’s model (1999) showing PER/TIM and CLK/CYC feedback loop.</td>
</tr>
<tr>
<td>4-2</td>
<td>Schematic diagram of Smolen’s model (2004) showing PER, CLK, and VRI/PDP1 feedback loops.</td>
</tr>
<tr>
<td>4-3</td>
<td>Schematic diagram of Ruoff model (2005) showing PER/TIM, VRI/PDP1, and CLK feedback loops.</td>
</tr>
<tr>
<td>4-4</td>
<td>Schematic diagram of Xie and Kulasiri model (2007) showing PER/TIM, VRI/PDP1, and CLK/CYC feedback loops.</td>
</tr>
<tr>
<td>5-1</td>
<td>Conceptual model A: in the above network diagram of conceptual model A, blocked red lines show the repressor activity and green arrows show the activator component.</td>
</tr>
<tr>
<td>5-2</td>
<td>Conceptual model B: in the above network diagram of conceptual model B, blocked red lines show the repressor activity and green arrows show the activator component.</td>
</tr>
<tr>
<td>5-3</td>
<td>A part of the NetPhosK output image showing phosphorylation sites between amino acid number 301 -480 of CWO protein, here C-indicates PKC, K(green)-indicates CKII, K(pink)-indicates CKI &amp; X-indicates GSK3.</td>
</tr>
<tr>
<td>5-4</td>
<td>A 3D chart of largest period difference effected by 6 parameters are shown: (a) In model A and B; (b) In model C.</td>
</tr>
<tr>
<td>5-5</td>
<td>Different steps of parameter estimation used in our models.</td>
</tr>
<tr>
<td>6-1</td>
<td>Robustness to parameter variations: (a) In model A, (b) In model B, and (c) In model C.</td>
</tr>
<tr>
<td>6-2</td>
<td>A 3D chart of largest period difference effected by 6 parameters are shown: (a) In model A and B; (b) In model C.</td>
</tr>
<tr>
<td>6-3</td>
<td>LD oscillation: (a) In model A, (b) In model B, and (c) In model C.</td>
</tr>
<tr>
<td>6-4</td>
<td>LD oscillation showing phase shift: (a) In model A, (b) In model B, and (c) In model C.</td>
</tr>
</tbody>
</table>
Figure 7-7: Damped protein oscillation in high k light: (a) In model A, (b) In model B, and (c) In model C.

Figure 7-8: Damped mRNA oscillation in per\(^{01}\) tim\(^{01}\) mutants: (a) In model A, (b) In model B, and (c) In model C.

Figure 7-9: Damped mRNA oscillation in clk\(^{4}\) mutants: (a) In model A, (b) In model B, and (c) In model C.

Figure 7-10: Period changes in PER protein oscillation of per\(^{L}\) and per\(^{S}\) mutants: (a) In model A, (b) In model B, and (c) In model C.

Figure 7-11: Robust oscillation of proteins with CWO loop removed: (a) In model A, (b) In model B, and (c) In model C.

Figure 7-12: A 3D histogram of Local sensitivity results with CWO loop removed and normal: (a) In model A, (b) In model B, and (c) In model C.

Figure 7-13: Comparing Local sensitivity ratios with CWO loop removed.

Figure 7-14: Time evolution (72 h) of mRNAs cwo\(^{B9}\) mutants: (a) In model A, and (b) in model B.

Figure 7-15: Time evolution (10 d) of mRNAs cwo\(^{B9}\) mutants: (a) In model A and (b) in model B.

Figure 7-16: Time evolution of mRNAs cwo\(^{B9}\) mutants in model C: (a) 72 h and (b) 10 d.

Figure 7-17: Time evolution of CLK protein concentration in wildtype and cwo Double knockout mutant.

Figure 7-18: Oscillation (10 d) of mRNAs in cwo\(^{CWPT}\) mutant in model C.

Figure 7-19: Time evolution (48 h) of mRNAs cwo\(^{CWPT}\) mutants in model C: (a) cwo mRNA and (b) per mRNA.

Figure 7-20: Time evolution transcripts in cwo\(^{CWPT}\) mutants in model C: (a) PER protein oscillation and (b) Protein oscillations for up to 10 d.

Figure 8-1: Time evolution of circadian models in Drosophila melanogaster.
Abbreviations

bHLH - Basic helix-loop-helix
CK2 - Casein Kinase 2 protein
CLK - CLOCK protein
clk - clock gene
cwo - Clockwork orange gene
CWO - Clockwork orange protein
LL - Constant light
COPASI - Complex pathway simulator
CYC - CYCLE protein
eye - cycle gene
CWD - CWO/HP complex
CWPT - CWO/PER/TIM complex
DNA - Deoxyribonucleic acid
DBT - DOUBLETIME
XML - Extensible Markup Language
GRN - Genetic regulatory network
HP - Hypothetical protein
LM - Levenberg-Marquardt
in silico - In the computer
in vitro - In the cell
in vivo - In a living organism
LD - Light-Dark
mRNA - Messenger RNA

LL - Constant light
ODE - Ordinary differential equation
PDP1 - PAR Domain Protein 1
pdp1 - PAR Domain Protein 1 gene
PAS - PER ARNT SIM
PER - PERIOD protein
per - period gene
PRC - Phase response curve
PT - PER and TIM protein complex
RNA - Ribonucleic acid
RNAP - RNA polymerase
SGG - Shaggy protein
SLMB - Slimb protein
SBML - Systems Biology Markup Language
TIM - TIMELESS protein
tim - timeless gene
TF - Transcription factor
VRI - VIRLLE protein
vri - virle gene
WSS - Weighted sum of squares
WT - Wild-type
DD - Constant Darkness
Chapter 1: Introduction

1.1 Circadian rhythm

Circadian rhythms are a response of living organisms to the sinusoidal variations in day-length, temperature and associated physiological changes over yearly periods caused by the inclined axis of the planet Earth with respect to the plane of the latter’s orbit around the sun. They can be defined biologically as entrained oscillations of gene, mRNA, and protein activity patterns within a period of approximately twenty-four hours that results in observable physiological changes in the organism. Such oscillations can be affected by changes in environmental cues (Figure. 1), most significantly those of light and temperature, hence most biological organisms from some prokaryotes such as metazoans to plants and eukaryotes including mammals exhibit circadian rhythm. Disruption of which causes physiological effects, for instance for humans travelling rapidly around the Earth’s axis in passenger airliners resulting in the people experiencing disorientation in their perception of the time of the day in what is colloquially called “jet-lag.” Circadian rhythms are the underlying mechanism that triggers aspects such as ovulation in animals, and flowering times in plants thus enabling the organism to function efficiently at reproductive maturity, and allowing the continued viability of the species in nature against the losses due to attrition.

Where environmental conditions are constant for extended periods of time circadian rhythm demonstrates recognisable characteristics as follows; Tts rhythmicity is entrained by temperature yet the period output is in the same phase over a variety of temperatures (Helfrich Forster et al. 2005). Circadian rhythm is also synchronized with day and night cycles with, both temperature and light as the major input signals, which are known as zeitgebers.

A good way to observe the circadian rhythm at physiological, molecular and behaviour scales is in insect models, notably the fruit fly Drosophila melanogaster that has been used in laboratory genetic studies for the last hundred years on account of the ease and economy with which laboratory strains may be maintained. The
Circadian Rhythm

Figure 1-1 A signalling pathway of human Circadian rhythm entrained by light
circadian clock responsible for regulating overt behavioural rhythms has been identified as localised in the brain of insects e.g., the optic lobe in crickets, cockroaches, and beetles is the clock locus (Page et al. 1982; Tomioka et al. 1992; Fleissner et al. 1982), whereas for moths, flies, and mosquitoes the central brain is postulated as the locus of the clock (Truman et al. 1974; Kasai et al. 1987).

As almost all species exhibit endogenous daily rhythms and a high level of scientific interest has been generated aiming to decipher molecular basics behind the clock that generates circadian rhythms, and their physiological implications. The importance of circadian rhythmicity in maintaining health and well-being in humans is attracting scientific and lay interest in recent years. Abnormal circadian behaviour induced by lack of sleep in many typical contemporary human lifestyles is linked to changes in hormone signals and metabolic pathways that in turn become one of the factors leading to severe cardiovascular diseases.

The thresholds of sensitivity to circadian rhythm vary according to the ecological niche exploited by the organism. Marine life in the extreme ocean depths, as well as cave dwelling aquatic animals experience constant temperature and constant darkness, and hence having no requirement to monitor seasonal changes, have devolved away from circadian rhythm sensitivity as all biochemical reactions have a metabolic cost (Kim et al. 1997).

1.2 Research motivation

This thesis covers the most important molecular components involved in circadian rhythm of *Drosophila melanogaster*, their positive and negative feedback loops, and in particular the recently discovered feedback loop involving the “clockwork orange” (CWO) protein.

Although there is ample scientific literature available that describes circadian rhythm, and its components, in diverse organisms as *Arabidopsis*, *Cyanobacteria*, *Insecta*, *Neurospora*, *Insecta*, and *Mammalia*; the reason *Drosophila melanogaster* was
chosen for this particular study is that we have a previous model of *Drosophila* circadian rhythm, and its linked feedback loops done by our group. This study thus expands on the existing model by incorporating an additional feedback loop of CWO, and hence this newly developed and extended model satisfies the various criteria set in the wet-lab data results. CWO, a typical helix loop orange-domain protein is a newly discovered clock component in *Drosophila*, its mammalian homologues being known as DEC1 and DEC2. The structure of the CWO peptide and its functional domains increased our curiosity to check different roles of CWO in regulating the clock. It resulted in this study being developed into three models with varying CWO molecular roles. Biological reasons for such an approach will be discussed in detail in following chapters. Before which, I would like to state the objectives of this current study.

### 1.3 Objectives of the research

Although our initial approach in this study is to develop a new model with an additional feedback loop, the versatility of clock components, structural and functional motifs of CWO gave us a variety of research questions to follow. In addition the choice of using detailed modelling gave us the flexibility to test various hypotheses with limited parameter changes. To satisfy these requirements overall research objectives were formulated as such.

1) To collect and review the current biological data available on *Drosophila* circadian rhythm.
2) Expanding the existing model by incorporating an additional feedback loop.
3) Develop three conceptual models each with an additional feedback loop to include different CWO activity.
4) Interpreting conceptual models mathematically.
5) To solve the models by estimating parameters.
6) Use the simulation results to answer CWO ambiguity.
7) Interpreting the *in silico* results with various *in vitro* and *in vivo* results.
8) Test the sensitivities of local parameters to interpret them biologically.
It is necessary to mention that, we accomplished all the major objectives of this study with necessary results. With that note, I would like to start with introducing basic eukaryotic cellular components.

1.4 Overview of chapters

The current chapter starts with introducing circadian rhythms, motivation and objectives for the current research. In chapter two, basic molecular biology concepts are touched upon including cooperativity. In chapter three mathematical basics of modelling biochemical systems are discussed. In chapter four, a detailed review of circadian models is given. In chapter five important clock components and feedback loops of circadian rhythm in the Drosophila model organism are presented in great biological detail. In chapter six, three conceptual models are developed from the molecular basis, which include the new clock component CWO, then the conceptual models are converted into mathematical models, the mathematical models are simulated in COPASI by estimating the parameters. In the seventh chapter various tests and mutations are carried out and the results are compared to appropriate wet-lab data. And finally in the eighth chapter the whole research is discussed in detail and appropriate future directions are given.
Chapter 2: Background

2.1 Central dogma of molecular biology

It is well established scientifically that the basic molecular components of cells are composed of deoxyribonucleic acid (DNA), ribonucleic acid (RNA) and proteins (Crick. 1970). DNA encodes the genetic information in the form of subsets of codons that recognise specific aminoacids at three receptor sites in transfer RNA (tRNAs). Such subsets are usually three bases long and the code is called the triplet code. On messenger RNA (mRNA) aminoacids are assembled through the ribosome complex to form proteins or enzymes. Hence, DNA is a macromolecule needed for the viability of all forms of life. The functional units of DNA are called genes, and it is understood in practice that the mRNA transcript is a gene derivative (Basham et al.1995).

DNA is composed of nucleotides made up nitrogen bases (an inorganic molecule with a nitrogen atom), adenine [A] guanine [G], cytosine [C], and thymine [T], but in RNA uracil [U] replaces thymine. Series of nucleotides are linked to the DNA macromolecule through a bridge of phosphate and sugar complexes. Genes are considered to be conservative in that they are copied during mitosis with an extremely small error rate (ohno. 1972). Non-coding sections of DNA were called “junk DNA” formerly, but are now referred to as “non-coding DNA”. They were until recently considered to be merely packing material between genes, but the current understanding is that at least some of the junk DNA is functional to some degree (Biémont et al. 2006).

The mRNA is read by the ribosome complex (Figure 2-1) in the five-prime (5’) to three-prime direction (3’) in both prokaryotes and eukaryotes always error free (Iwasaki et al. 1968). However, the human mitochondrial genome of about 16,000 bases codes for tRNAs in both 5’ - 3’ and 3’ - 5’ directions without duplication in either directions (Kocher et al. 1989).
Figure 2-1 A flowchart depicting the central dogma of molecular biology.
2.2 Gene expression

The process whereby various generic and sequence specific enzymes assist in converting the double-stranded DNA template into a functional protein is called gene expression. The nucleotide sequence from double stranded DNA is copied to produce a single stranded messenger RNA (mRNA), this step is called transcription (Studier et al. 1986). RNA polymerase II (RNAP II) is a major component in this step. The enzyme RNAP II reads the nucleotides from the coding part of DNA (gene) (Parker et al. 1984). A new single stranded mRNA molecule is created which has bases complementary to the DNA molecule.

This mRNA is used in the synthesis of a protein molecule in a process called translation. In translation a protein sequence is assembled based on the mRNA “triplet code” (Hayes. 1998). The mRNA sequence is “read” in sets of three nucleotides, which are called “codons” and each codon is related to a specific amino acid. A protein sequence with a chain of amino acids assembled by the ribosome, starts from the start codon AUG and the chain grows until it encounters the stop codons, either UAA or UAG (Skuzeski et al. 1991). The resulting protein chain is highly unstable and usually not active. Post-translational modification such as phosphorylation and de-phosphorylation are needed to form a functional stable protein molecule, there by completing the process of gene expression.

2.3 Regulation of gene expression

Gene expression can be regulated at the level of transcription and post-translational modifications. Transcription can be regulated through the number of mRNA copies synthesised from DNA. Transcriptional regulation happens at a region called “promoter” and it is explained in detail at the next section. The amount of functional protein level is determined by the presence of other enzymes which are involved in post-translational modifications.
2.3.1 Regulation of transcription

In eukaryotes such as *Drosophila melanogaster*, the regulation happens at the stage of transcription initiation. During the start of the transcription, specific molecular components bind to the DNA sequence upstream of the coding region, and are called as transcription factors (TFs). These regulatory proteins also interact with RNAP II (Zhang et al. 2001), resulting in up-regulation or down-regulation of transcription initiation. In most eukaryotes, transcriptional regulation involves a combinational regulation of different TFs. This interaction between different TFs and RNAP II leads to increase or decrease in mRNA copy numbers. In addition, certain changes in DNA properties do occur, where some cytosine bases are converted to methyl cytosine (5mC) during the chromosome formation, i.e. methylation (Ng H et al. 1999). In summary, regulation of transcription is an elaborate phenomenon which plays a major part in regulating the gene expression in various cell types.

2.3.2 Cis-regulatory regions

As mentioned in the previous section, transcriptional control of eukaryotic genes is exhibited by the combinational influence of various TFs binding to certain regulatory sequences. These regions are 6 to 8 base pairs (bp) in length and are called Cis-regulatory regions or motifs, e.g. -CCAAT (CAT box) and GGGCGG (GC boxes). These cis-acting regulatory sequences are generally located upstream of the TATA box (Figure 2-2) within a range of 100 bp (Wray et al. 2007). Various TFs can bind to the same Cis-regulatory region in a specific gene, thereby exhibiting a phenomenon called cooperativity which is discussed in detail in section 2.4.2.

2.3.3 Enhancers

Sometimes certain regulatory sequences can be located more than 10kbp upstream of the promoter region, or in some instances on a different chromosome altogether. These sequences are called enhancers (Su et al. 1990), are always Cis-acting (Figure 2-2), and like Cis-regulating regions TFs can bind to these enhancers and interact with RNAP II to influence transcription initiation. That interaction over such a substantial distance (10kbp) in DNA terms is possible because of the helical and the consequent
Figure 2-2 Regulation of gene expression: (a) A model of a Eukaryotic promoter, and (b) Transcription initiation complex on a promoter showing cooperativity between various TFs.
super-helical looping of DNA. Further, enhancer sequences can occur within introns (non-coding region of a gene) of the gene and also upstream or downstream of the whole gene itself. In eukaryotes such as *Drosophila melanogaster*, gene expression can be controlled by multiple enhancer sequences and a variety of TFs bound to these enhancers (Sen et al. 1986).

### 2.4 Transcription factors

The function of a transcription factor (TF) is to initiate transcription by identifying and binding to specific regions in a gene promoter such as enhancers and *Cis*-regulatory regions. Furthermore, TFs recruit other regulatory proteins and undergo a conformational change in order to accomplish DNA binding (Hunter et al. 1992). Research has shown that TFs are capable of binding tightly to DNA, both *in vitro* and *in vivo* (Pahl et al. 1999). The reason for TFs to recruit other regulatory proteins is to attract RNAP II and form a transcription initiation complex (Figure 2-2). Remarkably, the interaction of various TFs through their functional domains determines the nature of the whole transcription initiation complex (Rao et al. 1997). Such interactions determine whether the complex regulates transcription in a positive or negative fashion. Universally, a TF will have several structural domains in order to interact with accompanying TFs as well to successfully bind to regulatory regions of a gene. Collectively, all the TFs in an initiation complex which interact and cooperate to enable the transcription process of a same gene, are called co-activators.

#### 2.4.1 Transcriptional activators and repressors

Transcriptional activators are the group of TFs which start the transcription process. A typical activator protein has two distinct “functional domains”, which are necessary for their activator role. One domain is needed to recognise and bind to specific regulatory regions of the gene, and the other domain is used to interact with accompanying TFs (Sakura et al. 1989). These functional domains have a distinct structure and are evolutionarily conserved in various organisms and cell types. Based on their structure and their target sequences, transcriptional activators are classified in various TF “families”. For example, there are three known transcriptional activators
(CLK, CYC and PDP1) involved in circadian rhythm of *Drosophila melanogaster* (Rutila. 1998). Among them, CYC and CLK TFs belong to the basic helix-loop-helix protein family and bind to a consensus E-box of 6 bp sequence (CACGTG). In contrast, PDP1, which binds to 8 bp VP-box (TTATGTAA), belongs to the basic leucine-zipper protein family. Some TFs play a role in inhibiting transcription, and are called as transcriptional repressors. Similar to activators, repressor proteins have two functional domains and have the same role from a molecular standpoint (Hardin. 2005). Repressors stop the transcription by two ways. First, by binding and occupying the regulatory regions thus, denying any open sequences for activator proteins to bind. Next, by interacting with activator proteins directly by phosphorylation or dephosphorylation resulting in degradation of activator TFs. E.g. in *Drosophila melanogaster* CWO is a basic helix-loop-helix repressor and VRI is a basic leucine-zipper repressor. A detailed review of all TFs involved in *Drosophila* circadian rhythm is provided in various sections of chapters 4 and 5.

### 2.4.2 Cooperativity

Enzymes which are known to exhibit varied binding sites are reported in various literature as having the phenomenon of cooperativity, meaning functional co-operation. Proteins are made up of a sequence of amino acids and subunits. Many of the subunits are known to be identical. Such proteins are termed as oligomeric. E.g. a protein known as phosphofructokinase has four subunits which are identical in *E. coli*. Every subunit in phosphofructokinase has multiple binding sites for other interacting compounds such as ATP and ADP (Walker. 1982). If two or more binding sites in a subunit are close to each other, then one binding site occupied by a ligand can disturb the binding strength of other neighbouring binding sites. This phenomenon of ligands to influence other binding sites is known as cooperative binding (Lovell. 1998).

Such cooperative binding can be negative or positive. If due to the influence of a bound ligand the strength of nearby binding sites increases, it is called positive cooperativity and vice versa (Goldbeter et al. 1990). The unique nature of the positive cooperativity on the rate of a reaction is the generation of an “S” shaped sigmoid curve. Positive cooperativity is shown in figure 2-3, and a corresponding Michaelian curve is shown for comparison.
Figure 2-3 A plot comparing positive cooperativity (red curve) to a hyperbolic response (blue curve).
In a system involving gene promoters and TFs, the transcriptional process is controlled by cooperativity. Based on varied signals, regulation of transcription happens through a myriad of cooperation events among various TFs (Perutz et al. 1989). Effect of these signals can be either positive or negative and in addition it can lead to competition between similarly primed TFs (Levitzki et al 1969). If these binding signals which are usually activated by changes in environmental changes, are positive, cooperation (i.e., affinity) between different TFs is primed to increase and they come together readily in order to form a complex (Levitzki et al. 1969). And vice versa in the case where environmental signals are negative. In biological systems at the level of transcription, non-cooperativity is rarely the case, where in a gene promoter TFs can regulate transcription by binding singularly. Estimation of cooperativity is often done as a Hill regulation as shown in Eqn 1.4 (Bell et al. 2007; Werner et al. 2007).

2.4.3 Hill equation

The Hill equation was mathematically formulated as a way of characterising the binding of haemoglobin (oxygen-transport metalloprotein) and oxygen molecules in the blood stream (Hill. 1913). Subsequently, this mathematical representation was found to be not correct with respect to oxygen transport. Nevertheless, the Hill equation provided a start to further mathematical modelling approaches.

Let’s consider a hypothetical protein with multiple subunits \(n\), with each having a ligand \(S\) binding site. As mentioned in the previous section, if one binding site is occupied by a ligand, it results in change of binding strength of all other existing nearby binding sites \((n-J)\). This can be represented as the following reaction,

\[
E + nS \leftrightarrow ES
\]  

(1.1)

where, concentrations of enzyme which has the binding site is denoted by \(E\), \(S\) is the ligand and ligand-enzyme complex is denoted by \(ES\). Assuming that the reaction reaches equilibrium rapidly we can write the above reaction (1.1) as follows,

\[
K = \frac{ES}{E.S^n}
\]  

(1.2)
in which, $K$ is the association constant for the binding reaction.

If the total enzyme concentration is denoted by $E_t$, then it can be represented as $E_t = E + ES$, and by substituting the value of $E$ by $E_t$ in Eqn (1.2), the relationship of the binding reaction can be derived as follows,

$$\frac{ES}{E_t} = \frac{S^n}{1 / K + S^n} = \frac{S^n}{K_d + S^n}$$  \hspace{1cm} (1.3)

The above derivation (1.3) is called the Hill equation in which, $K_d$ is known as the disassociation constant and $n$ is the Hill coefficient. The Hill equation can also be expressed as follows,

$$v = \frac{V_{max} S^n}{K_d + S^n}$$  \hspace{1cm} (1.4)

The binding of a ligand ‘$S$’ to a binding site ‘$n$’ as expressed in the Hill equation poses a major challenge, since in biological systems; the number of binding sites is not a concrete integer, but only estimates of ‘$n$’ are always known. Nevertheless, using the Hill equation in a mathematical model serves the purpose by giving a sigmoid “S” shaped output, which is used to model cooperativity of TFs binding to promoter regions. The Hill equation severely lags behind other approaches when the complexity increases, for e.g. by adding additional regulatory components such as multiple TFs and protein elements. Further, Hill type regulation is not suitable for realistic *in vivo* or *in vitro* interactions which are complex reversible reactions.

### 2.5 Translation

Translational regulation of circadian mRNAs has not been considered for model development due to lack of necessary information in the current literature, but it is important to understand the process of protein synthesis. The process of expressing the gene embedded in DNA, through to production of the protein in eukaryotes is discussed briefly at this point to promote an understanding of the process.
2.5.1 Three step process

As in the case of transcription, translation is also typically described as occurring in three stages: initiation, elongation, and termination (Soneneber et al. 2009). The binding of initiation factors (IF) to DNA is the first of these translation steps. In eukaryotes there are at least ten such proteins (eIFs) designated as initiation factors, and a combination of eIFs attracts the two ribosomal subunits 40S and 80S resulting in the formation of an 80S initiation complex (Pestova et al. 2002).

The binding and formation of the initiation complex usually occurs at specific initiation sites referred to as 5’ un-translated regions (UTR). A UTR is a part of the coding region that is occupying positions from the initial nucleotide at the 5’ end and the start codon (AUG), with AUG coding for aminoacid methionine (Kevil et al. 1995). After the initiation complex has assembled, the second step in translation is extension of the polypeptide chain in a process called elongation in the literature.

2.5.2 Achieving critical mass

Several ribosomes can be translating the same mRNA strand at different positions on it. As the first ribosome moves downstream proceeding to translate the gene, the next ribosome occupies the initiation site to start assembling another identical peptide chain (Ringquist et al. 1992). Consequently the multiple ribosomes on the mRNAs, can be spaced at intervals from 100 to 200 nucleotides apart. The identical, though separate, polypeptide chains produced by a series of multiple ribosomes results in increasing concentration of protein copies in the cytoplasm in a time-dependent manner (Kennel et al. 1977), which is denoted in this model as time evolution, else transient concentration of the circadian proteins.
Chapter 3: Concepts in biosystems modelling

3.1 Concept of expression rates

The environment imposes severe limitations on the radius of action of all living organisms, and the eukaryotes in particular exhibit an astonishing number of genetic responses to their environment. For example the TFs involved in regulating circadian rhythm of *Drosophila melanogaster* are directly influenced by environmental cues such as light, temperature etc. (zeitgebers), and by deduction it can be established that TFs control important aspects of gene expression patterns (Reppert et al. 2002). Additionally, by association with chromatin the TFs exhibit more broad transcriptional regulation of a gene at various layers resulting in different rates of expression, a factor described in this study as the rate of transcription (Mitchell et al. 1989).

Even though transcriptional control is established as the initial stage of gene expression, the regulation of mRNA is also effected by other specific factors (Shen-Orr et al. 2002). Generally, regulation of transcription is moderated through specific proteins which binds at the 5’ un-translated region of the mRNA sequence thus blocking the translation of the gene downstream on the DNA, e.g. the translational regulation of ferritin, an iron storing intracellular protein (Cairo et al. 1995).

Although translational regulation at the mRNA level as currently described in the literature demonstrates a lack of resolution and understanding compared to the detailed knowledge concerning transcriptional regulation, it is established that post-translational modifications are major factors affecting the synthesis of functional proteins (Beevers. 1982). For modellers, these multiple regulations of protein synthesis offer added freedom during conceptual model development and these regulations are collectively represented mathematically in this study as the rate of translation. The known post-translational modifications with respect to circadian proteins are discussed in detail in the following chapters.
3.2 Enzyme-substrate interactions

“Enzymes are biological catalysts”. i.e. like any other catalysts in a chemical reaction, they reduce the energy of a reaction, but don’t change their shape or form in terms of their overall concentration in the reaction. In \textit{in vivo} and \textit{in vitro} systems every biochemical process takes place with the help of unique enzymes (Fersht. 1999). In molecular biology, enzymes are nothing but proteins with distinct shape and structure to accommodate other proteins (substrates). They form a reversible enzyme-substrate complex (protein dimers) resulting in the formation of a product (Hayashi et al. 1963). Two known models of enzyme-substrate complex (E-S) are as follows.

3.2.1 Lock-key model

The earlier model representing the formation of E-S is the \textit{lock-key} model (Koshland et al. 1995). In the lock-key model, only a specific region of the enzyme is primed to form an E-S complex with the substrate. This region, usually a few amino acids long is known as the active site (Figure 3-1). A major aspect of the lock-key model is that, the active site is specific to only its matching substrate. This differs from, the \textit{induced-fit} model, in which the active site of an enzyme has the ability of minute shape changes in order to form E-S complex (Benkovic et al. 2003).

3.2.2 Induced-fit model

The induced-fit model is the contemporary and widely agreed model of E-S interaction (Figure 3-1). In this model, the initial E-S complex is the result of a weak interaction between an enzyme and substrate, but progressively this weak interaction leads to change in shape of the enzyme. Change in the structure of the enzyme reduces the distance between enzyme catalytic sites and covalent bonds of the substrate, which culminates in forming a strong E-S complex. Enzyme catalysis brings about a temporary transition E-S complex and products.

Four types of enzyme catalysis are possible according to the induced-fit model.

\textbf{1. Bond strain:} In this type of catalysis, the conformational change of an enzyme as a result of the E-S complex happen due to weak covalent bonds resulting in the fast formation of a temporary E-S complex. This initial E-S complex results in bringing
Figure 3-1 Enzyme-substrate interactions: (a) Lock-key model and (b) Induced-fit model.
the stronger substrate atoms in large amino acid groups closer in order to form a more stable E-S complex, e.g. aspartate, glutamate, etc.

2. **Proximity and orientation:** In this type of E-S catalysis the formation of the initial E-S complex brings highly reactive catalytic group of amino acids such as aspartate in close proximity to one another. The orientation and proximity of the enzyme and substrate active regions determine their usefulness in forming the E-S complex.

3. **Proton donors and acceptors:**

   This type of catalysis is generally characterised by the involvement of a proton donor (acid) or acceptor (base) molecule. The temporary E-S complex is achieved by the donation of a proton from an acid molecule and accepting the donated proton by a base molecule. In acid base catalysis, enzyme and substrate can be either acid or base.

4. **Covalent mechanism:** In covalent catalysis, orientation of the substrate towards the active site of the enzyme is important. At the right orientation, the E-S complex is formed by a series of E-S covalent intermediates.

3.2.3 **Multiple substrate enzyme reactions**

Multi-substrate reactions follow complex rate equations that describe how the substrates bind and in what sequence. The analysis of these reactions is much simpler if the concentration of one substrate is kept constant and the other substrate is varied. Under these conditions, the enzyme behaves just like a single-substrate enzyme. For an enzyme that takes two substrates A and B and turns them into two products P and Q, there are two types of mechanism, a) sequential and b) ping-pong.

a) **Sequential mechanism:**

In these enzymes, both substrates bind to the enzyme at the same time to produce an EAB ternary complex (Figure 3-2). The order of binding can either be random (in a random mechanism) or substrates have to bind in a particular sequence (in an ordered mechanism). When a set of v by [S] curves (fixed A, varying B) from an enzyme with a ternary-complex mechanism are plotted in a Lineweaver–Burk plot, the set of lines produced will intersect.
Figure 3-2 In the above figures, plot A shows a Lineweaver–Burk plot of multiple substrate enzyme reactions in sequential mechanism and plot B shows a Lineweaver–Burk plot of multiple enzyme substrate reactions in a ping-pong mechanism.
b) Ping-Pong mechanism:

In this mechanism, one substrate bind first to the enzyme followed by product P release. Typically, product P is a fragment of the original substrate A. The rest of the substrate is covalently attached to the enzyme E, which is designated as E’. Now the second reactant, B, binds and reacts with the enzyme to form a covalent addition with the A as it is covalently attached to the enzyme to form product Q (Figure 3-2). This is now released and the enzyme is restored to its initial form, E. When a set of v by [S] curves (fixed A, varying B) from an enzyme with a ping–pong mechanism are plotted in a Lineweaver–Burk plot, a set of parallel lines will be produced.

3.3 Kinetics of expression rates

Whether oscillatory, or not, bio-chemical pathways, are complex as a consequence of non-linearly behaved components, and this establishes the requirement for a mathematical framework that represents such dynamics accurately (Tyson et al. 1978). The first analytical step to is to consider these components as a set of chemical species that are intra-convertible through chemical processes. Although these processes are usually described as a set of reversible binding steps, else irreversible catalytic steps, this does constitute a network of elementary reactions amenable to mathematical analysis (Deville et al. 2003). Every chemical reaction in a cell is catalysed through enzymes, and these enzymes are selective and particular in the choice of the reaction they participate in, which are usually reversible. They rapidly increase the reaction rates without getting consumed themselves. A specific reaction is catalysed by a specific enzyme both in forward and reverse directions.

In biological systems, reactants are termed as substrates. In a typical enzyme catalysed reaction, the ratios of enzyme molecules to products and substrates are usually in the order of one enzyme molecule to a thousand (or more) substrate or product molecules. Therefore, one enzyme molecule can catalyse the production of an enormous number of products from substrates (Cleland. 1963). As discussed in the previous section, the catalytic event of substrate to product conversion is mediated by a transitional E-S complex. This catalytic interaction usually starts at the specific active region on the enzyme molecule. Products are synthesised from a reaction by the breakdown of an E-S complex releasing the enzyme molecule to catalyse a new
reaction (Hammes. 2002). An overall picture of various basic reaction types and associated rate laws is shown in figure 3-3. The most frequent method for expressing E-S is through Michaelis-Menten kinetics.

3.3.1 Michaelis-Menten kinetics

The derivation of Michaelis-Menten kinetics to describe an enzyme catalysed biological reaction is a major milestone in describing reaction mechanisms in biochemistry. This mathematical description was established firmly by Michaelis and Menten as proposed by Brown and Henry earlier (Michaelis et al. 1913). It offered a reaction scheme for the formation of an enzyme substrate complex (ES) by the binding of a free enzyme to the substrate. Ensuing the E-S complex further transforms, synthesising a product with free enzyme. For example, consider the following reaction.

\[ E + S \xrightleftharpoons{\kappa_{-1}}^{k_1 \, \kappa_2} ES \rightarrow E + \text{products} \]  

(1.5)

where, enzyme is denoted by E, substrate (reactant) is denoted by S, and the intermediate E-S complex is denoted by ES. Further, k1, k2 and k-1 are individual reaction rates. For the above reaction, the rate of synthesis of the product is known as reaction rate and it is represented as follows,

\[
\frac{d(\text{product})}{dt} = k_2[ES] 
\]  

(1.6)

As we know before, the E-S complex (ES), is a short lived temporary step. Hence, steady state approximation can be used to get the overall reaction rate or the rate of change of ES and it can be derived as follows;
Figure 3-3 Comparing Kinetics: (a) Michaelis-Menten Kinetics and (b) Hill Kinetics.
\[ \frac{d[ES]}{dt} = k_1[E][S] - k_{-1}[ES] - k_2[ES] \approx 0 \]  
(1.7)

By solving \([ES]\), we get,

\[ [ES] = \frac{k_1[E][S]}{k_{-1} + k_2} \]  
(1.8)

And by substituting \([ES]\), in Eqn (1.6) we get,

\[ Rate = \frac{d[product]}{dt} = k_2 \frac{k_1[E][S]}{k_{-1} + k_2} \]  
(1.9)

If the Michaelis-Menten constant \(K_M\) is denoted as follows,

\[ \frac{1}{K_M} = \frac{k_1}{k_{-1} + k_2} \]  
(2.0)

Then by substituting it in Eqn (1.9) the rate becomes,

\[ Rate = \frac{k_2}{K_M} [E][S] \]  
(2.1)

In general biochemical reactions, the amount of free (un-complexed) enzyme \([E]\) cannot be quantified. But usually, the total enzyme concentration \([E]_o\) is known and it can be expressed as follows,

\[ [E]_o = [E] + [ES] = [E] + \frac{[E][S]}{K_M} \]  
(2.2)

\[ = [E] \left( 1 + \frac{[S]}{K_M} \right) \]  
(2.3)
From which we obtain the free enzyme concentration \([E]\) as follows,

\[
[E] = \frac{[E]_0}{1 + \frac{[S]}{K_M}}
\]  
(2.4)

By substituting the value of \([E]\) in the Eqn (2.1) the rate becomes,

\[
Rate = \frac{K_2 [S]}{K_M} \left( \frac{[E]_0}{1 + \frac{[S]}{K_M}} \right)
\]

\[
= K_2 \frac{[E]_0 [S]}{K_M + [S]}
\]  
(2.5)

(2.6)

And by defining the reaction velocity as \(v = Rate\) we get,

\[
v = K_2 \frac{[E]_0 [S]}{K_M + [S]}
\]  
(2.7)

From the above Eqn (2.7) it can be seen that when \([S] = 0\), the reaction velocity \(v = 0\).

And reaction velocity \(v\) will increase proportionally to increase in substrate concentration \([S]\). The maximum reaction velocity \(v_{max}\) is achieved when the amount of \([S]\) tends to infinity. By using this, we can derive the maximum velocity, \(v_{max}\), as follows,

\[
v_{max} = \lim_{[S] \to \infty} K_2 \frac{[E]_0 [S]}{K_M + [S]} = k_2 [E]_0
\]  
(2.8)

And by substituting the value of \(k_2[E]_0\) in rate Eqn (2.7) we get,

\[
v = \frac{v_{max} [S]}{K_M + [S]}
\]  
(2.9)
From the above Eqn (2.9) it can be seen that the reaction kinetics are determined by the parameters, $v_{\text{max}}$ and $K_M$. In models using Michaelis-Menten kinetics, these are the parameters that usually denote kinetics of biochemical reactions.

A plot of substrate concentration versus reaction velocity is provided in Figure 3–4. The points A, B and C of the plot denote key aspects of the reaction. The point C represents the reaction rate at $v_{\max}$ where, the substrate concentration $[S]$ is at the maximum. At this point, difference in substrate concentration will be inconsequential to the reaction rate. At point C the reaction is happening at zero-order kinetics and the reaction rate is entirely under the influence of enzyme concentration. At point C, the entire amount of the enzyme is utilised and there are a very few if any free enzyme molecules. At enormous concentration of the substrate, the reaction rate is $v_{\text{max}}$ in the Michaelis-Menten plot. But the reaction rate could still be affected by any other reactant e.g. if the reaction has one or more substrates, at which point the other reactions may not be in zero-order. At points A and B, the substrate concentrations are lower compared to point C. From Eqn (3.1) we can see that at these two points the reaction velocities will be considerably lower than point C. Thus, only a fraction of enzyme is utilised at this stage of the reaction and the concentration of E-S complex will be very low. As we can see from the plot, at point B, the reaction rate is exactly 50%. Here, only half of the total enzyme is utilised for the E-S complex. At this point the reaction rate is half of $v_{\max}$ and equal to $K_M$, the Michaelis-Menten constant. The reaction rate is zero and known as first-order when the substrate concentration is at its lowest, which is shown in the plot as point A.

In order to deal with experimental data, the Michaelis-Menten equation was reconstructed by Lineweaver and Burk (Lineweaver et al. 1934) as follows.

$$\frac{1}{v} = \frac{K_M}{v_{\text{max}}[S]} + \frac{1}{v_{\text{max}}} \tag{3.0}$$

The usefulness of the above Eqn (3.0) is that, in a biochemical reaction, reaction rate $v$ can be measured as a function of total substrate concentration $[S]$. Hence, a plot of $1/[S]$. 27
Figure 3-4 A Plot of substrate concentration versus reaction rate.
against 1/ν will give a straight line with “slope” \( K_M/ν_{\text{max}} \) and intercept 1/\( ν_{\text{max}} \). This is useful to avoid curvilinear plots.

\[
K_M = \text{slope} \times ν_{\text{max}} = \frac{\text{slope}}{\text{intercept}}
\]  

(3.1)

Eadie-Hofstee transformation (Hofstee. 1952), is a different rearrangement of the Michaelis-Menten equation which is as follows.

\[
\frac{ν}{[S]} = ν \left[ \frac{1}{K_M} \right] + \left[ ν_{\text{max}} \right] \frac{1}{K_M}
\]  

(3.2)

In addition to Michaelis-Menten equation, Eadie-Hofstee’s and Lineweaver-Burk transformations are used to describe enzyme catalysed reactions.

A major disadvantage in using Michaelis-Menten kinetics in genetic regulatory networks is that, while the E-S complex formation will always lead to the synthesis of a product and release of free enzyme, there is no reverse reaction wherein E-S complex can unbind to revert back to the original components of enzyme and reactant in the reaction. Michaelis and Menten considered that this reverse reaction is not energetically robust and that it will spend more energy than the forward reaction. In biological systems this is not the case e.g. in a living cell i.e. in \textit{in vivo} gene expression regulation, TFs can bind and unbind dynamically, and can exist in multiple bound or unbound states at any given time point. This can be overcome by characterising the reaction with mass action kinetics. Using a mass action law, a biochemical reaction can be split into many differential equations, with each equation representing a particular substrate (reactant) irrespective of whether this substrate leads to product or reverts back to the initial state.

\textbf{3.3.2 Mass action kinetics}

In chemistry and chemical engineering mass-action kinetics is used mathematically to describe the dynamics of systems of chemical reactions, i.e. reaction networks (Horn et al. 1972). Such models are a special form of compartmental systems, that involve mass and energy balance relations. Mass-action kinetics as a mathematical technique
has numerous analytical properties that are of inherent interest to a biologist studying
dynamical-systems perspectives. For example mass-action kinetics gives rise to
systems of differential equations having polynomial non-linearities, and these
equations are notorious for their intricate analytical properties in even low-
dimensional cases (Szederkényi et al. 2010). Yet mass-action kinetics have special
properties, such as non-negative solutions, that are useful for analysing their
behaviour.

Mass action kinetics can be used to model (approximately) the behaviour of
reversible, and irreversible enzymatic reactions for unregulated enzymes that are not
near saturation point.

Consider the following reaction,

\[ E + S \rightleftharpoons_{k_2}^{k_3} ES \rightarrow E + P \]  \hspace{1cm} (3.3)

Thermodynamically in a biochemical reaction, any two substrates (reactants) will
have enough energy for a collision to take place. These collisions will increase with
the number of substrates.

In the above mass action reaction (3.3), the mass action kinetics state that the rate of
an enzyme (E) reaction (E+S) is the product of a rate constant \( k \) and mass of the
reactants (E) and (S). The mass in our case is concentration of the substrate [S] and
Enzyme [E] of the reaction.

\[ v = k [E][S] \]  \hspace{1cm} (3.4)

As a result of using the mass action rate law, this reaction leads to the following four
differential equations.

\[ \frac{d[S]}{dt} = -k_1[E][S] + k_{-1}[C] \]  \hspace{1cm} (3.5)
\[ \frac{d[E]}{dt} = -k_1[E][S] + (k_{-1} + k_2)[C] \]  

(3.6)

\[ \frac{d[C]}{dt} = k_1[E][S] - (k_{-1} + k_2)[C] \]  

(3.7)

\[ \frac{d[P]}{dt} = k_2[C] \]  

(3.8)

For biochemical network models mass action kinetics is usually applied to study the effects of the concentrations of the chemical species in these reactions, and by implication this usually requires explicit mathematical expressions for the velocity of each reaction. The most popular representation for these models, which is also the method that we followed in this thesis, is based upon using ordinary differential equations (ODEs) governed by mass action kinetics to describe any variations in the concentrations of the chemical species.

### 3.4 Genetic regulatory networks

The relationships between average expression rates and average concentrations of transcription factors are the basis of our analysis of the gene regulatory network. As proteins in a cell are often present in relatively low levels, stochastic fluctuations in the copy numbers of molecules can have important effects on the dynamics of gene regulation (Elowitz et al. 2002; Pedraza et al. 2005) and the analysis of the noise produced by such fluctuations has been the subject of many recent advances in the analysis of regulatory networks (Munsky et al. 2009). The word stochastic in a set of chemical reactions the word refers to a random variable whose successive values are not independent.

Expressed genes are under regulatory control as described above, and TFs themselves are subject to the same kind of control mechanisms thus giving rise to complex genetic regulatory networks (GRN). A typical genetic regulatory network involves
interaction between proteins (TFs) and DNA sequences (Promoters) in a positive or negative fashion to form feedback loops. Depending on the positive or negative influence of the TFs on gene expression, the resulting control is called a positive or negative feedback loop. Sometime an aggregate of many positive and negative regulatory feedback signals result in a complex feedback network.

3.4.1 Modelling genetic regulatory networks

Genetic regulatory network models start with a particular research question and consist of various components. Individual gene expression in GRN models generally follow a typical framework wherein certain proteins (TFs) act upon a particular gene sequence resulting in the synthesis of RNAs and eventually proteins. The major aim of this linear approach (TFs \( \Rightarrow \) Promoters \( \Rightarrow \) Proteins) is to estimate manageable and biologically realistic parameters in order to reflect \textit{in vivo} findings. A model which satisfies the vigorous constraints of \textit{in vivo} or \textit{in vitro} data could be used to analyse the system at the appropriate scale. Such models can also be used to test various hypotheses resulting in assigning functions to unknown components (TFs or gene products) or contribute towards deeper understanding of the research question. Researchers achieve this goal through stochastic or deterministic approaches. Irrespective of the mathematical approaches used, every model has a positive or negative trade off due to the inherent scalar issues in modelling GRNs which is discussed in detail in section 3.4.3. A brief summary of modelling options is discussed as follows.

The type of research question determines the selection of a modelling technique to a large extent. There are many detailed and widely known modelling approaches for different GRN levels (De Jong et al. 2002), such as Boolean networks and ODE models of prokaryotes (Smolen et al. 2000), modelling GRNs of eukaryotes (Bolouri and Davidson et al. 2002), GRN modelling from a molecular level to a network level (Schlitt and Brazma et al. 2005), and developmental GRNs specific to large scale developmental pathways (Longabaugh et al. 2005). Similarly there is a comprehensive knowledge of the mathematical tools available for implementing and analysing genetic regulatory networks (Alves et al. 2006).
3.4.2 Noise in GRNs

Organisms with similar genotype under the same circadian cues exhibit a variety of change in phenotype (Delbruck et al. 1945). This phenomenon is observed from simple prokaryotes to complex eukaryotes. Such phenotypic characteristics are important in evolutionary terms, since they are known to be a major factor through which organisms adapt to change in environment. For example, the sensitive phenotypic response of plants to their environment such as flowering, shoot length and colour can be observed with the naked eye. The genome of any living organism does not remain static throughout its lifetime, and the changes although generally not observable in the phenotypes are nevertheless present in the genes. In biology stochasticity is universal, for example even identical human twins sharing near-identical genomes at birth will differ genetically, but very subtly, in their old age.

Hence, stochasticity of individual gene expression and the overall patterns of gene expression in an individual cell are studied in detail (Jaakkola et al. 1994; Gupta et al. 1990). Until a decade ago, quantifying the stochasticity of individual genes was not possible. But thanks to the cloning method of fluorescent protein markers, it can be quantified in real-time (Elowitz, Levine et al. 2002; Ozbudak, Thattai et al. 2002; Raser et. al 2004). Genetic regulatory networks and their associated control can be studied in a more detailed fashion to understand the implications of stochasticity (Elowitz et al. 2002; Austin et al. 2006).

Stochasticity in dynamical systems such as bio-systems is the result of two types of perturbations, which are inherent and omnipresent at any levels, intrinsic noise and extrinsic noise. Intrinsic noise can be defined as local variations present within any system e.g. intracellular. In contrast, extrinsic noise is controlled by external factors i.e. outside the cell e.g. cell-to-cell variation (Swain, Elowitz et al. 2002; Paulsson 2004). There are various theoretical studies, done by a variety of groups trying to differentiate intrinsic and extrinsic noise (Elowitz et al. 2002; Paulsson et al. 2004), and experimental studies (Ozbudak, Thattai et al. 2002). In biological systems both prokaryotic and eukaryotic extrinsic noise play a major role (Elowitz et al. 2002; Raser et al. 2004). A brief discussion of intrinsic and extrinsic noise is as follows.
• **Intrinsic noise**

Intrinsic noise is usually the product of stochasticity inherent in biological systems e.g. genetic regulatory networks. It is assumed that, in a living cell, molecular components such as proteins need to diffuse randomly in order to find their appropriate functional locations. Proteins and other biomolecules perform their biological function through collisions with complementary partners. Even under the presumption that functional proteins are well mixed in a particular cell, the process by which these collisions take place are probabilistic and not continuous and deterministic. This randomness leading to intrinsic noise is extremely important in order to understand the system theoretically, since participating molecules in a genetic regulatory network are of very low copy numbers and impossible to measure in real time (McAdams et al. 1999). An example of a low copy number GRN component is lac operon, where; it is found that one of its major components, the lac repressor copy number is only ten (Lewin et al. 2004).

• **Extrinsic noise**

Extrinsic noise is the stochasticity emanating under the influence of environmental factors. These environmental factors play an important role in the control and behaviour of an organism both at the gene expression and physiological levels (Elowitz et al. 2002; Raser et al. 2004). Various external factors are known to influence the parameters in a molecular reaction (Dolomanov et al. 2003), such as light, temperature, food, etc. Extrinsic noise is extremely important in circadian rhythms, since it is known that the environmental cues (zeitgebers) determine the amount and timing of molecular components of a cell in their respective time of the day. Thus, stochasticity of extrinsic noise is autonomous to the actual components of a GRN and the molecular collisions (Elowitz et al. 2002).

• **Roles of noise in biological systems**

The fear of noise in physical and engineering systems leading to destruction or failure is somewhat contrasting to noise in biological systems; where, it has both positive and negative connotations. Noise can lead to harmful variations from uniformity (Meir et
al. 2000); yet in evolutionary terms, intrinsic noise at the level of genetic mutations can lead to physiological adaptation and evolution of a new species (Kaneko et al. 2008). Theoretically, in gene expression networks noise can fine tune the role of various molecular components by making them more sensitive to minor changes in concentration and time (Paulsson, Berg et al. 2000). Another role for noise in biological systems is that, through the production of genetic stochasticity, it can lead to switching of phenotypic states in a cell (Mc Adams et al. 1999; Hartwell et al. 1999). This stochastic stabilisation enables a cell or an organism to adapt to different environmental conditions (Kussell et al. 2005).

3.4.3 Multi-scale issues in modelling biochemical systems

In mathematical modelling of biological systems, simplistic characterisation is inevitable as for every biological rule there is inevitably an exception (Gold. 1977). Hence, it is necessary to make various assumptions and disregard the influence of less important system components. It is important to keep in mind to include the effects of certain major rate limiting factors so that the overall behaviour of the system is not compromised (Bellouquid et al. 2005). Physical properties of a molecule are governed by well-known physics, e.g. molecules at a visible macroscopic level behave with the rules of classic Newtonian law. But universally, to understand the behaviour of molecules at the microscopic and mesoscopic levels quantum mechanics is needed. A major problem in modelling biological systems at molecular level is the lack of measurable data (Aderem. 2005). Similar to any other models, modelling molecular pathways is done to focus on a particular research question. For the same research question, a modeller can choose to follow different time and size units depending on the availability of validation data as long as the measurable output is similar to wet-lab results. Different scalar levels in bio-systems are as follows.

- **Macroscopic:** It is generally assumed on the macroscopic level that the components are uniformly mixed and consistently identical. Macroscopic system reactions are of highest order compared to microscopic and mesoscopic scales (Dada et al. 2011). In these systems, all participating molecules behave similarly to corresponding molecules and are expressed using ordinary differential equations. Here, the biochemical reactions are time dependent and continuous. The model output is usually
measured by interpreting the increase or decrease of reaction species with respect to
time.

• **Microscopic:** At microscopic level, chemical reactions are investigated at the
atomic and ionic elementary levels (Dada et al. 2011). Here, the unit to represent the
colliding atoms is known as mole (mol). One mole of elementary particles is given by
the Avogadro constant which is \(6.022141 \times 10^{23}\) molecules. The dynamics of a
microscopic system is of stochastic nature since, it is defined by the individual
dynamics of each elementary particle in this system.

• **Mesoscopic:** In mesoscopic reactions of biological systems, the scale of measurable
participating molecules falls roughly between the atomic and macroscopic level (Dada
et al. 2011). Molecules at this level have more in common with macroscopic than
microscopic counterparts, since the individual properties of atoms are ignored and the
collisions takes place at a molecular level. Similar to the macroscopic scale, at
mesoscopic level we can assume that the system is uniformly mixed. But the system
is not homogenous, thus every molecular collision can be random leading to
stochastic dynamics.

### 3.4.4 Randomness of biochemical reactions

It is known that identical cells grown in the same environment frequently exhibit
distinct characteristics. These differences are the result of random fluctuations in
biochemical reactions. Biologists had always thought of such biochemical fluctuations
as unwanted, but recent studies suggest that cells and bacteria sometimes utilize this
randomness to their benefit. Small systems such as cells are inherently sensitive to the
random effects scientists call stochasticity or noise because they contain only a few
active copies of individual proteins or nucleic acids. Minor fluctuations in the levels
of some cellular components, for example, affect whether a particular gene turns on
and makes a protein. Such noise seems to suggest that some aspects of cell fate are
left to randomness. This lack of control forces cells to evolve redundant biochemical
pathways in order to survive.

The eye of the common fruit fly, *Drosophila melanogaster*, comprises smaller units,
each consisting of eight cells. When each cell develops, it makes a choice determined
by the presence or absence of a regulatory protein. This protein becomes active only in a random subset of the cells, and its occurrence determines whether the cell will respond to a particular hue of ultraviolet light. Random expression of this regulatory protein ensures that the two cell types are apportioned throughout the eye by chance so as to avoid repetitive patterns that could limit the fly’s overall vision. Even though the cells are in an identical environment and they all come from an identical ancestor, they acquire different phenotypes.
Chapter 4: Circadian clocks

4.1 Importance of clock

The digestive system of most organisms with feeding behaviour during the day is an excellent example of circadian regulation in various tissue types and cells. Forage for food during the day involves regulation of physiology and behaviour (Moore et al. 1986). These regulatory signals, mostly molecular components of a circadian clock, control the awareness of sense organs to detect food, the usage of motor organs to reach the food and finally the resulting digestive process after the consumption of food. Even with minor changes to temperature and light e.g. during full moons, this locomotor activity rhythm is greatly affected (Panda et al. 2002). In addition to timing of day light and night, a circadian clock helps an organism in adapting to seasonal changes (Dan et al. 1975).

To follow the locomotor activity, circadian clocks need a light-input system that synchronises the clock to day–night cycles, a biochemical and cellular oscillator to measure the passing of time and output mechanisms to relay this timing information to the primary systems that regulate physiology and behaviour. Research in several model organisms such as Synechococcus elongates, Cyanobacterium, Neurospora crassa, Arabidopsis thaliana, Drosophila melanogaster, Mus musculus and humans uncovered many aspects of these basic mechanisms of oscillator function (Bell-Pedersen et al. 2005).

The light entrainment systems of various organisms differ due to their different photoreceptive systems. But, the circadian clock components, their oscillatory behaviour in various biological systems and their methods of regulating locomotor activity have a common framework, showing that the circadian clock is evolutionary conserved (Wager-Smith et al. 2000). Clocks also have other interesting properties such as temperature compensation and genetic robustness. As temperature increases, the speed of most biochemical reactions also increases. But the daily rhythms still exist in organisms living in environments where the temperature may change tens of degrees during the day and even more between seasons. For example cold blooded
animals like the common frog (Rana temporaria) have a circadian clock that accounts for this (Harri. 1972). Interestingly, clocks in warm blooded animals like mammals are also temperature compensated, hinting at evolutionary conservation. Another interesting property is genetic robustness. As clocks are so crucial for life, organisms need a genetic architecture that accommodates mutations without losing clock function (Panda et al. 2002). These properties of endogenous and robust oscillations show that they are produced by an architecture of genes and proteins interacting together rather than being traits conveyed by single genes.

4.2 Model organism Drosophila

Drosophila commonly known as ‘fruit fly’ is a famous model organism whose genetics has been studied in a great fashion. Starting from the famous Hunt Morgan’s Fly room in Columbia University, this small insect helped transform the whole gambit of our knowledge in genetics and developmental biology, and even more recently in the field of chronobiology (Morgan et al. 1925).

Since regulation of circadian rhythm is a daily cycle and several features of Drosophila behaviour and physiology are confined to certain intervals of the day, it was encouraged as the appropriate model for investigating circadian control at molecular level (Winfree. 1974). During its lifecycle, the evolution from pupa to adult stage happens during early morning at cooler temperatures. The emerging adult Drosophila flies shed their pupa and later during the day their cuticle gets harder and wings are visibly expanded from their folded condition. After training the pupae to 24 hour light/dark (LD) cycle, it was observed that even by changing the exposure to constant darkness (DD), the adult flies emerge from their pupal cases in the early morning (Konopka et al. 1971). This indicates that an internal pacemaker independent of light entrainment is present in Drosophila melanogaster. The foraging behaviour of the flies is a daily cycle where, mating and forage tendencies are observed during the day, and they remain relatively unresponsive to sensory stimuli and exhibit homeostasis during the night (konopka et al. 1989).
4.2.1 Organisation of *Drosophila* clock neurons

Organisation of the *Drosophila* circadian clock is well known through major studies using various molecular techniques, like fluorescent microscopy and, immunoblotting, a technique where, antibodies are synthesised against circadian proteins and genes with regulatory sequences driven by desired circadian gene promoters are identified subsequently. More recently functional microarray techniques are used to screen the molecular components of the clock. In the brain of *Drosophila melanogaster*, about 150 cells express the circadian components (Matsumoto et al. 2007) and they are divided into major subgroups (Figure 4-1). Between the central brain of the fly and its optic lobe there is a group of neurons which express clock genes and are known as lateral neurons (LNs). This group is further subdivided based on the size of neuron groups, a larger group known as large lateral neurons (lLNs) and a smaller group known as small lateral neurons (sLNs). In addition, a group of LNs located on the dorsal region of the brain are known as LNd. Apart from LNd, dorsal neurons are further grouped in to three, DN1, DN2 and DN3, which are collectively called as DNs.

The group of neurons found in the posterior section of the brain are known as lateral posterior neurons (LPNs). Every neuronal group is known to have specific functions in circadian control. sLNs primarily play a role in morning and evening activity rhythms (Stoleru et al. 2004; Grima et al. 2004) and by contrast lLNs are involved in phase shifting during late night (Shang et al. 2008). It has been observed that, locomotor activity rhythms still persist in flies under constant darkness with total deletion of LNs and DNs (Dushay et al 1989; Murad et al. 2007). In the pupal stage, entrainment of temperature is primarily controlled by DN2 in the DD cycle, but in the LD cycle temperature entrainment is regulated in LNs through a photoreceptor known as “Pigment Dispersing Factor” (PDF), which is the principal transmitter of light signal in *Drosophila* (Picot et al. 2009).
Figure 4-1 Schematic diagram of *Drosophila* cerebral cells showing seven main groups of neurons.
4.3 Summary of *Drosophila* clock components

The first clock component identified in *Drosophila* is the “Period” (*per*) gene, found through EMS mutagenesis (Konopka et al. 1971). Based on the mutational studies and observed period rhythmicity, three alleles of per mutant were discovered as follows; an arrhythmic *per* null allele (*per*\(^0\)), a rhythmic *per* allele with daily shortened period (*per*\(^s\)) and a rhythmic *per* allele with daily elongated period (*per*\(^l\)) (Konopka et al. 1971). By subsequent cloning and characterisation of the *per* gene product, the PER protein was identified to have a functional domain similar to known eukaryotic proteins such as, “Single Minded” (SIM) and “Arhl hydrocarbon Receptor Nuclear Transport” (ARNT) proteins (Bargiello et al. 1984; Thomas et al. 1988; Hoffman et al. 1991). Aiding in protein-protein interaction was later found to be the exact function of this domain, and it was named PAS domain (Huang, et al. 1993).

Protein hybridisation studies identified that the PER protein exists as a heterodimer with the product of another clock gene *tim* forming a PER/TIM dimer complex through their PAS domain (Myers et al. 1995; Gekakis et al. 1995). A *tim* null deletion mutant (*tim*\(^0\)) was found to be arrhythmic indicating that, *tim* is a major component needed for circadian oscillation (Sehgal et al. 1994). Even though these studies indicated that *tim* and *per* genes are important for rhythmic time keeping, their crucial role and functional aspects were not known until their exact molecular role was discovered (Hardin et al. 1990).

It was observed that PER auto regulates its own mRNA levels concurring with previous observations that mutations in the *per* gene disrupt the cyclic rhythmicity of *per* mRNA synthesis. It was detected that, *per* mRNA and protein levels oscillate in opposite phase to each other, i.e. when PER protein level is at a peak, *per* mRNA level is at a trough and vice versa (Hardin et al. 1990). This result indicated that PER could be an auto repressor, repressing its own expression. This auto regulatory effect was observed with *tim* gene and TIM protein levels as well, hence the inference was made that, TIM and PER oscillation were obviously due to the *tim* and *per* mRNAs oscillations respectively (Sehgal et al. 1995). Promoter dissection studies were carried out to understand the process by which transcription of *tim* and *per* genes is
controlled. These experiments found that, in the per gene promoter ~500 “bp” (basepair) upstream from the coding region and the transcription initiation site, a ~70 bp enhancer sequence was present (Hao et al. 1997). In this enhancer sequence a consensus E-box of 6 bp sequence (CACGTG) was detected to be crucial for the activation of transcription.

Similarly, upstream from the tim gene promoter, an enhancer sequence necessary for tim transcription was found with a consensus E-box element “CACGTG” (McDonald et al. 2001). Later mutational experiments found two genes which played an important role in rhythmicity of the circadian clock, “clock” (clk) and “cycle” (cyc). This new discovery helped to understand the TIM and PER feedback loops even better. The expression of tim and per genes were reduced and arrhythmic in both cyc and clk mutant flies (Allada et al. 1998). Thus it was proposed that, the molecular role of CLK and CYC proteins is to positively regulate the transcription of tim and per genes.

It was seen from the sequence of the circadian proteins CLK and CYC that, they have a PAS domain and a basic helix-loop-helix domain (bHLH). It was known that other proteins belonging to the bHLH TF family are able to bind to E-box sequence. Thus, a mechanism for CLK and CYC transcriptional control was proposed i.e. CLK and CYC proteins can bind to the tim and per gene E-box containing promoters through their bHLH domains resulting in transcriptional activation of per and tim genes. Upon subsequent translation and synthesis, PER and TIM proteins can bind through their PAS domains to CLK and CYC proteins, thereby removing CLK and CYC from E-box resulting in termination of transcription (Darlington et al. 1998).

It was observed that per and tim mRNA levels peak roughly in anti-phase with clk mRNA levels (Darlington et al. 1998). Just after dawn clk mRNA concentrations are at their maximum, but per and tim mRNAs peak roughly before dark. In per01 and tim01 null mutants, the mRNA concentration of the clk gene is very less compared to the wild type. Initially, this result was misinterpreted as PER and TIM being the activators for clk transcription (Bae et al. 1998). Surprisingly, in clkbrk, a type of mutant which produces non-functional CLK, the clk mRNA concentrations were found to be extremely high (Glossop et al. 1999). Hence, it was proposed that CLK might be self-repressing its own expression, which was soon discouraged. Since,
investigation of \(clk\) gene promoter indicated the absence of any E-box sequence needed for bHLH binding (Darlington et al. 1998).

The ambiguity of transcriptional regulation of \(clk\) was increased when in the \(per^{01}\) \(clk^{brk}\) double mutant, it was observed that \(clk\) mRNA transcripts are consecutively high (Glossop et al. 1999). These results indicated a new \(clk\) gene activation component. A new rhythmically expressed circadian protein with a “PAR domain factor” was identified and named “vrille” (VRI). The promoter region of the \(vri\) gene showed a consensus E-box indicating its activation by CLK/CYC. Considerably reduced levels of \(clk\) mRNA concentration were found in VRI over-expression and also, the discovery of a VRI specific consensus sequence (“VRI Box”) in the \(clk\) promoter suggested a direct \(clk\) transcriptional repression by VRI (Glossop et al. 1999).

It was not until 2003 that a new “PAR Domain Protein 1” (PDP1) was identified through its similarity with VRI and their common PAR Domain Factor (Cyran et al. 2003). E-box specific sequence CACGTG was spotted in \(pdp1\) and \(vri\) genes thus, throwing light on their transcriptional control. Similar to \(tim\) and \(per\) genes, CLK/CYC heterodimers regulate the transcription of \(pdp1\) and \(vri\) genes by binding to their respective E-box promoters (Cyran et al. 2003; Glossop et al. 2003). Sequence analysis of PDP1 and VRI proteins showed that, they belong to a family of TFs having a highly conserved basic-leucine zipper DNA binding domains, implying that both PDP1 and VRI should bind to the same gene sequences. Similarly, wet-lab observations implied that VRI and PDP1 can bind to the same region in the \(clk\) gene promoter (subsequently named as V/P box). Thus, PDP1 and VRI, compete and complement each other’s activation and repression of \(clk\) gene transcription by competitive binding to the same V/P box (Cyran et al. 2003).

In brief, excluding the newly discovered known circadian component (CWO), six gene products in total (Table 1-1) were known to be important in circadian regulation of \textit{Drosophila} until 2007. They can be broadly classified into two groups based in their molecular function as transcriptional activators and repressors. Transcriptional activators include PDP, CYC and CLK. The repressors include TIM, PER, VRI and
Table 1-1 Molecular components of *Drosophila* circadian clock.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Type</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLK</td>
<td>bHLH PAS Domain</td>
<td>Activator</td>
</tr>
<tr>
<td>CYC</td>
<td>bHLH PAS Domain</td>
<td>Activator</td>
</tr>
<tr>
<td>PDP1ε</td>
<td>Basic-leucine Zipper</td>
<td>Activator</td>
</tr>
<tr>
<td>PER</td>
<td>PAS Doimain</td>
<td>Repressor</td>
</tr>
<tr>
<td>TIM</td>
<td>PAS Domain</td>
<td>Repressor</td>
</tr>
<tr>
<td>VRI</td>
<td>Basic-leucine Zipper</td>
<td>Repressor</td>
</tr>
<tr>
<td>CWO</td>
<td>bHLH Orange</td>
<td>Repressor</td>
</tr>
</tbody>
</table>
CWO. Since CWO molecular function is the main focus of this study, it is discussed in detail in section 4.4. These circadian clock components form an intricate genetic regulatory network by either repressing or activating other clock genes or even their own.

4.3.1 PER/TIM-CLK/CYC feedback loop

The molecular components involved in this feedback loop are “PERIOD” (per), “TIMELESS” (tim), “CLOCK” (clk), and “CYCLE” (cyc) genes respectively (Hardin et al. 2006, Stanewsky et al. 2002). Products of clk and cyc genes i.e. CLK and CYC proteins, are involved in the transcriptional activation of tim and per genes. The roughly 24 hours molecular cycling of the Drosophila circadian rhythm starts with late evening/early night transcriptional activation of per and tim PAS domain TF genes by the binding of CLK/CYC heterodimers to the E-boxes in tim and per promoters. Subsequently, their protein products i.e. TIM and PER protein concentrations, are at maximum level peaking towards late night (Figure 4-2). During this time, TIM and PER form a PER/TIM heterodimer complex to stabilise each other. The reason for such heterodimerisation is discussed in detail in section 4.3.3. Upon forming the dimer PER/TIM can be trans-located to the nucleus either together or independent of each other (see section 4.3.3 and 4.3.4), following which, PER/TIM interact with CLK/CYC through their PAS domain and remove CLK/CYC from their E-box resulting in transcriptional self-repression. As a consequence of their self-repression, PER and TIM protein levels fall below the threshold needed to accommodate existing CLK/CYC dimers. Hence, the free CLK/CYC complex now reactivates and starts the next cycle of per and tim transcription unhindered during the following day.

4.3.2 VRI/PDP1 feedback loop

The VRI/PDP feedback loop plays an important role in robust clk oscillation (Glossop et al. 2003; Cyran et al. 2003). Similar to the E-box having tim and per, during late evening/early night the CLK/CYC complex induces transcriptional activation of vri and pdpl Basic-leucin zipper TF genes (Figure 4-2). Even though, pdpl and vri gene transcriptions are initiated at the same time, their protein products accumulate in
Figure 4-2 Relative 24 h concentrations of circadian clock components in *Drosophila melanogaster*: (a) mRNA levels and (b) Protein levels.
different time periods. VRI protein starts to accumulate in sync with PER and TIM protein concentrations during middle night, but PDP1 accumulation happens at late night. There is 3-4 hour time delay for PDP1 protein to concentrate in cytoplasm. The reason for such a time lag is not known. Following the cytoplasmic accumulation of VRI, it is transported to the nucleus and binds to a V/P box element in the \textit{clk} promoter sequence, in order to repress \textit{clk} transcription resulting in a fall of CLK protein levels. But VRI encounters a completion for the V/P box when PDP1 concentrations increase and eventual binding of transcriptional activator PDP1 by early morning results in increased levels of \textit{clk} mRNA rest of the day. Interestingly, a PDP1 deletion \textit{Drosophila} mutant, lacking the PDP1 activation loop was able to generate sustained oscillations (Benito et al. 2007). This shows that, there are many discoveries waiting to happen in the circadian system of \textit{Drosophila} to eventually better understand its molecular stochasticity. This result shows that even with increased circadian research in \textit{Drosophila} new clock components and feedback loops are waiting to be discovered.

\subsection*{4.3.3 Post-translational modifications of clock components}

The CLK/CYC aided transcriptional activation is repressed by a negative feedback loop formed by TIM and PER proteins. It was assumed that, since TIM is needed for PER accumulation in cytoplasm, the same might be true for its nuclear localisation (Kloss et al. 1998, 2001). DBT (doubletime), a “casein kinase Iε homolog” in \textit{Drosophila}, aids in rapid degradation of PER protein in cytoplasm by phosphorylation (Kloss et al. 1998). It was observed that upon hyperphosphorylation by DBT, PER protein is primed for degradation due to the binding of F-box protein SLIMB (Chiu et al. 2008). An exact domain in the PER protein where DBT binds was mapped to a narrow region in the C-terminal of the PAS domain about 27 to 54 amino acids long (Nawathean et al. 2007, Kim et al. 2007). Degradation through hyperphosphorylation of PER is prevented by the formation of PER/TIM heterodimers (Vosshall et al. 1994; Fang et al. 2007; Kivimae et al. 2008)). Cytoplasmic accumulation of hypophosphorylated PER protein can be noticed in mutant \textit{dbt} partial loss flies, which lack a functional DBT protein (Rothenfluh et al. 2000). Even in \textit{tim}^{01}\textit{-dbt} double mutants, which lack both TIM and functional DBT proteins, this same result for PER was observed (Cyran et al. 2005; Weber et al. 2003). Both these \textit{dbt} mutant studies
showed that PER nuclear transport did happen resulting in repressed CLK/CYC activation. From these results we can infer that, independent of TIM, hypo-phosphorylated PER can repress CLK/CYC activation by nuclear transport. Apart from the degradation of PER through hyper-phosphorylation, it is also prevented from nuclear entry by DBT (Cyran et al. 2005). It was discovered in an immunohistochemistry experiment that “Protein phosphatase 2a” (PP2a) which belongs to HEAT repeat kinase family helps in dephosphorylating and stabilising cytoplasmic PER from hyper- to hypo-phosphorylated state (Shafer et al. 2002). Aided by the partial dephosphorylation through PP2A, PER protein is translocated to the nucleus (Sathyanarayanan et al. 2004). In addition to PP2a and DBT, another kinase CK2 is known to play a role in PER nuclear accumulation (Weber 2003). As seen before, DBT promotes cytoplasmic but prevents nuclear accumulation of PER meanwhile CK2 supports PER nuclear translocation (Smith et al. 2008). CK2 is known to phosphorylate both PER and TIM in the PER/TIM dimer; this phosphorylation leads to the disassociation of PER and TIM proteins, after which PER nuclear translocation happens (Landskron et al. 2009). It was observed that the inhibition of CLK/CYC transcriptional activation of E-boxes by PER is dependent on CLK phosphorylation by DBT, mediated by PER (Yu et al. 2006). A diagram of various kinases interaction is shown in (Figure 4-3).

This result is consistent with the observation in partial 
*dbt* mutant flies where, PER was able to bind to CLK/CYC, but effective transcriptional repression was absent (Weber et al. 2003; Cyran et al. 2005). A new kinase protein, “Oscillation Protein Phosphatase 1” (PP1) was discovered, similar to PP2A. This new protein was seen to dephosphorylate both TIM and PER (Fang et al. 2007). Even though TIM dephosphorylation by PP1 can happen spontaneously, similar action on PER requires the participation of TIM. In addition, rhythmic accumulation of PER might be dependent on dephosphorylation by PP1 (Fang et al. 2007). While PP2A controls nuclear translocation of PER/TIM, PP1 may be necessary for rhythmic accumulation of PER (Fang et al. 2007). Added to that, “Shaggy” (SGG) a *Drosophila* homolog of mammalian protein “Glycogen Synthase Kinase 3” (GSK3) was found to phosphorylate TIM and affect nuclear transport of TIM and PER proteins (Harms et al. 2001).
Figure 4-3 Network diagram for phosphorylation interaction of PER.
In summary, a handful of kinases are known to interact in phosphorylation of PER and TIM proteins (Figure 4-3), e.g. DBT, PP2A, CK2 and PP1 with PER, and SGG and PP1 with TIM respectively. PER, TIM and the PER/TIM complex are seen to cohabit with these kinase proteins in cytoplasm and nucleus (Kloss et al. 2001). Further studies to understand the exact composition and location of these kinases with respect to the 24 hour oscillation of circadian proteins, will help in better understanding of circadian system in *Drosophila*.

We did not consider the nuclear import or phosphorylation of *Drosophila* clock components in our new models, since there are too many candidates and severe uncertainty in the role of various kinases. Contesting ideas with respect to multiple kinases involved in PER and TIM phosphorylation contrasts with total lack of knowledge on phosphorylation of other clock proteins like VRI, PDP1, CYC and CWO.

**4.3.4 Period homodimer hypothesis**

A photo bleaching study (FRET) was carried out to monitor the activity of PER and TIM proteins in *Drosophila* “S2 cells” (commonly used lab grown *Drosophila* cell lines). The study showed that in cytoplasm, the synthesised TIM and PER proteins were able to form a heterodimeric complex, but during the nuclear translocation stage they entered and accumulated separately (Meyer et al. 2006). This shows that TIM protein is not important for PER nuclear function. The mechanism by which PER performs its repressor function is by either directly altering CLK conformation upon interaction with the CLK/CYC hetrodimer, or indirectly as a transporter of the known phosphorylation kinases DBT into the proximity of CLK. In order to perform either of these functions PER should be translocated to the nucleus (Shafer et al. 2002, Shafer et al. 2004). Even though, PER/TIM heterodimerisation is important for their individual stabilisation, their eventual nuclear translocation can happen independent of each other (Cryan et al. 2005; Meyer et al. 2006).

Since PER protein is vital for rhythmic oscillation of clock proteins, the question arises how, in the absence of TIM, can PER protect itself from DBT hyperphosphorylation? A possible explanation could be the existence of a PER/PER
homodimer which was earlier reported (Dietz et al. 1996), but no biological significance or relevance was attributed to it. PER/PER homodimers were observed in wild-type flies and disrupting the formation of such dimers using single amino acid substitution resulted in reduced rhythmicity of clock components in mutants (Schotland et al. 2000, Yildiz et al. 2005). Yet, the mechanism by which such homodimers affect the clock components was not known. Recently in another study, a mutant PER\textsuperscript{L} protein was synthesised using single amino acid substitutions in the \textit{per} gene. The nature of PER\textsuperscript{L} is such that, it can interact with phosphorylating kinases DBT and CK2 but not with TIM (Landskron et al. 2009). It was observed that in PER\textsuperscript{L} flies homodimerisation was still happening and the resultant dimers were accumulating in the nucleus. But PER\textsuperscript{L} flies showed drastically impaired behavioural and molecular rhythmicity. Furthermore, a mutant PER protein without a complete “αF-helix”, due to the deletion of 512-568 amino acids from its “C-domain” was unable to homodimerise and the interaction between PER and CK2 was lacking (Landskron et al. 2009). Hence it is suggested that, PER/PER homodimer formation is dependent upon its interaction with phosphorylating kinases. A protein crystallographic study of the PER PAS domain was done in which the PAS domains of PER isoform A and isoform B were hybridised and crystalised (King et al. 2011). The resulting tertiary structure showed that the αF-helix of one isoform was able to completely wrap around the αE-helix of another and vice versa inferring the exact nature of PER/PER homodimerisation. We simulated a possible PER homodimer using a COTH dimer prediction server and a representational picture of the output is provided in figure 4-4 and table 1-2. The prey used was a 38 amino acid sequence of the PAS domain in a PER protein isoform A transcript, and the bait was the PAS domain sequence in PER protein isoform B. The predicted result showed a very high confidence of PER forming a homodimer through its PAS domain with the best alignment convergence score of 0.97. For more details about the COTH server the reader is referred to Zhang et al. 2011. Although, based on these observations and results we were tempted to include a PER/PER homodimer in our models, it was discouraged due to three reasons. 1) Even though PER homodimers do exist, their exact molecular role is still in a hypothetical stage. 2) Including a blind hypothesis will lead to considerable increase in model complexity. 3) It is nearly impossible to estimate the parameters without any wet-lab data to validate.
Figure 4-4 COTH output result showing predicted secondary structure of PER PAS domains, and one of the output image is used.
The above table is explained as follows:
(From www.zhanglab.ccmf.med.umich.edu/COTH)

Column 1: The rank of the template as determined by the Z-score. The template is with the highest Z-score is ranked 1.
Column 2: Name of the template PDB from which the particular model was generated. Templates are ranked according to the Z-score of the alignment.
Column 3: The length of the threading alignment between the query sequence and the template.
Column 4: The coverage of the alignment in calculated by dividing the length of the alignment (previous column) by the total length of the query complex.
Column 5: Z-score, or statistical significance, of the alignment upon which the ranking of the templates is based.
Column 6: The sequence identity between the query sequence and the template sequence.
Column 7: The confidence of the threading alignment is based on the Zscore and is interpolated from the COTH benchmarking results. The confidence is considered High if the Zscore of the alignment is greater than 2.5 or Low otherwise.
Column 8: The threading alignment between the query sequence and the template.
Column 9: The models generated from COTH threading alone. These models are generated before the superposition of the MUSTER individual chain templates onto the dimer templates identified by COTH threading.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Template</th>
<th>Align_length</th>
<th>Coverage</th>
<th>Zscore</th>
<th>Seq_id</th>
<th>Confidence</th>
<th>Target-template alignments</th>
<th>Models from Dimeric COTH threading only</th>
<th>Final models by COTH after MUSTER template superposition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1tabf1_1tabf2</td>
<td>72</td>
<td>0.947</td>
<td>5.342</td>
<td>0.083</td>
<td>high</td>
<td>alignment 1</td>
<td>threading 1</td>
<td>model 1</td>
</tr>
<tr>
<td>2</td>
<td>2sf9A1_2sf9A2</td>
<td>69</td>
<td>0.908</td>
<td>4.253</td>
<td>0.130</td>
<td>high</td>
<td>alignment 2</td>
<td>threading 2</td>
<td>model 2</td>
</tr>
<tr>
<td>3</td>
<td>2dx8A0_2dx8B0</td>
<td>75</td>
<td>0.987</td>
<td>4.063</td>
<td>0.107</td>
<td>high</td>
<td>alignment 3</td>
<td>threading 3</td>
<td>model 3</td>
</tr>
<tr>
<td>4</td>
<td>1krla0_1knlb0</td>
<td>75</td>
<td>0.987</td>
<td>3.997</td>
<td>0.133</td>
<td>high</td>
<td>alignment 4</td>
<td>threading 4</td>
<td>model 4</td>
</tr>
<tr>
<td>5</td>
<td>1ufiA0_1ufib0</td>
<td>74</td>
<td>0.974</td>
<td>3.627</td>
<td>0.189</td>
<td>high</td>
<td>alignment 5</td>
<td>threading 5</td>
<td>model 5</td>
</tr>
<tr>
<td>6</td>
<td>2gpea0_2gpeb0</td>
<td>75</td>
<td>0.987</td>
<td>3.565</td>
<td>0.080</td>
<td>high</td>
<td>alignment 6</td>
<td>threading 6</td>
<td>model 6</td>
</tr>
<tr>
<td>7</td>
<td>1e6ff0_1e6ff0</td>
<td>76</td>
<td>1.000</td>
<td>3.464</td>
<td>0.211</td>
<td>high</td>
<td>alignment 7</td>
<td>threading 7</td>
<td>model 7</td>
</tr>
<tr>
<td>8</td>
<td>1hx2f0_1hx2h0</td>
<td>60</td>
<td>0.789</td>
<td>3.339</td>
<td>0.183</td>
<td>high</td>
<td>alignment 8</td>
<td>threading 8</td>
<td>model 8</td>
</tr>
<tr>
<td>9</td>
<td>1hx2b0_1hx2d0</td>
<td>60</td>
<td>0.789</td>
<td>3.273</td>
<td>0.183</td>
<td>high</td>
<td>alignment 9</td>
<td>threading 9</td>
<td>model 9</td>
</tr>
<tr>
<td>10</td>
<td>1hx2d0_1hx2f0</td>
<td>59</td>
<td>0.776</td>
<td>3.254</td>
<td>0.203</td>
<td>high</td>
<td>alignment 10</td>
<td>threading 10</td>
<td>model 10</td>
</tr>
</tbody>
</table>
4.3.5 Cryptochrome repressor hypothesis

The main difference between *Drosophila* and mammalian functional clock components is the hypothesised role of “Cryptochrome” (CRY), a photoreceptor. In “Light-Dark” (LD) cycles of *Drosophila cry*\(^b\) mutants (a strongly hypomorphic mutation) entrainment by light was reduced (Stanewsky et al. 1998). It was known that, light cannot be entrained totally in *cry*\(^b\) mutants which lack optic nerve functions (Helfrich-Foster et al. 2005). Thus, the fundamental function of CRY protein in *Drosophila* was established as the transmitter of light signals (zeitgeber) in order to entrain the daily locomotor activity. In mammals, the CRY protein (mCRY) is considered a core clock protein. By contrast in *Drosophila* the CRY photoreceptor is considered only needed for signal transduction.

Among all the circadian clock exhibiting cells in *Drosophila*, pacemaker neurons contribute only a small portion. “Peripheral clocks” do exist in the remaining non-pacemaker cells like malphigian tubules, antennae, etc. Peripheral clocks in these organs like antennae are proposed to affect their local physiology (Kirshnan et al. 1999). In peripheral clocks of *cry*\(^b\) mutants, oscillation of molecular components ceases to exist (Krishnan et al. 2001; Ivanchenko et al. 2001). This indicates that, in addition to being the main photoreceptor in pacemaker neurons, CRY protein has a primary function in peripheral circadian clocks too. Later investigations of the *cry*\(^b\) eye mutant clock gene showed that, similar to the *per*\(^01\) null mutation, CLK/CYC activated transcription of E-box containing gene promoters is not repressed (Collins 2006). Furthermore, CLK/CYC transcriptional activation was repressed with over-expression of both PER and CRY proteins, whereas CRY could not repress CLK/CYC alone.

It has been reported through mutational analysis during development of *Drosophila* larval brain cells that, high levels of CLK/CYC transcriptional activators were found but due to the repression of CRY and PER proteins, rhythmic expression of E-box having clock genes was not initiated (Collins et al. 2006). Hence in some specific clock neurons, CRY in cooperation with PER seems to display repression of CLK/CYC activation.
In order to test the probable repressor activity of CRY, we analysed the CRY transcript using phosphorylation prediction server NetphosK. It is known from literature that CRY repressor activity is through its interaction with PER, hence we checked for DBT mammalian homolog CKI binding sites in the whole CRY protein and found more than 30 hits for these sites (figure 4-5). Based on this observation it will be interesting to consider the hypothesis of the PER homodimer, or the CRY repressor in an appropriate model without traditional circadian clock loops. Even though bioinformatics analysis of the CRY peptide sequence gives hope for the CRY repressor hypothesis, lack of detailed biological understanding favoured non-inclusion of the CRY repressor in our model.

4.4 New clock component CWO

A new clock component “Clockwork Orange” (CWO), a basic helix-loop-helix transcription factor (TF) coded by an E-box having gene (cwo), has been discovered in Drosophila (Kadener et al. 2007; Lim et al. 2007; Matsumoto et al. 2007; Richier et al. 2008). The protein CWO belongs to the “Hairy-Orange” domain TF family, whose other members are known DNA binding transcriptional repressors, e.g. mammalian transcription factors like HES-1, HES-2, HES-3.

5.4.1 CWO molecular function, the research question.

In pacemaker neurons, CWO protein levels were found to be oscillating in phase with CLK/CYC activated PER and TIM protein with a slightly reduced amplitude (Kadener et al. 2007). Subsequent investigation of the cwo promoter showed the presence of up to 20 CLK/CYC target E-boxes in the first intron of the 5’ region of the cwo gene (Lim et al. 2007). Hence it was proposed that transcription of the cwo gene might be influenced by CLK/CYC and its mRNA levels were found to peak in phase with vri, pdp1, tim and per. In the same year it was showed that in clk^{ck} mutants the cwo mRNA levels were low and in per^{01} mutants the cwo mRNA levels were high. This result is similar to results for other known E-box having clock genes, thus confirming the CLK/CYC transcriptional activation of cwo (Kadener et al. 2007).
Figure 4-5 Screening CRY protein sequence: A Part of the NetPhoseK output image showing phosphorylation sites towards the N terminus of CRY protein, here C-indicates PKC, K(green)-indicates CKII, K(pink)-indicates CKI & X-indicates GSK3.
Analysis of a cwo mutant which synthesises a defective CWO protein in constant darkness (DD) revealed that, the oscillation of CLK/CYC activated clock gene products were greatly perturbed and phenotypes observed were similar to the loss of VRI/PDP1 feedback loop. Compared to the wild-type flies, the mutant flies showed, damped rhythms and lengthened periods of E-box having clock genes (Lim et al. 2007). Thus, CWO was established as an important clock component necessary to maintain clock amplitude.

Since CWO belongs to the basic helix-loop-helix TF family, it was suspected that CWO might bind to the E-boxes of other core clock genes. To investigate this assumption, immunoprecipitation studies were carried out using a reporter tagged CWO protein. This study proved that CWO not only binds to the E-boxes of vri, pdp1, tim and per genes, but it also binds to its own E-boxes (Matsumoto et al. 2007). In addition, transcription assays in Drosophila S2 cells in both LD and DD conditions established that CWO forms a negative feedback loop repressing all E-box having gene transcription as well as self-repressing its own expression (Matsumoto et al. 2007).

A cwoB9 null mutant was created using EMS mutagenesis which produced a defective CWO of 36 amino acids in length compared to 698 amino acid length wild-type CWO protein. Further analysis of cwoB9 mutants in LD and DD conditions showed an approximate 2.5 hour lengthening of activity rhythms. The resultant gene transcript concentration levels showed severely damped rhythms for E-box having pdp1, vri, tim and per genes but remarkably, cwo gene transcripts in cwoB9 mutants were of high level (Richier et al. 2008).

From the initial findings from cell culture experiments it was believed that, CWO might be a universal transcriptional repressor which binds to clock gene E-boxes and represses transcriptional activation initiated by CLK/CYC (Matsumoto et al. 2007; Kadener et al. 2007). But reduced per, tim, vri and pdp1 mRNAs hint that CWO might be a transcriptional activator (Richier et al. 2008).

In summary, these experimental findings indicate an uncertain functional role of CWO protein (Figure 4-6) and thus became our primary focus and research question.
Figure 4-6 Schematic diagram of circadian oscillation of clock components in *Drosophila melanogaster* and red blocked lines denote the CWO ambiguity.
We were inspired by this challenge to decipher a probable molecular function and ending the ambiguity of CWO by putting forward a convincing biological argument based on a detailed mathematical model.

4.4.2 Fathallah-Shaykh model

The only known mathematical model which incorporates the new clock component was published in 2009. Fathallah-Shaykh, Bona and Kadener (2009) proposed a model based on a system of non-linear ordinary differential equations to simulate the *Drosophila* circadian regulatory network using regulatory weights that are represented by single parameters. The purpose of their model is to study the regulatory effects of CWO, and to suggest a resolution for the contradicting effects of CWO on direct target genes. The model predicts that the actions of CWO on the interconnected loops elevate the level of CLK/CYC; and the assumption is that elevated level of CLK/CYC in turn generates the positive transcriptional signals on *per, tim, vri, and pdpl* that outweigh the direct repressive actions of CWO, but the model fails to give any biological explanation for this assumption. Also, if the elevated levels of CWO can reflect on the elevated levels of *per, pdpl, vri* and *tim* mRNA levels, then biologically the same should happen for the *cwo* transcript in a similar fashion, since CLK/CYC elevated levels will give the same positive signal to the *cwo* gene promoter too. The deficiency of their conceptual model can be explained with the mathematical approach as follows.

Fathallah-Shaykh Bona and Kadener (2009) introduced a non-linear, autonomous, first-order system of ordinary differential equations assuming that genes/proteins $j$, \( j \in \{1, \ldots, n\} \) regulate the production of gene/protein $i$, as denoted in the following equation

\[
\frac{dx_i(t)}{dt} = \rho \sum_{j=1}^{n} \lambda_{ji} x_j(t) - \delta x_i(t),
\]

\(1 \leq j \leq n,\)
where $x_i$ is the state vector representing the concentration of molecule $i$ at its site of action, $\lambda_{ji}$ are regulatory weights of real parameters showing the effects of molecule $j$ on production rate of molecule $i$. Hence, any increase or decrease of $\lambda_{ji}$ reflects strength of activation or repression.

The sum of the regulatory influences is modulated by an odd sigmoid function,

\[ g : R \rightarrow R \]

where,

\[ g(u) = \frac{u}{\sqrt{1 + u^2}} = \tanh \left(\ln(u + \sqrt{1 + u^2})\right) \]

The function $g$ together with $\rho_i$ indicates the rate of formation of $i$. Thus the Fathallah-Shaykh model has a direct sigmoid function for activation and repression of a gene, as well as the strength of this activity.

We disagree with this modelling approach since there can be either activation or repression of a gene by a particular transcription factor (in this case CWO). It was not shown in *in vivo* or *in vitro* that the strength of activation or repression of the same TFs in the same gene promoters can differ at different time periods. Further, a sign function was designed to model the opposing transcriptional signals of CWO as an enhancer or a repressor.

\[ \text{sign} \left( \Delta Y_g \right) = \text{sign} \left[ \Delta x_{C/C} \left( t_g \right) \lambda_{C/C,g} - x_{CWO} \left( t_g \right) \lambda_{CWO,g} \right] \]

where, $\lambda_{C/C,g}$ and $\lambda_{CWO,g}$ are regulatory weights governing positive and negative transcriptional signal of CWO proteins and the strength of $\lambda_{C/C,g}$ and $\lambda_{CWO,g}$ is used to replicate the opposing positive and negative signals respectively.
Hence, they used $\Delta x_{C/C} (t_g) \lambda_{C/C,g} > x_{CWO} (t_g) \lambda_{CWO,g}$ to model transcriptional enhancement of CWO on per, tim, vri and pdpl gene promoters and $\Delta x_{C/C} (t_g) \lambda_{C/C,g} < x_{CWO} (t_g) \lambda_{CWO,g}$ for CWO repression on its own cwo gene expression.

We differ from this approach of designing special functions to model opposing transcriptional signals or any other regulatory components. There can be a discrimination in the choice of the E-box, but once a TF is bound, it is a switch type result of either trans-activation (ON) or trans-repression (OFF) of the target gene (Genetta et al. 1994). Experience in analysing in vivo dynamics tells us that there can be both positive and negative signals in the form of cooperativity between different TFs but not “reduced/increased” signals of transcription (Prendergast et al. 1992).

The main difference of our model from the Fathallah-Shaykh model is at the conceptual level and in the choice of detailed modelling.

Because of using differential sign functions, the Fathallah-Shaykh model can replicate the opposing CWO signals. But the model makes no effort in explaining the reason for such effect. The authors argue that by removing the CWO loop they can show the increase in CLK/CYC concentration and thereby increase in per, tim, vri and pdpl mRNA levels. This is entirely obvious due to their model’s architecture, but it has to be noted that in in vivo (Richier et al. 2008), loss of CWO leads to a 2-3 h time delay in unphosphorylated PER and TIM protein accumulation, and this is not reproduced in the Fathallah-Shaykh model.

In our model, we not only replicate the biological observation (Richier et al. 2008) that mentions the ambiguous role of CWO protein. In addition, we take into account the biological assumptions that CWO might play a post-translational role along with PER and TIM in the presence of DBT (Richier et al. 2008; Matsumoto et al. 2011). Our choice of detailed modelling gave us the flexibility to test various mutations in addition to a hypothetical mutant which predicts the opposing role CWO plays in loop regulation.
4.4.3 Remodelling the circadian rhythm with new CWO feedback loop

Following from the previous section, clockwork orange (CWO), an orange family TF having a basic helix-loop-helix domain was recently discovered. The cwo mRNA expression is dependent on transcription activation by CLK/CYC and through a self-repressing feedback loop, it controls its own expression. Available literatures attributes a dual role for the CWO protein as a repressor in some and activator in one another publication, in terms of CWO activity on E-boxes of other core clock genes like pdp1, vri, tim and per (Lim et al. 2007; Kadener et al. 2007; Richier et al. 2008). Thus it is important to remodel the circadian rhythm in Drosophila include the new CWO feedback loop and check the role of CWO.

Due to the importance of the CWO component in this thesis, and its ability to play a variety of roles in clock oscillation, we developed three models. The reason for developing three models is that we wanted to conduct various in silico experiments in order to answer the molecular mechanism by which CWO performs its dual role. We based our hypothesis on available wet-lab inference and bioinformatics analysis of the CWO protein. Thus we have three conceptual models with separate sets of probability-based ODEs. Each model is solved and the parameters estimated, and these are presented in great detail in the following chapters. Before we proceed to model development, it is helpful to know the name for each model for reader’s convenience.

We designated the names for our models as Model A, B, and C. Their properties are as follows

Model A: CWO binds to E-boxes alone.

Model B: CWO binds to E-boxes in the form of a CWD heterodimer with a hypothetical protein HP.

Model C: CWO binds to E-boxes alone, but a new CWPT post translational Complex is introduced (explained in section 6.1.3)
In the following chapter, three mathematical models based on probability based ordinary differential equations (ODEs) are developed to include different CWO activity. In Section 6.1, all three conceptual models are developed based on the available biological data and certain assumptions are used to simplify the models. In Section 6.2, choices of the modelling methods used in previous circadian models are discussed and an appropriate method is picked. In section 6.3, a mathematical model is developed from conceptual models in the form of non-linear ODEs.
Chapter 5: Review of circadian models in *Drosophila melanogaster*

5.1 First clock component PER

The molecular components of the *Drosophila* circadian clock have been very well understood. Initially screening of genes was done to identify the first of many key molecular components in the *Drosophila* clock (Hardin et al. 1990). Endogenous oscillation of the clock provided an ideal reference point to screen for molecules whose biochemical structure might shed light on the nature of this self-sustaining oscillator (Czeisler et al., 1982). The first candidate with oscillatory properties was identified as *per*. The discovery of the *per* gene revealed a negative feedback loop, in which the *per* transcript and PER protein oscillate rhythmically (Baylies et al., 1987; Hardin et al., 1990). It was realised that the concentration peaks of *per* mRNA and PER protein are separated by a 4–6 h gap; subsequently activation of *per* expression declines, suggesting that the PER protein either directly or indirectly inhibits the expression of its own gene.

5.2 Circadian models

Various mathematical models for these endogenous clocks were proposed for different organisms like *Arabidopsis, Neurospora, Drosophila* and *Mammalia* (Goldbeter et al. 1995; Smolen et al. 2001; Leloup et al. 2003, Xie et al. 2007; Locke et al. 2005). Similar properties of all the previous models are that, they include at the very least a single negative or positive feedback loop and had the ability to simulate continuous oscillations for a minimum of one clock component in a feedback loop with desired parameter settings (Goldbeter et al. 2002).

5.2.1 Goldbeter model

The first model of circadian rhythms in *Drosophila melanogaster* reproduced the self-repression of the period (PER) protein on its own gene (*per*) promoter in a feedback loop (Goldbeter et al. 1995). The well-known Goldbeter model used a continuous deterministic approach in a mathematical model of a single negative feedback loop of
the *Drosophila* circadian clock. A schematic diagram of Goldbeter model is shown in figure 5-1. The time delay was shown as a function of PER phosphorylation, which delays entry of mature protein into the nucleus, thereby repressing *per* transcription.

Goldbeter’s model contained five variables to describe the rate limiting phosphorylation steps of PER protein. This model can produce sustained negative feedback loop oscillations of PER protein under given parameter settings. Since the photoreceptor TIM protein was not known then, the entrainment of light was not included in Goldbeter’s model.

**5.3 Discovery of PER/TIM feedback loop**

Another circadian gene expression was observed to be entrained by light and was named timeless (*tim*) whose protein TIM binds with PER (Gekakis et al. 1995). It was later known that the accumulation of the PER/TIM complex in cytoplasm is a necessary prerequisite for the transport of both PER and TIM to the nucleus (Saez et al. 1996), inability to form a heterodimer made the flies arrhythmic. The phase shift of circadian oscillation required degradation of TIM in the presence of light (Hunter-Ensor et al., 1996; Albrecht et al. 1997). TIM’s response to light is mediated by a photoreceptor CRY; it oscillates in phase with light. CRY protein expression peaks during daylight hours (Stanewsky et al., 1998), which leads to degradation of TIM and resetting the circadian clock.

**5.3.1 Leuloup and Goldbeter model**

After the discovery of the PER/TIM complex a new model was built based on the auto regulation of the PER/TIM protein dimer, self-repressing their own expression by *per* and *tim* genes (Leloup et al. 1998; Leloup et al. 2000). This model had a light entrainment compartment by incorporating the input of a light signal with a degradation rate for TIM protein. But the exact mechanism through which light entrained CRY to degrade phosphorylated TIM was not known.
Figure 5-1 Schematic diagram of Goldbeter’s model (1995) showing PER feedback loop.
5.3.2 Tyson Model

Second model incorporating the PER/TIM feedback loop was described for *Drosophila* circadian rhythms by (Tyson et al. 1999) as shown in figure 5-2. It differed from Leuloup and Goldbeter model, by the addition of a PER stabilisation positive feedback loop after the PER dimer is formed. This new model replicated several qualities of observed biology, like expression of the *per*\(^L\) mutant and temperature compensation, generating interlocked feedback loops.

5.4 Discovery of CLK/CYC

A new *Drosophila* circadian clock gene (*clk*), an ortholog of the circadian mammalian clock gene *bmal1* was discovered (Bae et al. 1998). A previously identified 69-bp promoter region (or E-box) upstream of the *per* gene suggested the presence of a transcriptional activator (Hao et al., 1997), and it was identified that the *Drosophila* clock protein (CLK) activates the genes of both PER and TIM in association with another circadian protein CYCLE (CYC) by forming a heterodimer, hence the CLC/CYC role is understood as transcriptional activation.

CLK/CYC activated *per* and *tim* gene promoters and subsequent transcription was seen to be inhibited by the accumulation of the PER/TIM dimer in the nucleus (Darlington et al. 1998). These discoveries complemented the observed effect that PER and TIM do not bind directly to their own DNA (Sasson-Corsi, et al. 1998). Another protein kinase coding gene double-time (*dbt*), a mammalian homolog of Ceasin Kinase 1 (CK1) was soon discovered and it was found to encode for a protein kinase DBT, which positively regulates the accumulation and nuclear transport of PER/TIM through cytoplasmic phosphorylation of the dimer. In a fly with a mutation in the *dbt* gene resulting in loss of functional DBT protein, *per* and *tim* gene transcripts cease to oscillate (Kloss et al., 1998).
Figure 5-2 Schematic diagram of Tyson model (1999) showing PER/TIM and CLK/CYC feedback loop.
5.4.1 Smolen model

Immediately after the discovery of the *clk* gene and subsequent understanding of the function of its product CLK protein (Glossop et al. 1999), a new elaborate model with CLK/CYC and PER/TIM interlocked feedback loops was proposed (Smolen et al. 2001). Smolen’s model was different from other previous models majorly with the inclusion of *clk* mRNA and CLK protein and its ability to produce continuous free running oscillations of these substances. Robustness of circadian rhythms in *Drosophila melanogaster* was attributed to this new CLK component.

5.5 Discovery of VRI/PDP1 feedback loop

In addition to the known interlocked feedback loops (Glossop et al. 1999), it was observed that protein concentrations of CLK oscillate in counter phase to PER and TIM concentrations, suggesting the presence of an unknown feedback loop with a new component either transcriptionally activating or repressing the *clk* gene resulting in CLK oscillating independent of PER and TIM proteins. These predictions were supported by reports of a novel clock-controlled gene vrille (*vri*) whose transcript VRI, a basic-leucine zipper protein, oscillates in the same phase as *per* and *tim* (Blau et al. 1999). Later on the PAR-domain-protein (*pdp1*) a gene, whose product PDP1, another basic-leucine zipper protein, participates as part of feedback loop that positively controls the the *clk* gene expression was identified (Cyran et al. 2003). CLK activates the expression of *vri* and *pdp1* genes. CLK is a core clock component, flies lacking functional CLK protein become arrhythmic and other clock components cease to oscillate.

4.6 Later models with two feedback loops

A revised mathematical model of the *Drosophila* circadian clock by including this additional *clk* feedback loop was proposed (Ueda et al. 2001). The mathematical framework used in Ueda’s model was similar to Goldbeter’s model. This model confirmed the feasibility of interlocked feedback loops, but it was not until 2003 when the molecular function of the VRI protein as a repressor of *clk* gene transcription was confirmed (Cyran et al., 2003).
Following the discovery of the VRI/PDP1 feedback loop, an updated mathematical model of the *Drosophila* clock, that incorporated the negative feedback loop of VRI protein and subsequently the newly discovered positive feedback loop of PDP1 protein in the regulatory network was proposed (Smolen et al. 2004).

This new model (Figure 5-3) contained a network based on interconnected feedback loops of the recently known *vri* and *pdp1* as well as the previously known *per* and *clk* gene transcripts and their products. Importantly, the model replicated the positive feedback loop of PDP1 protein with the biologically observed time lag. A rate-limiting two-step phosphorylation of PER protein, one in the cytoplasm and a second in the nucleus, was added, differing from Goldbeter’s first mathematical model with four rate-limiting PER phosphorylation steps. In contrast to Smolen’s previous model, the role of TIM protein was not included in this model. Simulation showed sustained oscillation of all participating components; in addition, the auto regulatory PER negative feedback loop was found to be important, but not the *pdp1* and *vri* feedback loops, to generate limit cycle oscillations. Even with the non-inclusion of the TIM protein, phase response curves can be generated similar to experimental data and the model can replicate null mutations (mutations resulting in loss of functional gene) observed in wet-lab including *clk, pdp1, vri* and *per* null mutants (viable organisms having a null mutation).

**4.6.1 Ruoff model**

Similar to Smolen’s 2004 model, a new model was proposed in which PDP1 and VRI proteins, positively and negatively regulate the expression of *clk* respectively (Ruoff et al. 2005), and CLK protein in turn activates the transcription of *pdp1* and *vri* (Figure 5-4). A novel characteristic of this model is that, CLK activation of *per* and *tim* gene expressions were combined to form a single function, represented by the PER/TIM complex. This assumption reduces the mathematical complexity without losing the overall biological picture that CLK activation is needed for PER and TIM protein production. In contrast to Smolen’s model where, *pdp1* and *vri* regulatory feedback loops were shown as unnecessary for maintaining robust oscillations, in Ruoff’s model under given parameter settings *pdp1* and *vri* feedback loops were
Figure 5-3 Schematic diagram of Smolen’s model (2004) showing PER, CLK, and VRI/PDP1 feedback loops.
Figure 5-4 Schematic diagram of Ruoff’s model (2005) showing PER/TIM, VRI/PDP1, and CLK feedback loops.
known to be vital for sustained oscillation of the PER/TIM complex and its amplitude. This importance of pdp1 and vri feedback loops regulation is more in sync with observed biology than Smolen’s model. In addition, the per phase shift due to temperature compensation in short \(per^s\) and long \(per^l\) mutants can be replicated, which make this model more valid. But a major drawback is that, due to the robustness of the parameters appropriate outputs for light entrainment cannot be achieved, since the parameters were too stable to induce any observable changes from perturbation.

5.6.2 Xie and Kulasiri Model

A new model using a novel set of continuous, deterministic equations with binding probabilities, was proposed (Xie et al. 2007). Figure 5-5 provides a schematic diagram of the additional feedback loops used in this model by Xie and Kulasiri. In this model two negative loops (PER/TIM and VRI) and one positive loop (PDP1) regulated the network of 14 molecules. Activation of transcription in E-box containing gene promoters by CLK/CYC is repressed by the binding of the PER/TIM (PT) heterodimer complex to CLK/CYC (CC), resulting in the formation of a CLK/CYC-PER/TIM (CCPT) super complex. VRI binds to the \(clk\) promoter (VP-box) to inhibit transcription of the \(clk\) gene. Meanwhile, PDP1 competes with VRI to activate transcription of \(clk\). For the purposes of simplification, phosphorylation and compartmentalisation are ignored in this model, as well as the effects of both light on TIM degradation and the protein DBT on PER/TIM transport.

5.6.3 Highlights of Xie and Kulasiri’s Model

The major aspect of Xie and Kulasiri’s model is the inclusion of all main TFs discovered in wet-lab experiments (CLK, CYC, PDP1, PER, VRI and TIM). As the previous models skipped one component or another, this Drosophila circadian model was able to simulate sustained oscillations of all six proteins and mRNAs known at that time (Xie et al. 2007). This model was based on a \(per/tim\) positive transcriptional feedback loop and a \(vri/pdp1\) regulatory feedback loop similar to previous models (Cryan et al. 2003). Various assumptions and simplifications were used to simplify the model, but important biological observations were not compromised. The two
Figure 5-5 Schematic diagram of Xie and Kulasiri’s model (2007) showing PER/TIM, VRI/PDP1, and CLK/CYC feedback loops.
feedback loops were joined by a CLK/CYC transcriptional activation loop (Figure 5-5). CLK/CYC activator complex binding to E-boxes of vri, pdp1, per and tim genes initiates the two loops. In the first loop, the PER/TIM heterodimer auto regulates its own expression and in the second loop PDP1 and VRI proteins, activate and repress the expression of the clk gene by binding to its V/P box promoter respectively. Like Goldbeter’s model, which produced sustained oscillation of circadian components, this model, which employed very different types of rate equations compared to Goldbeter’s, is also capable of producing sustained 24-h periodic oscillations of the interacting circadian network molecules. Also similarly to other proposed models, this model replicates observed biology by responding to changes in light cycles with phase-adjustments (or entrainment) in the periodic rhythms. The results of various simulated mutations in the per gene were also consistent with experimental observations (Xie and Kulasiri et al. 2007). In conclusion, this model validates the idea that models with very different mathematical equations can replicate the same observed biological behaviours of an intracellular regulatory network (Murray et al. 2002).

Hence in the current research, probability-based ODEs governed by mass action kinetics are used as employed by Xie and Kulasiri (2007). In addition various other assumptions and modelling techniques used in the previous models are taken forward with valid in vitro/in vivo reasoning.
Chapter 6: Developing three new *Drosophila* circadian rhythm models using CWO.

6.1 Development of conceptual models

It is necessary to note that our models are conceptualised at the microscopic cellular level and the components such as proteins and mRNAs are assumed to be well mixed and homogenous in the system. And the system volume unit is \( nL \) (nanolitre) and quantity unit is \( nM \) (nanomol). The relative level of clock components in our models is the average concentration level in the whole *Drosophila* brain and not at individual neuronal level. The following arguments and assumptions are the same for all three models A, B and C until we reach sections 6.1.2 and 6.1.3 where some more assumptions and molecular observations are made for conceptual model B and C respectively.

6.1.1 Conceptualising model A

As discussed in the previous chapters, a new conceptual model as shown in Figure 6-1 involving core clock components in pacemaker neurons was designed for model A. The basic model is identical to the previous molecular networks shown (Cyran et al. 2003; Xie et al. 2007; Fathallah-Shaykh et al. 2009). This new model has *cwo*, *vri/pdp1* and *per/tim* loops, in total three feedback loops. As shown in Figure 6-1, all the three loops are initiated by the CLK/CYC transcriptional activator. In the *per/tim* feedback loop, CLK/CYC activates the *per* and *tim* E-box having promoters and the subsequent TIM and PER proteins for a heterodimer. This heterodimer binds to CLK/CYC and represses its own expression. Similarly in the next loop, CLK/CYC initiates the E-box having pdp1 and vri gene promoters and among the resultant gene products, PDP1 and VRI bind to the consensus V/P box in the *clk* gene promoter enabling the activation and repression of *clk* transcription respectively. In the third loop, transcription of the E-box having *cwo* gene is initiated by CLK/CYC activators and its resultant gene product CWO represses all five core clock components *per*, *tim*, *vri*, *pdp1* and *cwo* by binding to their E-boxes.
Figure 6-1 Conceptual model A: In the above network diagram of conceptual model A, blocked red lines show the repressor activity and green arrows show the activator component. Dotted arrows indicate transcription and translation processes. Where, rectangles with concave points denote gene names and round objects denote proteins and protein complexes respectively.
The following assumptions were used while developing the conceptual model in order to simplify and concentrate on the primary research question.

1. Even though, nuclear and cytoplasmic localisation of core clock components do occur, the same was not taken in to consideration in our models. Rather, we ignored the same in order to avoid compartmentalisation. Reason for this approach is that in order to include separate compartments we need to include the kinases which play a role in priming the protein for nuclear entry. But, due to lack of a concrete argument in favour of exact roles kinases play to participate in nuclear transport, it was decided not to include the roles of kinases. Also previous theoretical studies (Kurosawa et al. 2002; Xie et al. 2007; Fathallah-Shaykh et al. 2009) showed that, even without separate cellular compartments, models with appropriate feedback loops can produce sustained rhythmic oscillations.

2. Following from the previous assumption, phosphorylation and dephosphorylation by kinases like DBT, CK2, PP1, etc. are not considered at a detailed level. Notwithstanding the fact that phosphorylation is an important rate limiting step which governs molecular collisions and timing, it is impossible to estimate parameters for the stochasticity involved in such interactions. Yet, the degradation rate of clock proteins in our model can be considered depicting the phosphorylation of PER and TIM proteins by DBT and PP1 respectively, since, hyperphosphorylation by these kinases leads to degradation.

3. With regard to the number of E-boxes, six E-boxes were found while sequencing a 4 kb region in the pdp1 gene promoter upstream of the transcription initiation site. (Cyran et al. 2003). Similarly, up to four functional E-boxes in the vri promoter, and up to five in the per and tim promoters were discovered (Blau et al. 1999; McDonald et al. 2001; Cyran et al. 2003). Also, up to 20 CLK/CYC target E-boxes were found in the first intron 5’ end of the cwo gene (Lim et al. 2007). But we ignored the E-box numbers in this model since, TFs and their E-box binding behaviour are still under investigation. In
order to model the competitive binding of CWO and CLK/CYC at E-boxes, we assumed that CLK/CYC and CWO bind to an E-box separately at any given time, since even if one E-box is bound by CLK/CYC the signal is positive for transcription and if CWO is bound it is a negative transcriptional signal. Due to the detailed probability functions the number of E-boxes was chosen as 1 in all the promoters.

4. The next assumption is that, if CLK/CYC dimers are binding to E-boxes then they are free to initiate transcription without the repression from PER/TIM heterodimers. Activation through CLK/CYC, cannot be repressed by PER/TIM when, the former is already bound to an E-box (Yu et al. 2006). However, direct inhibition takes place in mammals where, the transcriptional activation by CLK/BMAL1 the mammalian analogue of *Drosophila* CLK/CYC can be directly repressed by photoreceptor cryptochrome (CRY) even while CLK/BMAL1 is being bound to the target E-boxes and there is no need for the activation complex to move away from the E-box (Lee et al. 2003). We chose this simple approach in our model since in *Drosophila*, there are various models which propose a direct and indirect inhibition of transcription by PER/TIM dimers which are contradictory to each other.

5. Data from wet-lab findings of pacemaker neurons elucidated that, compared to other clock proteins CYC activator concentrations are high repeatedly (Glossop et al. 1999). In previous models, CYC concentrations were fixed at 1 nM to maintain its constant high levels (Xie et al. 2007). The same assumption is used in these models (A, B and C) as well.

6.1.2 Conceptualising model B

The main reason for this conceptual model is the nature of the basic helix-loop-helix domain (bHLH) present in CWO protein. It is known that in these proteins near the HLH sequence a 15 amino acid basic region is present. The primary role of the HLH domain is to aid in formation of homo or hetero protein dimmers. Such dimerisation is
vital for DNA binding repressor activity, since two basic regions are needed for repression. Other proteins related to CWO with a hairy-orange domain have a WRPW motif at their C-terminal, through which they interact with similar repressor protein to function as transcriptional repressors e.g. GROUCHO. But analysis of the CWO peptide sequence shows that CWO lacks the WRPW motif. Hence, it is possible that CWO does not need any cofactors to repress transcription or maybe some yet to be discovered domains do exist in CWO protein through which it can interact with lesser known cofactors.

Based on the assumption made in the wet-lab literatures about probable role for CWO functioning as dimer (Matsumoto et al. 2011), a new conceptual model B was developed as shown in (figure 6-2). It was assumed that CWO protein might have a molecular function as a heterodimer (Matsumoto et al. 2011). In Drosophila, other basic helix-loop-helix orange domain proteins are known to play a transcriptional repressor role by binding to E-boxes after forming a heterodimer (Davis et al. 2001). Based on the findings in molecular studies, CWO was proposed to form a heterodimer with potential proteins like Mγ, SIDE and Mβ. In addition CWO was proposed to probably homodimerise in order to function as a transcriptional repressor (Matsumoto et al. 2011).

Even though the above assumptions are valid, there is less evidence towards the CWO homodimer, hence we decided to assume a CWO heterodimer (CWD) with a hypothetical protein (HP). Note that this hypothetical protein could be considered as SIDE, Mβ or Mγ as postulated in Matsumoto et al. 2011.

The core framework of this new model B is similar to model A with the only exception being the addition of another component CWD that is formed by dimerisation of CWO and HP. Instead of CWO binding to E-boxes in model B, CWD binds to E-boxes in all CLK target genes, thereby competing with CLK/CLK for E-box space.
Figure 6-2 Conceptual model B: In the above network diagram of conceptual model B, blocked red lines show the repressor activity and green arrows show the activator component. Dotted arrows indicate transcription and translation processes. Where, rectangles with concave points denote gene names and round objects denote proteins and protein complexes respectively.
6.1.3 Conceptualising model C

It has been observed by all three research papers related to CWO discovery (Lim et al. 2007; Kadener et al. 2007; Richier et al. 2008) that CWO can act as both an activator and repressor in an unknown fashion. The main objective of this research is to satisfy our curiosity in explaining the ambiguity of CWO protein. Towards this effect we went ahead to develop this new model.

Consider a large lateral ventral neuron in cwoB9 mutant flies. Here the null mutant is not arrhythmic and the mutation can be observed by detecting reduced mRNA concentrations of E-box genes. The proposed reason for this effect is agreed as loss of CWO activation of transcription on E-boxes having genes (Richier et al. 2008). Even though CWO activation by binding to E-box is tested *in vivo*, it is entirely possible that CWO acts as a pure transcriptional repressor true to its nature of belonging to a Myc-type, basic helix-loop-helix domain family (bHLH). But how can an E-box-binding transcriptional repressor play an activator role? In order to answer this anomaly we made an assumption that CWO acts as an activator indirectly by destabilising other transcriptional repressors rhythmically expressed in phase with CWO.

Since PER and TIM protein oscillations are in phase with CWO we were curious to check any post translational interactions that can take place between CWO, PER and/or TIM. Phosphorylation is a major component of post-translational modifications. Phosphorylation and de-phosphorylation are discussed in detail in section 4.3.5. As shown in section 4.3.5, the DBT binding sites play a major role in functional PER protein increase during dark. Hence we checked for DBT phosphorylation sites if any in CWO sequence using GPS 2.0 (“Group-based Prediction System”) (Xue et al. 2008) software (Table 1-3). Phosphorylation sites and their associated kinase candidate prediction using this bioinformatics software was immensely helpful for us in developing this conceptual model. It is well known that GPS 2.0 is one of the first known phosphorylation site prediction computational tools.

GPS 2.0 was implemented in Java. It is able to scan and predict kinase-specific phosphorylation sites in relevant rankings and has an inventory of 408 mammalian
protein kinases in its database. As stand-alone software, GPS 2.0 has the ability to predict on a large scale with high fidelity of > 13000 phosphorylation sites. Hence, GPS 2.0 is an excellent tool to test for CWO phosphorylation sites, and to construct our assumption of the post-translational component. For further reading about the prediction tool and its features please refer to (Xue et al. 2008). Since the GPS library only has mammalian kinases, we can use CK1 the mammalian homolog of DBT as prediction choice. GPS2.0 predicted in total more than 25 binding sites for CK1 in the *Drosophila* CWO protein transcript (Table 1-3). This positive result was encouraging.

We then used NetPhosK, a server based phosphorylation-site prediction application. NetPhosK can predict probable phosphorylation sites at tyrosine, threonine and serine residues in peptide sequences, and it computes this using a neural network-based method (Blom et al. 2004). Earlier versions of NetPhosK were developed between 1997 and 1998 but were not released publicly. The recent version 2.0 was released after training the algorithms for a large number of established phosphorylation sites.

Currently, NetPhosK covers the following kinases CKII, Cam-II, PKA, PKG, PKC, Cdc2, etc. CKI and GSK3 binding sites can also be predicted using NetPhosK, hence it was decided to compare the GPS 2.0 results with NetPhosK (Figure 6-3). The results were identical and both tools confirmed the existence of many potential phosphorylation sites. Though it is a fact that these sites are present, without knowledge of the territory structure of CWO it is difficult to determine the exact available phosphorylation sites that will be used. However, based on these data, we proceeded to include a post-translational interaction component (CWPT) involving CWO, PER and TIM in model C (figure 6-4). The core framework of model C is similar to models A and B. The phosphorylation-interaction component between CWO, PER and TIM was added, and it is shown from figure 6-4 that, like in model A, CWO protein competes with CLK/CYC in binding to E-boxes in *per, tim, cwo, pdp1 and vri* promoters.
<table>
<thead>
<tr>
<th>Position</th>
<th>Code</th>
<th>Kinase</th>
<th>Peptide</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>S</td>
<td>CK1</td>
<td>STSATAYSEDDAEYA</td>
<td>2.143</td>
</tr>
<tr>
<td>448</td>
<td>S</td>
<td>CK1</td>
<td>ELHDEETSEHCPSV</td>
<td>1.623</td>
</tr>
<tr>
<td>514</td>
<td>S</td>
<td>CK1</td>
<td>SSNAGASSANRLDKP</td>
<td>1.766</td>
</tr>
<tr>
<td>190</td>
<td>S</td>
<td>CMGC</td>
<td>NVSASSGSPHQAYHP</td>
<td>5.802</td>
</tr>
<tr>
<td>258</td>
<td>S</td>
<td>CMGC</td>
<td>YAVANGSPASNAVG</td>
<td>6.574</td>
</tr>
<tr>
<td>563</td>
<td>T</td>
<td>CMGC</td>
<td>AAGGQLSTPSSTTAP</td>
<td>6.867</td>
</tr>
<tr>
<td>634</td>
<td>S</td>
<td>CMGC</td>
<td>NVNFMPPSPPASLLA</td>
<td>9.52</td>
</tr>
<tr>
<td>34</td>
<td>S</td>
<td>CK1/CK1</td>
<td>STSATAYSEDDAEYA</td>
<td>2.274</td>
</tr>
<tr>
<td>45</td>
<td>S</td>
<td>CK1/CK1</td>
<td>STSATAYSEDDAEYA</td>
<td>2.744</td>
</tr>
<tr>
<td>362</td>
<td>S</td>
<td>CK1/CK1</td>
<td>HHHHTDSHHDFFES</td>
<td>1.973</td>
</tr>
<tr>
<td>258</td>
<td>S</td>
<td>CMGC/GSK</td>
<td>YAVANGSPASNAVG</td>
<td>4.951</td>
</tr>
<tr>
<td>383</td>
<td>S</td>
<td>CMGC/GSK</td>
<td>TDTSNMHSPPRDLL</td>
<td>4.123</td>
</tr>
<tr>
<td>563</td>
<td>T</td>
<td>CMGC/GSK</td>
<td>AAGGQLSTPSSTTAP</td>
<td>5.889</td>
</tr>
<tr>
<td>565</td>
<td>S</td>
<td>CMGC/GSK</td>
<td>GGLSTPSSTAPLP</td>
<td>4.568</td>
</tr>
<tr>
<td>567</td>
<td>T</td>
<td>CMGC/GSK</td>
<td>QLSTPSSTAPLPR</td>
<td>3.667</td>
</tr>
<tr>
<td>633</td>
<td>S</td>
<td>CMGC/GSK</td>
<td>NVNFMPPSSPASLLA</td>
<td>4.074</td>
</tr>
<tr>
<td>634</td>
<td>S</td>
<td>CMGC/GSK</td>
<td>NVNFMPPSSPASLLA</td>
<td>6.79</td>
</tr>
<tr>
<td>42</td>
<td>T</td>
<td>CK1/CK1/CK1a</td>
<td>LNFSTSATAYSEDDA</td>
<td>2.059</td>
</tr>
<tr>
<td>45</td>
<td>S</td>
<td>CK1/CK1/CK1a</td>
<td>STSATAYSEDDAEYA</td>
<td>1.824</td>
</tr>
<tr>
<td>211</td>
<td>S</td>
<td>CK1/CK1/CK1a</td>
<td>DMLATASDVEHSDQ</td>
<td>1.588</td>
</tr>
<tr>
<td>294</td>
<td>S</td>
<td>CK1/CK1/CK1a</td>
<td>VPSNSTGSGAAACA</td>
<td>3.824</td>
</tr>
<tr>
<td>310</td>
<td>S</td>
<td>CK1/CK1/CK1a</td>
<td>GNSNSGSNSNNSAS</td>
<td>2.706</td>
</tr>
<tr>
<td>469</td>
<td>S</td>
<td>CK1/CK1/CK1a</td>
<td>HSHLQALSEHSKDG</td>
<td>2.176</td>
</tr>
<tr>
<td>514</td>
<td>S</td>
<td>CK1/CK1/CK1a</td>
<td>SSNAGASSANRLDKP</td>
<td>2.471</td>
</tr>
<tr>
<td>697</td>
<td>S</td>
<td>CK1/CK1/CK1a</td>
<td>LEQAMNQSW******</td>
<td>2.824</td>
</tr>
<tr>
<td>185</td>
<td>S</td>
<td>CK1/CK1/CK1d</td>
<td>CHMPDNVSAASSGPH</td>
<td>5.333</td>
</tr>
<tr>
<td>406</td>
<td>S</td>
<td>CK1/CK1/CK1d</td>
<td>HSHHTQDSLMVSRR</td>
<td>5.111</td>
</tr>
<tr>
<td>39</td>
<td>T</td>
<td>CK1/CK1/CK1e</td>
<td>ESSLNFSTSATAYSE</td>
<td>1.143</td>
</tr>
<tr>
<td>45</td>
<td>S</td>
<td>CK1/CK1/CK1e</td>
<td>STSATAYSEDDAEYA</td>
<td>3.143</td>
</tr>
<tr>
<td>86</td>
<td>S</td>
<td>CK1/CK1/CK1e</td>
<td>NSCLADLSRLPPQY</td>
<td>2</td>
</tr>
<tr>
<td>190</td>
<td>S</td>
<td>CK1/CK1/CK1e</td>
<td>NVSASSGSPHQAYHP</td>
<td>1.5</td>
</tr>
<tr>
<td>209</td>
<td>S</td>
<td>CK1/CK1/CK1e</td>
<td>LRDMLATASDVEHSDV</td>
<td>1.357</td>
</tr>
<tr>
<td>237</td>
<td>S</td>
<td>CK1/CK1/CK1e</td>
<td>HLNQLQRSQQAAAAA</td>
<td>1.571</td>
</tr>
<tr>
<td>277</td>
<td>T</td>
<td>CK1/CK1/CK1e</td>
<td>PLTNGGGTGAPPA</td>
<td>1.286</td>
</tr>
<tr>
<td>333</td>
<td>T</td>
<td>CK1/CK1/CK1e</td>
<td>GSCPAPKTVPLAAHQ</td>
<td>1.643</td>
</tr>
<tr>
<td>349</td>
<td>S</td>
<td>CK1/CK1/CK1e</td>
<td>HQAPVITSTAPHH</td>
<td>1.714</td>
</tr>
<tr>
<td>350</td>
<td>T</td>
<td>CK1/CK1/CK1e</td>
<td>QAPVITSTAPHH</td>
<td>1.357</td>
</tr>
<tr>
<td>383</td>
<td>S</td>
<td>CMGC/GSK/GSK3A</td>
<td>TDTSNMHSPPRDLL</td>
<td>6.667</td>
</tr>
<tr>
<td>563</td>
<td>T</td>
<td>CMGC/GSK/GSK3A</td>
<td>AAGGQLSTPSSTTAP</td>
<td>4.333</td>
</tr>
<tr>
<td>258</td>
<td>S</td>
<td>CMGC/GSK/GSK3B</td>
<td>YAVANGSPASNAVG</td>
<td>6</td>
</tr>
<tr>
<td>333</td>
<td>T</td>
<td>CMGC/GSK/GSK3B</td>
<td>GSCPAPKTVPLAAHQ</td>
<td>4.49</td>
</tr>
<tr>
<td>383</td>
<td>S</td>
<td>CMGC/GSK/GSK3B</td>
<td>TDTSNMHSPPRDLL</td>
<td>5.061</td>
</tr>
<tr>
<td>563</td>
<td>T</td>
<td>CMGC/GSK/GSK3B</td>
<td>AAGGQLSTPSSTTAP</td>
<td>5.776</td>
</tr>
<tr>
<td>634</td>
<td>S</td>
<td>CMGC/GSK/GSK3B</td>
<td>NVNFMPPSSPASLLA</td>
<td>5.898</td>
</tr>
</tbody>
</table>

Table 1-3 GPS2.0 predicted phosphorylation sites in CWO protein.
Figure 6-3 A part of the NetPhoseK output image showing phosphorylation sites between amino acid number 301-480 of CWO protein, here C-indicates PKC, K(green)-indicates CKII, K(pink)-indicates CKI & X-indicates GSK3.
Figure 6-4 Conceptual model C: In the above network diagram of conceptual model B, blocked red lines show the repressor activity and green arrows show the activator component. Dotted arrows indicate transcription and translation processes. Where, rectangles with concave points denote gene names and round objects denote proteins and protein complexes respectively.
6.2 Modelling transcriptional regulation

There are seven known rhythmic TFs in the molecular clock of *Drosophila* circadian rhythm with the inclusion of CWO, as discussed in detail in chapter 3. The CLK/CYC heterodimers, by binding to the consensus E-box promoters, initiate the transcription of *cwo*, *pdp1*, *vri*, *tim* and *per* genes. Competition between activator PDP1 and VRI repressor to bind at the consensus V/P box in the *clk* promoter leads to the rhythmic expression of the *clk* gene. At the same time PER/TIM dimers disable the CLK/CYC activators thereby repressing the transcription of *cwo*, *pdp1*, *vri*, *per* and *tim* genes. In addition CWO, CWD and CWPT bind to E-boxes in *cwo*, *vri*, *pdp1*, *per* and *tim* promoters competing with CLK/CYC in model A, B and C respectively.

Two common methods are used to model the transcriptional control. In the first method, as used in Fathallah-Shaykh’s CWO model, the rate of transcription of a gene is simulated using Hill functions or in terms of a decreasing or increasing function. The majority of the previous *Drosophila* circadian clock models use this method. In the second method, both forward and reverse reactions are used to explicitly model the binding and unbinding of TFs (Vilar et al. 2002; Forger et al. 2003). Like Hill functions, this method of detailed modelling is also used widely in both mammalian and circadian clock models (Xie et al. 2007).

6.2.1 The requirement for detailed modelling with explicit reactions

The transcriptional regulation of the circadian system in *Drosophila*, including activation and repression loops, was modelled by using Hill type or special functions in most of the well-known earlier models (Leloup et al. 1998; Smolen et al. 2001). At appropriate parameter settings using such functions, these models were able to replicate molecular oscillations of clock proteins (Ueda et al. 2001). In particular, to model the CLK/CYC activation of E-box having genes and PER/TIM repression of CLK/CYC, Hill functions were used with the assumption that TFs bind/unbind to E-boxes in promoters in a fast manner. Hill functions use is limited to oscillations of proteins with no time delay. Wet-lab results show that upon activation of *pdp1* and *vri* transcription by CLK/CYC, there exists a 3-4 hour time lag between the
accumulations of their respective mRNAs. Similarly, in the CWO mutants there is a 2-3 hour time delay in accumulation of proteins. Even though, the exact mechanism by which such time lags happen is not known, it is very important to simulate them in order to propose a more biologically relevant model. These observations cannot be reproduced using Hill functions.

In the absence of knowledge about participating components of such time delays, known models using equations representing detailed binding and unbinding processes have been shown to simulate observed time lags (Vilar et al. 2002; Forger et al. 2003; Xie et al. 2007). In addition, Hill functions cannot depict the competition between different transcription factors able to bind to the same regulatory sequence. In molecular regulation of *Drosophila* circadian rhythm such competitions do exist e.g. VRI and PDP1 competition to bind with the *clk* V/P box (Cyran et al. 2003) and; CWO and CLK/CYC competition to bind with E-box having promoters (Matsumoto et al. 2007).

In a previous model, the following equation represents the VRI and PDP1 regulation of *clk* gene expression (Smolen et al. 2004).

\[
R_{clk} = V_{clk} \left( \frac{[P DPI]^2}{[P DPI]^2 + K_{POC}^2} \right) \left( \frac{K_{V/P}^2}{[VRI]^2 + K_{V/P}^2} \right) + R_{bas}
\]  

(3.9)

where,

- \( R_{clk} \) is the transient rate of transcription in the *clk* gene,

- \( V_{clk} \) is the \( V_{max} \) of *clk* transcription rate, and

- \( R_{bas} \) is the transcription rate of the *clk* gene without any TFs binding to its consensus V/P box (Basal transcription rate).
A Hill function is used to reproduce PDP1 activation of the \( \textit{clk} \) gene, in which, \( K_{PDC} \) is the association constant for \( \textit{clk} \) promoter and PDP1 activator binding. Similarly, another Hill function is used to represent VRI repression of the \( \textit{clk} \) gene promoter, in which \( K_{VC} \) is the association constant for the V/P box regulatory sequence in \( \textit{clk} \) gene promoter and VRI binding. In this regulation equation, increase of VRI will be taken care by decrease in PDP1 and vice versa, like a switch. This behaviour was supposed to show the PDP1 and VRI competition. But, in vitro results show that, such competition between competing TFs can happen simultaneously during \( \textit{clk} \) gene transcription (Cyran et al. 2003).

### 6.3 Modelling transcriptional activation and repression

It is considered that, transcriptional rates of a gene in both activated and deactivated states, multiplied with the sum of probabilities of activated and deactivated promoters will give the efficient transcriptional rate for that particular gene in that particular state. Mathematical representation of transcriptional control this way is biologically more meaningful than using special functions to act as switch. The probability of an activated or deactivated promoter can be calculated, since promoters do not stay in both activated and deactivated states at the same time.

It has been discussed before that, \( \textit{cwo, tim, pdp1, per} \) and \( \textit{vri} \) genes have many E-boxes in their promoters (see section 6.1). With respect to E-box numbers, we made a crucial hypothesise that the transcriptional repressor CWO competes with the transcriptional activator CLK/CYC in order to bind to any E-box in gene promoters. Thus an assumption is made that just one E-box bound can trigger activation or repression and then we proceeded to derive the kinetic equations for the new model based on binding probabilities which are as follows.

The competition between CWO and CLK/CYC to bind or unbind from an E-box can be represented as shown in the following reactions:

\[
B + A \xrightarrow{b_{BA}} BA
\]
\[ B + C \xrightarrow{b_{cw}} BC \]

where \( B \) is the binding site in the gene promoter (E-box)

\( A \) is the activator CLK/CYC

\( BA \) is the CLK/CYC bound to the E-box

\( b_{cc} \) is the rate of CLK/CYC binding to the E-box

\( ub_{cc} \) is the rate of CLK/CYC unbinding from E-box

\( C \) is the repressor CWO

\( BC \) is the CWO bound to the E-box

\( b_{cw} \) is the rate of CWO binding to the E-box

\( ub_{cw} \) is the rate of CWO unbinding from E-box

Using mass action kinetics, the above reactions can be represented as the following ODEs:

\[
\frac{d[BA]}{dt} = [B][A]b_{cc} - [BA]ub_{cc} \tag{4.0}
\]

\[
\frac{d[BC]}{dt} = [B][C]b_{cw} - [BC]ub_{cw} \tag{4.1}
\]

If ‘\( V \)’ represents the total volume (moles) of the cell, the number (molecules) of substrates i.e. \( BA \), \( BC \) and \( B \) can be expressed as \([BA]V\), \([BC]V\) and \([B]V\) respectively. And if ‘\( n \)’ represents the total number of E-boxes (binding sites), we get,

\[
[B]V + [BA]V + [BC]V = n \tag{4.2}
\]

\[
[B]V = n - [BA]V - [BC]V \tag{4.3}
\]
\[ [B] = \left( \frac{n}{V} \right) - [BA] - [BC] \]  \hspace{1cm} (4.4)

Substituting the value of \([B]\) in Eqn. (4.0) and (4.1) we get,

\[ \frac{d[BA]}{dt} = \left( \left( \frac{n}{V} \right) - [BA] - [BC] \right) [A]b_{ce} - [BA]ub_{ce} \]  \hspace{1cm} (4.5)

\[ \frac{d[BC]}{dt} = \left( \left( \frac{n}{V} \right) - [BA] - [BC] \right) [C]b_{cw} - [BC]ub_{cw} \]  \hspace{1cm} (4.6)

If \(Pr_{ba}\) and \(Pr_{bc}\) is the probability of \(A\) binding to \(B\) and \(C\) binding to \(B\) respectively, then

\[ Pr_{ba} = \frac{[BA]V}{n} \Rightarrow [BA] = \left( \frac{n}{V} \right) Pr_{ba} \quad \text{Similarly} \quad [BC] = \left( \frac{n}{V} \right) Pr_{bc} \]

Substituting the values of \([BA]\) and \([BC]\) in Eqn. (4.5) and (4.6) we get

\[ \frac{d\left( \frac{n}{V} \right) Pr_{ba}}{dt} = \left( \left( \frac{n}{V} \right) - \left( \frac{n}{V} \right) Pr_{ba} - \left( \frac{n}{V} \right) Pr_{bc} \right) [A]b_{ce} - \left( \left( \frac{n}{V} \right) Pr_{ba} \right) ub_{ce} \]  \hspace{1cm} (4.7)

That simplifies to,

\[ \frac{dPr_{ba}}{dt} = (1 - Pr_{ba} - Pr_{bc}) [A]b_{ce} - Pr_{ba} ub_{ce} \]  \hspace{1cm} (4.8)
Similarly,

\[
\frac{d}{dt} \left( \frac{n}{V} \right) \Pr_{bc} = \left( \frac{n}{V} \Pr_{na} - \left( \frac{n}{V} \right) \Pr_{bc} \right) [C]b_{cw} - \left( \frac{n}{V} \right) \Pr_{bc} ub_{cw} \]  

(4.9)

and this simplifies to,

\[
\frac{d \Pr_{bc}}{dt} = (1 - \Pr_{na} - \Pr_{bc}) [C]b_{cw} - \Pr_{bc} ub_{cw} 
\]

(5.0)

Assuming that the E-box is bound by the activator CLK/CYC, the transcription rate of the gene is \( tC_{av} \). If CWO repressor is bound, the rate is \( tC_{dc} \) and if neither is bound the transcription occurs at the basal rate of \( tC_{dv} \).

The probability of no E-boxes being bound will be

\[
(1 - \Pr_{na} - \Pr_{bc})^n 
\]

(5.1)

Whereas the probability of an E-box being bound is

\[
\left[ 1 - (1 - \Pr_{na} - \Pr_{bc})^n \right] 
\]

(5.2)

The probability of only CLK/CYC being bound is

\[
\left[ \frac{(A)}{(A) + (C)} \right] \left[ 1 - (1 - \Pr_{na} - \Pr_{bc})^n \right] 
\]

(5.3)
Similarly, the probability of CWO being bound will be

\[
\left[ \frac{C}{A + C} \right] \left[ 1 - \left( 1 - \Pr_{ba} - \Pr_{dc} \right)^n \right]
\]

(5.4)

Consequently the transcription rate will be

\[
tC_{av} \left[ \frac{A}{A + C} \left[ 1 - \left( 1 - \Pr_{ba} - \Pr_{dc} \right)^n \right] \right] + tC_{dv} \left[ \frac{C}{A + C} \left[ 1 - \left( 1 - \Pr_{ba} - \Pr_{dc} \right)^n \right] \right] + tC_{dv} \left( 1 - \Pr_{ba} - \Pr_{dc} \right)^n
\]

(5.5)

and that simplifies to

\[
\left[ \frac{tC_{av} (A) + tC_{dv} (C)}{A + C} \right] \left[ 1 - \left( 1 - \Pr_{ba} - \Pr_{dc} \right)^n \right] + tC_{dv} \left( 1 - \Pr_{ba} - \Pr_{dc} \right)^n
\]

(5.6)

Similar equations describe the following,

1. Competition between VRI and PDP1 to a binding site in clk promoter in model A, B and C.
2. Competition between CWD and CWPT with CLK/CYC in binding to E-box of CLK target genes cwo, vri, tim, per and pdp1, also share the same probability rates.

6.3.1 Kinetic equations explained

A new mathematical model for *Drosophila* circadian rhythm was created using the conceptual models as shown in Figures 6-1, 6-2 and 6-4. Using general enzyme kinetics, time evolution of the circadian system with its participating components can be expressed by a framework of kinetic equations in a deterministic model, provided the initial concentrations of the components are known.
In the new models, effective transcription rates are used to determine the rate constants of transcriptional control as shown in Eqn 5.8. In contrast to the previous models where Michaelis-Menten kinetics was used (Ueda et al. 2001; Smolen et al. 2004)), in our model the mass action law is used to govern the reaction rates. This decision was taken to increase the simplicity of the model and reduce the parameter numbers that needed to be estimated. Furthermore, it is not known if the transcription process described using Michaelis-Menten kinetics is the right choice for the circadian system. It was known that, robustness of a model is enhanced by using Michaelis-Menten kinetics while simulating oscillation of a single negative feedback loop (Kurosawa et al. 2002). In another detailed circadian model using mass action kinetics, the estimated parameters were more sensitive to variations (Xie et al. 2007). This sensitivity of parameters is very important while validating the model and conduction various in silico experiments, thus a decision was made to use the mass action rate law in the current models.

An example of an ODE in our model system describing time evolution of per mRNAs and PER protein from Model A is:

\[
\frac{d(per_m)}{dt} = \left\{ \left[ a_{43}(CC) + a_{49}(CWO) \right](CC) + (CWO) \right\} \left[ 1 - \left( 1 - Pr_{5a} - Pr_{5b} \right)^{\alpha_p} \right] + a_{56}(1 - Pr_{5a} - Pr_{5b})^{\gamma_p}\right\}\times per_a - \left( a_{63} \times per_m \right) \tag{5.7}
\]

It was derived from rate equation 5.6 and the equations expressing time evolution of proteins are governed by mass action kinetics:

\[
\frac{d(PER)}{dt} = \left( a_{57} \times per_m \right) - \left( a_{63} \times PER \times TIM \right) + \left( a_{64} \times PT \right) - \left( a_{65} \times PER \right) \tag{5.8}
\]

The above rate equation is used in our model for the time evolution of PER protein in model A. The first part of the equation shows the rate of translation of per mRNA, followed by the rate of association of PT (PER/TIM) dimers and disassociation rate. The last term in the equation denotes the degradation part of time evolved PER.
6.3.2 Variable names and parameters explained

In the derived ODEs, rate constants and variable names are presented in subscript, mixed and normal fonts for better differentiation. Upper case letters were used for proteins and dimer complex names, while lower case letters with the subscript ‘m’ where used to denote the name of mRNAs.

Special abbreviations used for species names are: PT for PER/TIM dimer complex, CCPT for CLK/CYC/PER/TIM super complex, CC for CLK/CYC dimer complex, CWPT for CWO/PER/TIM interaction complex and PDP for PDP1. All other proteins are denoted by their three letter biological nomenclature. The biological description of the parameters is elaborated in Table 1-4, 1-5 and 1-6 for models A, B and C respectively.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Biochemical meaning for model A</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1</td>
<td>Transient rate of binding of CLK/CYC to an E-box promoter in per gene</td>
</tr>
<tr>
<td>a2</td>
<td>Transient rate of binding of CLK/CYC to an E-box promoter in tim gene</td>
</tr>
<tr>
<td>a3</td>
<td>Transient rate of binding of CLK/CYC to an E-box promoter in pdp1 gene</td>
</tr>
<tr>
<td>a4</td>
<td>Transient rate of binding of CLK/CYC to an E-box promoter in vri gene</td>
</tr>
<tr>
<td>a5</td>
<td>Transient rate of binding of CLK/CYC to an E-box promoter in cwo gene</td>
</tr>
<tr>
<td>a6</td>
<td>Transient rate of binding of PDP1 to the V/P box promoter in clk gene</td>
</tr>
<tr>
<td>a7</td>
<td>Transient rate of binding of VRI to the V/P box promoter in clk gene</td>
</tr>
<tr>
<td>a8</td>
<td>Transient rate of binding of CWO to an E-box promoter in cwo gene</td>
</tr>
<tr>
<td>a9</td>
<td>Transient rate of binding of CWO to an E-box promoter in per gene</td>
</tr>
<tr>
<td>a10</td>
<td>Transient rate of binding of CWO to an E-box promoter in tim gene</td>
</tr>
<tr>
<td>a11</td>
<td>Transient rate of binding of CWO to an E-box promoter in pdp1 gene</td>
</tr>
<tr>
<td>a12</td>
<td>Transient rate of binding of CWO to an E-box promoter in vri gene</td>
</tr>
<tr>
<td>a13</td>
<td>Transient rate of CLK/CYC unbinding from an E-box promoter in per gene</td>
</tr>
<tr>
<td>a14</td>
<td>Transient rate of CLK/CYC unbinding from an E-box promoter in tim gene</td>
</tr>
<tr>
<td>a15</td>
<td>Transient rate of CLK/CYC unbinding from an E-box promoter in pdp1 gene</td>
</tr>
<tr>
<td>a16</td>
<td>Transient rate of CLK/CYC unbinding from an E-box promoter in vri gene</td>
</tr>
<tr>
<td>a17</td>
<td>Transient rate of CLK/CYC unbinding from an E-box promoter in cwo gene</td>
</tr>
<tr>
<td>a18</td>
<td>Transient rate of PDP1 unbinding from the V/P box promoter in clk gene</td>
</tr>
<tr>
<td>a19</td>
<td>Transient rate of VRI unbinding from the V/P box promoter in clk gene</td>
</tr>
<tr>
<td>a20</td>
<td>Transient rate of unbinding of CWO from an E-box promoter in cwo gene</td>
</tr>
<tr>
<td>a21</td>
<td>Transient rate of unbinding of CWO from an E-box promoter in per gene</td>
</tr>
<tr>
<td>a22</td>
<td>Transient rate of unbinding of CWO from an E-box promoter in tim gene</td>
</tr>
<tr>
<td>a23</td>
<td>Transient rate of unbinding of CWO from an E-box promoter in pdp1 gene</td>
</tr>
<tr>
<td>a24</td>
<td>Transient rate of unbinding of CWO from an E-box promoter in vri gene</td>
</tr>
<tr>
<td>a25</td>
<td>Transient rate of association of CLK/CYC dimer complex</td>
</tr>
<tr>
<td>a26</td>
<td>Transient rate of association of CLK/CYC/PER/TIM complex</td>
</tr>
<tr>
<td>a27</td>
<td>Transient rate of dissociation of CLK/CYC dimer complex</td>
</tr>
<tr>
<td>a28</td>
<td>Transient rate of dissociation of CLK/CYC/PER/TIM complex</td>
</tr>
<tr>
<td>a29</td>
<td>Transient rate of CLK/CYC bound per gene transcription</td>
</tr>
<tr>
<td>a30</td>
<td>Transient rate of CLK/CYC bound tim gene transcription</td>
</tr>
<tr>
<td>a31</td>
<td>Transient rate of CLK/CYC bound pdp1 gene transcription</td>
</tr>
<tr>
<td>a32</td>
<td>Transient rate of CLK/CYC bound vri gene transcription</td>
</tr>
<tr>
<td>a33</td>
<td>Transient rate of PDP1 bound clk gene transcription</td>
</tr>
<tr>
<td>a34</td>
<td>Transient rate of CLK/CYC bound cwo gene transcription</td>
</tr>
<tr>
<td>a35</td>
<td>Transient rate of CLK/CYC bound pdp1 gene transcription</td>
</tr>
<tr>
<td>a36</td>
<td>Transient rate of CWO bound per gene transcription</td>
</tr>
<tr>
<td>a37</td>
<td>Transient rate of CWO bound tim gene transcription</td>
</tr>
<tr>
<td>a38</td>
<td>Transient rate of CWO bound pdp1 gene transcription</td>
</tr>
<tr>
<td>a39</td>
<td>Transient rate of CWO bound vri gene transcription</td>
</tr>
</tbody>
</table>
Table 1-4 cont. Biological description of the kinetic parameters in model A.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Biochemical meaning for model A</th>
</tr>
</thead>
<tbody>
<tr>
<td>a54</td>
<td>Transient rate of CWO bound cwo gene transcription</td>
</tr>
<tr>
<td>a55</td>
<td>Basal transcription rate of deactivated clk gene free of VRI and PDP1</td>
</tr>
<tr>
<td>a56</td>
<td>Basal transcription rate of deactivated cwo, vri, per, pdp and tim genes.</td>
</tr>
<tr>
<td>a57</td>
<td>Transient rate of translation of synthesised per mRNA</td>
</tr>
<tr>
<td>a58</td>
<td>Transient rate of translation of synthesised tim mRNA</td>
</tr>
<tr>
<td>a59</td>
<td>Transient rate of translation of synthesised pdp1 mRNA</td>
</tr>
<tr>
<td>a60</td>
<td>Transient rate of translation of synthesised vri mRNA</td>
</tr>
<tr>
<td>a61</td>
<td>Transient rate of translation of synthesised clk mRNA</td>
</tr>
<tr>
<td>a62</td>
<td>Transient rate of translation of synthesised cwo mRNA</td>
</tr>
<tr>
<td>a63</td>
<td>Transient rate of degradation of synthesised per mRNA</td>
</tr>
<tr>
<td>a64</td>
<td>Transient rate of degradation of synthesised tim mRNA</td>
</tr>
<tr>
<td>a65</td>
<td>Transient rate of degradation of synthesised pdp1 mRNA</td>
</tr>
<tr>
<td>a66</td>
<td>Transient rate of degradation of synthesised vri mRNA</td>
</tr>
<tr>
<td>a67</td>
<td>Transient rate of degradation of synthesised clk mRNA</td>
</tr>
<tr>
<td>a68</td>
<td>Transient rate of degradation of synthesised cwo mRNA</td>
</tr>
<tr>
<td>a69</td>
<td>Transient rate of degradation of synthesised PER protein</td>
</tr>
<tr>
<td>a72</td>
<td>Transient rate of degradation of synthesised VRI protein</td>
</tr>
<tr>
<td>a73</td>
<td>Transient rate of degradation of synthesised CLK protein</td>
</tr>
<tr>
<td>a74</td>
<td>Transient rate of degradation of synthesised CWO protein</td>
</tr>
<tr>
<td>a76</td>
<td>Transient rate of degradation of produced CLK/CYC dimer complex.</td>
</tr>
<tr>
<td>a79</td>
<td>Fixed E-box number considered in per and tim promoters</td>
</tr>
<tr>
<td>a80</td>
<td>Fixed E-box number considered in pdp1 promoter</td>
</tr>
<tr>
<td>a81</td>
<td>Fixed E-box number considered in vri promoter</td>
</tr>
<tr>
<td>a82</td>
<td>Fixed E-box number considered in cwo promoter</td>
</tr>
<tr>
<td>a85</td>
<td>Transient rate of degradation of synthesised PER protein</td>
</tr>
<tr>
<td>a86</td>
<td>Transient rate of degradation of synthesised TIM protein</td>
</tr>
<tr>
<td>a94</td>
<td>Transient rate of association of produced PER/TIM dimer complex</td>
</tr>
<tr>
<td>a95</td>
<td>Transient rate of disassociation of produced PER/TIM dimer complex</td>
</tr>
<tr>
<td>a103</td>
<td>Transient rate of degradation of produced PER/TIM dimer complex</td>
</tr>
<tr>
<td>Parameter</td>
<td>Biochemical meaning for model B</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>a1</td>
<td>Transient rate of binding of CLK/CYC to an E-box promoter in per gene</td>
</tr>
<tr>
<td>a2</td>
<td>Transient rate of binding of CLK/CYC to an E-box promoter in tim gene</td>
</tr>
<tr>
<td>a3</td>
<td>Transient rate of binding of CLK/CYC to an E-box promoter in pdp1 gene</td>
</tr>
<tr>
<td>a4</td>
<td>Transient rate of binding of CLK/CYC to an E-box promoter in vri gene</td>
</tr>
<tr>
<td>a5</td>
<td>Transient rate of binding of PDP1 to the V/P box promoter in clk gene</td>
</tr>
<tr>
<td>a6</td>
<td>Transient rate of binding of VRI to the V/P box promoter in clk gene</td>
</tr>
<tr>
<td>a7</td>
<td>Transient rate of binding of CWD to an E-box promoter in pdp1 gene</td>
</tr>
<tr>
<td>a8</td>
<td>Transient rate of binding of CWD to an E-box promoter in tim gene</td>
</tr>
<tr>
<td>a9</td>
<td>Transient rate of binding of CWD to an E-box promoter in pdp1 gene</td>
</tr>
<tr>
<td>a10</td>
<td>Transient rate of binding of CWD to an E-box promoter in vri gene</td>
</tr>
<tr>
<td>a11</td>
<td>Transient rate of binding of CWD to an E-box promoter in vri gene</td>
</tr>
<tr>
<td>a18</td>
<td>Transient rate of association of CLK/CYC dimer complex</td>
</tr>
<tr>
<td>a19</td>
<td>Transient rate of association of CLK/CYC/PER/TIM complex</td>
</tr>
<tr>
<td>a20</td>
<td>Transient rate of dissociation of CLK/CYC dimer complex</td>
</tr>
<tr>
<td>a21</td>
<td>Transient rate of dissociation of CLK/CYC/PER/TIM complex</td>
</tr>
<tr>
<td>a35</td>
<td>Transient rate of CLK/CYC bound per gene transcription</td>
</tr>
<tr>
<td>a44</td>
<td>Transient rate of CLK/CYC bound tim gene transcription</td>
</tr>
<tr>
<td>a45</td>
<td>Transient rate of CLK/CYC bound pdp1 gene transcription</td>
</tr>
<tr>
<td>a46</td>
<td>Transient rate of CLK/CYC bound vri gene transcription</td>
</tr>
<tr>
<td>a47</td>
<td>Transient rate of PDP1 bound clk gene transcription</td>
</tr>
<tr>
<td>a48</td>
<td>Transient rate of CLK/CYC bound cwo gene transcription</td>
</tr>
<tr>
<td>a49</td>
<td>Transient rate of CWD bound per gene transcription</td>
</tr>
<tr>
<td>a50</td>
<td>Transient rate of CWD bound tim gene transcription</td>
</tr>
<tr>
<td>a51</td>
<td>Transient rate of CWD bound pdp1 gene transcription</td>
</tr>
<tr>
<td>a52</td>
<td>Transient rate of CWD bound vri gene transcription</td>
</tr>
<tr>
<td>Parameter</td>
<td>Biochemical meaning for model B</td>
</tr>
<tr>
<td>-----------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>a54</td>
<td>Transient rate of CWD bound cwo gene transcription</td>
</tr>
<tr>
<td>a55</td>
<td>Basal transcription rate of deactivated clk gene free of VRI and PDP1</td>
</tr>
<tr>
<td>a56</td>
<td>Basal transcription rate of deactivated cwo, vri, per, pdp and tim genes.</td>
</tr>
<tr>
<td>a57</td>
<td>Transient rate of translation of synthesised per mRNA</td>
</tr>
<tr>
<td>a58</td>
<td>Transient rate of translation of synthesised tim mRNA</td>
</tr>
<tr>
<td>a59</td>
<td>Transient rate of translation of synthesised pdp1 mRNA</td>
</tr>
<tr>
<td>a60</td>
<td>Transient rate of translation of synthesised vri mRNA</td>
</tr>
<tr>
<td>a61</td>
<td>Transient rate of translation of synthesised clk mRNA</td>
</tr>
<tr>
<td>a62</td>
<td>Transient rate of translation of synthesised cwo mRNA</td>
</tr>
<tr>
<td>a63</td>
<td>Transient rate of degradation of synthesised per mRNA</td>
</tr>
<tr>
<td>a64</td>
<td>Transient rate of degradation of synthesised tim mRNA</td>
</tr>
<tr>
<td>a65</td>
<td>Transient rate of degradation of synthesised pdp1 mRNA</td>
</tr>
<tr>
<td>a66</td>
<td>Transient rate of degradation of synthesised vri mRNA</td>
</tr>
<tr>
<td>a67</td>
<td>Transient rate of degradation of synthesised clk mRNA</td>
</tr>
<tr>
<td>a68</td>
<td>Transient rate of degradation of synthesised cwo mRNA</td>
</tr>
<tr>
<td>a69</td>
<td>Transient rate of degradation of synthesised PER protein</td>
</tr>
<tr>
<td>a72</td>
<td>Transient rate of degradation of synthesised VRI protein</td>
</tr>
<tr>
<td>a73</td>
<td>Transient rate of degradation of synthesised CLK protein</td>
</tr>
<tr>
<td>a74</td>
<td>Transient rate of degradation of synthesised CWO protein</td>
</tr>
<tr>
<td>a76</td>
<td>Transient rate of degradation of produced CLK/CYC dimer complex.</td>
</tr>
<tr>
<td>a79</td>
<td>Fixed E-box number considered in per and tim promoters</td>
</tr>
<tr>
<td>a80</td>
<td>Fixed E-box number considered in pdp1 promoter</td>
</tr>
<tr>
<td>a81</td>
<td>Fixed E-box number considered in vri promoter</td>
</tr>
<tr>
<td>a82</td>
<td>Fixed E-box number considered in cwo promoter</td>
</tr>
<tr>
<td>a85</td>
<td>Transient rate of degradation of synthesised PER protein</td>
</tr>
<tr>
<td>a86</td>
<td>Transient rate of degradation of synthesised TIM protein</td>
</tr>
<tr>
<td>a94</td>
<td>Transient rate of association of produced PER/TIM dimer complex</td>
</tr>
<tr>
<td>a95</td>
<td>Transient rate of disassociation of produced PER/TIM dimer complex</td>
</tr>
<tr>
<td>a103</td>
<td>Transient rate of degradation of produced PER/TIM dimer complex</td>
</tr>
<tr>
<td>a37</td>
<td>Transient rate of association of CWD dimer complex</td>
</tr>
<tr>
<td>a41</td>
<td>Transient rate of disassociation of CWD dimer complex</td>
</tr>
<tr>
<td>a77</td>
<td>Transient rate of degradation of CWD dimer complex</td>
</tr>
</tbody>
</table>
Table 1-6 Biological description of the kinetic parameters in model C.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Biochemical meaning for model C</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1</td>
<td>Transient rate of binding of CLK/CYC to an E-box promoter in per gene</td>
</tr>
<tr>
<td>a2</td>
<td>Transient rate of binding of CLK/CYC to an E-box promoter in tim gene</td>
</tr>
<tr>
<td>a3</td>
<td>Transient rate of binding of CLK/CYC to an E-box promoter in pdp1 gene</td>
</tr>
<tr>
<td>a4</td>
<td>Transient rate of binding of CLK/CYC to an E-box promoter in vri gene</td>
</tr>
<tr>
<td>a5</td>
<td>Transient rate of binding of CLK/CYC to an E-box promoter in cwo gene</td>
</tr>
<tr>
<td>a6</td>
<td>Transient rate of binding of PDP1 to the V/P box promoter in clk gene</td>
</tr>
<tr>
<td>a7</td>
<td>Transient rate of binding of VRI to the V/P box promoter in clk gene</td>
</tr>
<tr>
<td>a8</td>
<td>Transient rate of binding of CWO to an E-box promoter in cwo gene</td>
</tr>
<tr>
<td>a9</td>
<td>Transient rate of binding of CWO to an E-box promoter in per gene</td>
</tr>
<tr>
<td>a10</td>
<td>Transient rate of binding of CWO to an E-box promoter in tim gene</td>
</tr>
<tr>
<td>a11</td>
<td>Transient rate of binding of CWO to an E-box promoter in pdp1 gene</td>
</tr>
<tr>
<td>a12</td>
<td>Transient rate of binding of CWO to an E-box promoter in vri gene</td>
</tr>
<tr>
<td>a13</td>
<td>Transient rate of CLK/CYC unbinding from an E-box promoter in per gene</td>
</tr>
<tr>
<td>a14</td>
<td>Transient rate of CLK/CYC unbinding from an E-box promoter in tim gene</td>
</tr>
<tr>
<td>a15</td>
<td>Transient rate of CLK/CYC unbinding from an E-box promoter in pdp1 gene</td>
</tr>
<tr>
<td>a16</td>
<td>Transient rate of CLK/CYC unbinding from an E-box promoter in vri gene</td>
</tr>
<tr>
<td>a17</td>
<td>Transient rate of CLK/CYC unbinding from an E-box promoter in cwo gene</td>
</tr>
<tr>
<td>a18</td>
<td>Transient rate of PDP1 unbinding from the V/P box promoter in clk gene</td>
</tr>
<tr>
<td>a19</td>
<td>Transient rate of VRI unbinding from the V/P box promoter in clk gene</td>
</tr>
<tr>
<td>a20</td>
<td>Transient rate of unbinding of CWO from an E-box promoter in cwo gene</td>
</tr>
<tr>
<td>a21</td>
<td>Transient rate of unbinding of CWO from an E-box promoter in per gene</td>
</tr>
<tr>
<td>a22</td>
<td>Transient rate of unbinding of CWO from an E-box promoter in tim gene</td>
</tr>
<tr>
<td>a23</td>
<td>Transient rate of unbinding of CWO from an E-box promoter in pdp1 gene</td>
</tr>
<tr>
<td>a24</td>
<td>Transient rate of unbinding of CWO from an E-box promoter in vri gene</td>
</tr>
<tr>
<td>a25</td>
<td>Transient rate of association of CLK/CYC dimer complex</td>
</tr>
<tr>
<td>a26</td>
<td>Transient rate of association of CLK/CYC/PER/TIM complex</td>
</tr>
<tr>
<td>a27</td>
<td>Transient rate of dissociation of CLK/CYC dimer complex</td>
</tr>
<tr>
<td>a28</td>
<td>Transient rate of dissociation of CLK/CYC/PER/TIM complex</td>
</tr>
<tr>
<td>a29</td>
<td>Transient rate of CLK/CYC bound per gene transcription</td>
</tr>
<tr>
<td>a30</td>
<td>Transient rate of CLK/CYC bound tim gene transcription</td>
</tr>
<tr>
<td>a31</td>
<td>Transient rate of CLK/CYC bound pdp1 gene transcription</td>
</tr>
<tr>
<td>a32</td>
<td>Transient rate of CLK/CYC bound vri gene transcription</td>
</tr>
<tr>
<td>a33</td>
<td>Transient rate of PDP1 bound clk gene transcription</td>
</tr>
<tr>
<td>a34</td>
<td>Transient rate of CLK/CYC bound cwo gene transcription</td>
</tr>
<tr>
<td>a35</td>
<td>Transient rate of CWO bound per gene transcription</td>
</tr>
<tr>
<td>a36</td>
<td>Transient rate of CWO bound tim gene transcription</td>
</tr>
<tr>
<td>a37</td>
<td>Transient rate of CWO bound pdp1 gene transcription</td>
</tr>
<tr>
<td>a38</td>
<td>Transient rate of CWO bound vri gene transcription</td>
</tr>
<tr>
<td>Parameter</td>
<td>Biochemical meaning for model C</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>a54</td>
<td>Transient rate of CWO bound cwo gene transcription</td>
</tr>
<tr>
<td>a55</td>
<td>Basal transcription rate of deactivated clk gene free of VRI and PDP1</td>
</tr>
<tr>
<td>a56</td>
<td>Basal transcription rate of deactivated cwo, vri, per, pdp and tim genes.</td>
</tr>
<tr>
<td>a57</td>
<td>Transient rate of translation of synthesised per mRNA</td>
</tr>
<tr>
<td>a58</td>
<td>Transient rate of translation of synthesised tim mRNA</td>
</tr>
<tr>
<td>a59</td>
<td>Transient rate of translation of synthesised pdp1 mRNA</td>
</tr>
<tr>
<td>a60</td>
<td>Transient rate of translation of synthesised vri mRNA</td>
</tr>
<tr>
<td>a61</td>
<td>Transient rate of translation of synthesised clk mRNA</td>
</tr>
<tr>
<td>a62</td>
<td>Transient rate of translation of synthesised cwo mRNA</td>
</tr>
<tr>
<td>a63</td>
<td>Transient rate of degradation of synthesised per mRNA</td>
</tr>
<tr>
<td>a64</td>
<td>Transient rate of degradation of synthesised tim mRNA</td>
</tr>
<tr>
<td>a65</td>
<td>Transient rate of degradation of synthesised pdp1 mRNA</td>
</tr>
<tr>
<td>a66</td>
<td>Transient rate of degradation of synthesised vri mRNA</td>
</tr>
<tr>
<td>a67</td>
<td>Transient rate of degradation of synthesised clk mRNA</td>
</tr>
<tr>
<td>a68</td>
<td>Transient rate of degradation of synthesised cwo mRNA</td>
</tr>
<tr>
<td>a69</td>
<td>Transient rate of degradation of synthesised PER protein</td>
</tr>
<tr>
<td>a72</td>
<td>Transient rate of degradation of synthesised VRI protein</td>
</tr>
<tr>
<td>a73</td>
<td>Transient rate of degradation of synthesised CLK protein</td>
</tr>
<tr>
<td>a74</td>
<td>Transient rate of degradation of synthesised CWO protein</td>
</tr>
<tr>
<td>a76</td>
<td>Transient rate of degradation of produced CLK/CYC dimer complex.</td>
</tr>
<tr>
<td>a79</td>
<td>Fixed E-box number considered in per and tim promoters</td>
</tr>
<tr>
<td>a80</td>
<td>Fixed E-box number considered in pdp1 promoter</td>
</tr>
<tr>
<td>a81</td>
<td>Fixed E-box number considered in vri promoter</td>
</tr>
<tr>
<td>a82</td>
<td>Fixed E-box number considered in cwo promoter</td>
</tr>
<tr>
<td>a85</td>
<td>Transient rate of degradation of synthesised PER protein</td>
</tr>
<tr>
<td>a86</td>
<td>Transient rate of degradation of synthesised TIM protein</td>
</tr>
<tr>
<td>a94</td>
<td>Transient rate of association of produced PER/TIM dimer complex</td>
</tr>
<tr>
<td>a95</td>
<td>Transient rate of disassociation of produced PER/TIM dimer complex</td>
</tr>
<tr>
<td>a103</td>
<td>Transient rate of degradation of produced PER/TIM dimer complex</td>
</tr>
<tr>
<td>a104</td>
<td>Transient rate of association of CWPT interaction complex</td>
</tr>
<tr>
<td>a105</td>
<td>Transient rate of disassociation of CWPT interaction complex</td>
</tr>
<tr>
<td>a106</td>
<td>Transient rate of degradation of CWPT interaction complex</td>
</tr>
</tbody>
</table>
6.3.3 Binding probabilities explained:

The following symbols denote probabilities of CLK/CYC (CC) transcriptional activator binding to an E box in:

- \( Pr_{c_{per}} \)
- \( Pr_{c_{ct}} \)
- \( Pr_{c_{cv}} \)
- \( Pr_{c_{pdp}} \)
- \( Pr_{c_{cwo}} \)

The following symbols denote binding probabilities of CWO or CWD or CWPT binding to an E box in:

- \( Pr_{c_{cuper}} \)
- \( Pr_{c_{cwt}} \)
- \( Pr_{c_{cwy}} \)
- \( Pr_{c_{cwpdp}} \)
- \( Pr_{c_{cwo}} \)

The following symbols denote binding probabilities of:

- VRI repressor binding to the V/P box in \( clk \) promoter is denoted by \( Pr_{c_{v}} \)
- PDP1 activator binding to the V/P box in \( clk \) promoter is denoted by \( Pr_{c_{p}} \)
- Binding Probability of DBT binding to PER is denoted by \( Pr_{c_{pd}} \)
- Binding Probability of PP2A binding to PER is denoted by \( Pr_{c_{pp}} \)
In the following equations:

- CC, CWO, CWD, PER, CCPT, CLK, TIM, PDP, VRI, CWPT, clk, pdp, per, tim, vri, cwo are initial conditions (concentrations) which are yet to be determined.
- CYC and HP are constant values which are yet to be determined.
- \( \text{Pr}_{\text{cper}}, \text{Pr}_{\text{cv}}, \text{Pr}_{\text{ct}}, \text{Pr}_{\text{ccwo}}, \text{Pr}_{\text{cwper}}, \text{Pr}_{\text{cwt}}, \text{Pr}_{\text{cwv}}, \text{Pr}_{\text{cwpdp}}, \text{Pr}_{\text{cwo}} \) are probabilities.

These notations will be same for sets of equations in all three models.

### 6.3.4 Models described by system of ODEs

The following ODEs are derived for model A. Refer to APPENDIX A and APPENDIX B for systems of ODEs derived for model B and model C respectively.

i) Binding Probabilities of TFs binding to regulatory elements in gene promoters

\[
\frac{d(\text{Pr}_{\text{cper}})}{dt} = (1 - \text{Pr}_{\text{cper}} - \text{Pr}_{\text{cwper}}) \times a_1 \times CC - \text{Pr}_{\text{cper}} \times a_{18}
\] (5.9)

\[
\frac{d(\text{Pr}_{\text{ct}})}{dt} = (1 - \text{Pr}_{\text{ct}} - \text{Pr}_{\text{cwt}}) \times a_2 \times CC - \text{Pr}_{\text{ct}} \times a_{19}
\] (6.0)

\[
\frac{d(\text{Pr}_{\text{cpdp}})}{dt} = (1 - \text{Pr}_{\text{cpdp}} - \text{Pr}_{\text{cwpdp}}) \times a_3 \times CC - \text{Pr}_{\text{cpdp}} \times a_{20}
\] (6.1)

\[
\frac{d(\text{Pr}_{\text{cv}})}{dt} = (1 - \text{Pr}_{\text{cv}} - \text{Pr}_{\text{cwv}}) \times a_4 \times CC - \text{Pr}_{\text{cv}} \times a_{21}
\] (6.2)

\[
\frac{d(\text{Pr}_{\text{ccwo}})}{dt} = (1 - \text{Pr}_{\text{ccwo}} - \text{Pr}_{\text{cwo}}) \times a_5 \times CC - \text{Pr}_{\text{ccwo}} \times a_{22}
\] (6.3)

\[
\frac{d(\text{Pr}_{\text{owo}})}{dt} = (1 - \text{Pr}_{\text{owo}} - \text{Pr}_{\text{cwo}}) \times a_6 \times CWO - \text{Pr}_{\text{owo}} \times a_{25}
\] (6.4)

\[
\frac{d(\text{Pr}_{\text{cwper}})}{dt} = (1 - \text{Pr}_{\text{cwper}} - \text{Pr}_{\text{cwpdp}}) \times a_9 \times CWO - \text{Pr}_{\text{cwper}} \times a_{26}
\] (6.5)

\[
\frac{d(\text{Pr}_{\text{cwt}})}{dt} = (1 - \text{Pr}_{\text{cwt}} - \text{Pr}_{\text{cwv}}) \times a_{10} \times CWO - \text{Pr}_{\text{cwt}} \times a_{27}
\] (6.6)
\[
\frac{d}{dt}(Pr_{cwpdp}) = (1 - Pr_{cwpdp} - Pr_{cwpdp}) \times a_{11} \times CWO - Pr_{cwpdp} \times a_{28} \\
(6.7)
\]

\[
\frac{d}{dt}(Pr_{cwv}) = (1 - Pr_{cwv} - Pr_{cwv}) \times a_{12} \times CWO - Pr_{cwv} \times a_{29} \\
(6.8)
\]

\[
\frac{d}{dt}(Pr_{vc}) = (1 - Pr_{vc} - Pr_{vc}) \times a_{17} \times VRI - Pr_{vc} \times a_{24} \\
(6.9)
\]

\[
\frac{d}{dt}(Pr_{pc}) = (1 - Pr_{pc} - Pr_{pc}) \times a_{16} \times PDP - Pr_{pc} \times a_{23} \\
(7.0)
\]

The above equations were derived based on Eqn 5.0.

ii) Time evolution of cwo, tim, vri, per, pdp1, and clk mRNA's

\[
\frac{d}{dt}(per_m) = \left\{ \left[ \left( \frac{a_{44}(CC) + a_{45}(CWO)}{(CC) + (CWO)} \right), \left( 1 - (1 - Pr_{ha} - Pr_{hc})^{\gamma_m} \right) \right] + a_{56} \left( 1 - Pr_{ha} - Pr_{hc} \right)^{\alpha_m} \right\} \\
\times per_m - (a_{63} \times per_m) \\
(7.1)
\]

\[
\frac{d}{dt}(tim_m) = \left\{ \left[ \left( \frac{a_{44}(CC) + a_{45}(CWO)}{(CC) + (CWO)} \right), \left( 1 - (1 - Pr_{ha} - Pr_{hc})^{\gamma_m} \right) \right] + a_{56} \left( 1 - Pr_{ha} - Pr_{hc} \right)^{\alpha_m} \right\} \\
\times tim_m - (a_{64} \times tim_m) \\
(7.2)
\]

\[
\frac{d}{dt}(pdp_m) = \left\{ \left[ \left( \frac{a_{44}(CC) + a_{45}(CWO)}{(CC) + (CWO)} \right), \left( 1 - (1 - Pr_{ha} - Pr_{hc})^{\gamma_m} \right) \right] + a_{56} \left( 1 - Pr_{ha} - Pr_{hc} \right)^{\alpha_m} \right\} \\
\times pdp_m - (a_{65} \times pdp_m) \\
(7.3)
\]

\[
\frac{d}{dt}(vri_m) = \left\{ \left[ \left( \frac{a_{46}(CC) + a_{47}(CWO)}{(CC) + (CWO)} \right), \left( 1 - (1 - Pr_{ha} - Pr_{hc})^{\gamma_m} \right) \right] + a_{56} \left( 1 - Pr_{ha} - Pr_{hc} \right)^{\alpha_m} \right\} \\
\times vri_m - (a_{66} \times vri_m) \\
(7.4)
\]

\[
\frac{d}{dt}(cwo_m) = \left\{ \left[ \left( \frac{a_{48}(CC) + a_{49}(CWO)}{(CC) + (CWO)} \right), \left( 1 - (1 - Pr_{ha} - Pr_{hc})^{\gamma_m} \right) \right] + a_{56} \left( 1 - Pr_{ha} - Pr_{hc} \right)^{\alpha_m} \right\} \\
\times cwo_m - (a_{68} \times cwo_m) \\
(7.5)
\]
\[
\frac{d(\text{clk}_m)}{dt} = \left\{ \left( a_{47}(\text{PDP}) + a_{53}(\text{VRI}) \right) \left[ 1 - (1 - \text{Pr}_u - \text{Pr}_r) \right] + a_{55}(1 - \text{Pr}_u - \text{Pr}_r) \right\} \times \text{clk}_m - (a_{67} \times \text{clk}_m)
\]

(7.6)

The above equations were derived based on Eqn 5.6.

iii) Time evolution of CWO, CLK, PER, VRI, TIM, and PDP1 proteins:

\[
\frac{d(\text{PER})}{dt} = (a_{97} \times \text{per}_m) - (a_{94} \times \text{PER} \times \text{TIM}) + (a_{95} \times \text{PT}) - (a_{85} \times \text{PER})
\]

(7.7)

\[
\frac{d(\text{TIM})}{dt} = (a_{98} \times \text{tim}_m) - (a_{94} \times \text{PER} \times \text{TIM}) + (a_{95} \times \text{PT}) - (a_{86} \times \text{TIM})
\]

(7.8)

\[
\frac{d(\text{PDP})}{dt} = (a_{59} \times \text{pdp}_m) - (a_{71} \times \text{PDP})
\]

(7.9)

\[
\frac{d(\text{VRI})}{dt} = (a_{60} \times \text{vri}_m) - (a_{72} \times \text{VRI})
\]

(8.0)

\[
\frac{d(\text{CLK})}{dt} = (a_{61} \times \text{clk}_m) - (a_{35} \times \text{CLK} \times \text{CYC}) + (a_{39} \times \text{CC}) - (a_{73} \times \text{CLK})
\]

(8.1)

\[
\frac{d(\text{CWO})}{dt} = (a_{62} \times \text{cwo}_m) - (a_{74} \times \text{CWO})
\]

(8.2)

In the above equations, the first and last variables denote translational and degradation rates respectively. The second function in Eqn 7.7, 7.8, and 8.1 represents formation of association complexes and the third term denotes the dissociation of synthesised complexes.

iv) Time evolution of CLK/CYC, PER/TIM, and PER/TIM/CLK/CYC complex:

\[
\frac{d(\text{PT})}{dt} = (a_{44} \times \text{PER} \times \text{TIM}) - (a_{95} \times \text{PT}) - (a_{38} \times \text{PT} \times \text{CC}) + (a_{42} \times \text{CCPT}) - (a_{103} \times \text{PT})
\]

(8.3)

\[
\frac{d(\text{CC})}{dt} = (a_{35} \times \text{CLK} \times \text{CYC}) - (a_{39} \times \text{CC}) - (a_{38} \times \text{PT} \times \text{CC}) + (a_{42} \times \text{CCPT}) - (a_{76} \times \text{CC})
\]

(8.4)
\[
\frac{d(CCPT)}{dt} = (a_{38} \times PT \times CC) - (a_{42} \times CCPT) - (a_{78} \times CCPT)
\] 

(8.5)

Explaining from above equations, the first and second terms express formation of an association complex and its dissociation respectively, and the last term denotes degradation of the complex. In Eqn 8.3 and 8.4, the third and fourth term denotes association and dissociation complexes respectively.

6.4 Implementing the models with parameter estimation

In general, standard methodologies used for models are 1.) deterministic and 2.) stochastic. Our models are an example of the deterministic type, since integration of ordinary differential equations (ODEs) are carried out, while Gillespie's algorithm is more commonly used in stochastic type models (Gillespie et al. 1976). After defining the system of ODEs, these mathematical models should be implemented in the computer so, we can use the computing power in order to estimate the parameters governing the rate equations using associated algorithms. By simulating reaction equations, stoichiometric network analysis and the computation of steady states can be used in computing elementary nodes (Abrams et al. 1999) to perform sensitivity analysis (Fell et al. 1997; Klulppel et al. 1997), optimisation and parameter estimation.

There are various programming languages and simple GUI tools available to help solve many problems encountered by systems biologists like sensitivity analysis, parameter estimation and bifurcation analysis. Among which parameter estimation is the time consuming issue (see section 6.4.2). In the systems biology community, MATLAB (Guide. 1998) or Mathematica (Maerder. 1991) are the widely used model simulation programming languages.

Several tools are developed, tested and released to satisfy the need for modelling software (see http://www.sbml.org). Many softwares packages in the free domain offer specific functionalities like flux analysis and are of limited use (Le Novere et al. 2001). Nevertheless, some tools satisfy a range of requirements like stochastic simulations of reaction networks, flux analysis and control (Tomita et al. 1999; Sauro
et al. 2003). In order to take advantage of a range of free tools there is a need for a common markup language so that, a model created in one tool can be freely tested and exchanged with another. Almost all modelling tools for biology can read the two prominent markup languages: 1.) “Systems Biology Markup Language” (SMBL) 2.) CellML. These languages were created to provide the much needed compatibility between multiple modelling platforms (Hucka et al. 2003; Lloyd et al. 2004). Among them the most prominent and widely used is the SBML format (Hucka et al. 2003; Hucka, et al. 2004), because biochemical reactions governed by enzyme kinetics are the primary focus in SBML models. Due to this inherent nature of SBML representation, until December, 2013, we noticed over 250 varied tools supporting the SBML format (http://sbml.org, accessed on 03, Dec, 2013). Using SBML, a system of reaction networks can be created by linking various participating reaction substrates. Yet only locally identifiable networks of the SBML can be interpreted by a tool which can “read” SBML. To convert these reaction networks in sets of discrete or continuous ODEs by assigning deterministic or stochastic functions we need a software tool which can “write” or “create” models in SBML format.

To simulate our models, COPASI (Mendes et al. 1993), a platform-independent and user-friendly biochemical simulator is used. In addition to its excellent user interface, COPASI can import and export SBML codes.

6.4.1 COPASI (Complex pathway simulator)

COPASI (COmplex PAthway SImlulator), is one of the best freely downloadable tools available to simulate and analyse biochemical reaction networks. Apart from the fact that its free, COPASI has compatible versions to support common operating systems like Windows, Linux, Mac OS X and Solaris. In addition it satisfies all the requirements for a good modelling tool (as discussed in the previous section), thus we made an easy decision to use COPASI. The history of COPASI lies in its predecessor Gepasi, one of the famous modelling tools widely used in early 1990’s (Mendes et al. 1993, 1997). In COPASI, appropriate ODEs are automatically created from reaction equations which is very useful even for those users who are new to computational modelling. Due to the flexibility and easy graphical user interface (GUI) we used COPASI as the main modelling software in this research (www.copasi.org).
6.4.2 Parameter estimation

Parameter estimation is a typical problem while modelling molecular networks. The models initially have unknown parameters. There is no known tool which can estimate the parameters in one go, the only way to find the appropriate parameters is by trial and error method (Zwolak et al. 2005). Parameter estimation is used to solve the inverse problem in a model i.e. to gather appropriate parameter values in order to validate the model with observed biological data is the main aim of parameter estimation. Hence, it is highly important but time consuming and needs a lot of patience and intuition. There are little or no available rates or kinetic parameters in biological networks at the molecular level, in particular for reactions involving cooperativity and stochastic noise. This is compounded by the reality that, biochemical reactions in biological systems at the molecular level are almost always random and stochastic.

In circadian models, even though some data is known in qualitative terms such as rhythmic regulation of translation, transcription and degradation of proteins and mRNAs in *Drosophila*, quantitative kinetic parameters are not known. But measured reaction rates in the circadian clock like expression rates of genes and degradation rates of proteins, are available for other model organisms (Shu et al. 2004). As of now, no quantitative rates have been published for circadian components in *Drosophila*. The usefulness of such measured reaction rates from wet-lab experiments will be very limited, since factors like standard operating procedures, state of cells and cell types which are different from lab to lab will certainly influence these quantities. Thus, the parameter estimation problem is inescapable in modelling genetic regulatory networks.

In our models, all the kinetic parameters involved to solve the reaction ODEs are unknown and needed to be estimated. In order to find a group of parameters which can simulate the observed biology we devised the following steps. At the start, a set of initial parameters were estimated, which can roughly produce 24 hour oscillations. This was achieved by using a combination of parameter scan, sensitivity and parameter estimation functions in COPASCI. Initial motive was to produce damped oscillations for all clock components from no oscillation state. We realised that
damped oscillations were achieved by increasing the sensitivity of local parameters in the respective ODE’s, in particular the degradation rate parameters. In parameter scan function, a single parameter is chosen and an objective maximum and minimum are given and the time course simulation is performed to see if there is any change in curve pattern. Then two dimensional scans are performed where, two parameters are independently scanned with two independent maximum and minimum values. COPASI achieves this by holding the minimum of the first parameter constant while scanning the second parameter and this process is carried forward with different parameters to increase the sensitivities. Similar to parameter scan, random distribution widget in COPASI was used to scan for random parameter values which can produce oscillations. This is time consuming process since, to scan for two independent parameters with 10 intervals for both parameters needs a 100 time course simulations in a two dimensional scan. Once damped oscillations were achieved, we were able to use manual parameter adjustments to produce irregular but strong oscillations and different parameter behaviours are studied at this stage, which are used to choose minimum number of parameters for estimations using estimation algorithms in COPASI. We were particularly impressed with Levenberg–Marquardt algorithm which arrived at the approximate fit. This step requires substantial computing time and patience. In our case, the problem was enormous since we needed to estimate three models, model A, model B and model C each with 75, 68 and 69 transient parameters respectively in addition to initial conditions. It has to be noted that, since many parameters from model A were seen to be compatible for model B it made our task a little easier.

There are many modelling tools one can use to simulate the experimental data by adjusting the parameters (Hoops et al. 2006). As mentioned before, we used COPASI in this study. To fit the appropriate parameters according to experiment data, COPASI uses a function which is based on a “weighted sum of squares” (WSS). The WSS is the sum for all data points’ measures of differences between experimental input and model predictions. While measuring this average of data points, only the ones with low variance are given greater relevance, since the calculated WSS is the inverse of the standard deviation. The discrepancies between predicted and measured data are reported as least square error, formulated as follows,
\[ WSS = \sum_{i=1}^{n} \frac{(y_{i}^{ex} - y_{i}^{mod})^2}{\sigma(y_{i}^{ex})} \] (8.6)

where, \( y_{i}^{ex} \) is the inputted experimental data for variables in different time points, \( y_{i}^{mod} \) is the simulated value for the same variables in inputted time points. Here, WSS is measured for all required time points and data points. Since low variance is better, to get better goodness of fit value WSS needs to be low.

COPASI is armed with various local and global optimisation algorithms (Mendes et al. 1998), which can be used to minimise WSS. To optimise the ODE’s with least square error, one of the best known algorithms currently is the Levenberg-Marquardt algorithm (LM) (Zwolak et al. 2005). It was observed in similar non-linear circadian models having least square problems, that the Levenberg-Marquardt algorithm approximated quickly to the best possible parameter fit (Xie et al. 2007). These findings, saved us a lot of time since we started using the LM algorithm from the start. An outline of the parameter estimation procedure is shown in figure 6-5.

Eventually, we were able to produce both protein and mRNA oscillations for all core clock components in a roughly 24 hour range using the above logic. We used different numerical optimisation algorithms initially but they failed due to the considerable number of unknown parameters. This shows that, in modelling biological systems at the molecular level, estimating the parameters is a very challenging task. In the second step, we were able to adjust the parameters by comparing the observed biological data to simulated output plots. Experimental dataset was produced from protein and mRNA levels based on visible approximate estimates from figures in we-lab publications. A good parameter set should produce indefinite oscillation of circadian clock components and the oscillations should have phase and amplitude similar to experimental data set. In *Drosophila* peacemaker neurons, exact concentrations of circadian clock components are unknown. According to earlier models, around 100 RNA and 1000 protein core clock molecules were assumed to be in a whole neuron (Vilar et al. 2002; Xie et al. 2007). For the initial concentrations,
Figure 6-5 Different steps of parameter estimation used in our models.
the range for mRNA’s was maintain between 0.3 – 0.4 nM and for proteins it was 3 – 4 nM. The corresponding number of molecules were calculated as follows.

In *Drosophila*, the radius of a Lateral Ventral Neuron is about 5 – 6 μm (Frisch et al. 1992), but the radius of the long axon which contributes to the length of the cell is about 2 μm (Tuthill et al. 2009) and so an average of 3 μm was assumed to be the radius of the cell in our model. The volume of the cell is \( V = \frac{4}{3} \pi r^3 = 1.13 \times 10^{-13} \text{ L} \). From which, we can calculate the unit number of molecules present in 1 nM using the Avogadro’s constant as follows,

\[
1 \text{nM} = (6 \times 10^{23} \text{ molecules/mole}) \times (10^{-9} \text{ mole/L}) \times (1.13 \times 10^{-13} \text{ L})
= 68 \text{ molecules (Approx)} \tag{8.7}
\]

The parameters were estimated and validated to reproduce the maximum number of biological observations and are shown in tables 1-7, 1-8, and 1-9 for models A, B and C respectively. It is necessary to note that, for all the kinetic parameters the units are nM per hour.
Table 1-7 Estimated kinetic parameter values for model A.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>value</th>
<th>Parameter</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1</td>
<td>0.8115</td>
<td>a50</td>
<td>19.1582</td>
</tr>
<tr>
<td>a2</td>
<td>0.8115</td>
<td>a51</td>
<td>17.5853</td>
</tr>
<tr>
<td>a3</td>
<td>0.0471</td>
<td>a52</td>
<td>19.2598</td>
</tr>
<tr>
<td>a4</td>
<td>0.0211</td>
<td>a53</td>
<td>0.1190</td>
</tr>
<tr>
<td>a5</td>
<td>0.2476</td>
<td>a54</td>
<td>26.8582</td>
</tr>
<tr>
<td>a6</td>
<td>12.2126</td>
<td>a55</td>
<td>58.8412</td>
</tr>
<tr>
<td>a7</td>
<td>0.0072</td>
<td>a56</td>
<td>0.0001</td>
</tr>
<tr>
<td>a8</td>
<td>0.1250</td>
<td>a57</td>
<td>33.8068</td>
</tr>
<tr>
<td>a9</td>
<td>0.4799</td>
<td>a58</td>
<td>33.8068</td>
</tr>
<tr>
<td>a10</td>
<td>0.4799</td>
<td>a59</td>
<td>1.9222</td>
</tr>
<tr>
<td>a11</td>
<td>0.4852</td>
<td>a60</td>
<td>11.5477</td>
</tr>
<tr>
<td>a12</td>
<td>0.3383</td>
<td>a61</td>
<td>35.7967</td>
</tr>
<tr>
<td>a18</td>
<td>0.6586</td>
<td>a62</td>
<td>14.2465</td>
</tr>
<tr>
<td>a19</td>
<td>0.6586</td>
<td>a63</td>
<td>0.0698</td>
</tr>
<tr>
<td>a20</td>
<td>0.0963</td>
<td>a64</td>
<td>0.0698</td>
</tr>
<tr>
<td>a21</td>
<td>0.1408</td>
<td>a65</td>
<td>0.0720</td>
</tr>
<tr>
<td>a22</td>
<td>0.2074</td>
<td>a66</td>
<td>0.0654</td>
</tr>
<tr>
<td>a23</td>
<td>2.1580</td>
<td>a67</td>
<td>0.6666</td>
</tr>
<tr>
<td>a24</td>
<td>0.1250</td>
<td>a68</td>
<td>0.0841</td>
</tr>
<tr>
<td>a25</td>
<td>22.5000</td>
<td>a71</td>
<td>0.1528</td>
</tr>
<tr>
<td>a26</td>
<td>10.3028</td>
<td>a72</td>
<td>0.9329</td>
</tr>
<tr>
<td>a27</td>
<td>10.3028</td>
<td>a73</td>
<td>0.0785</td>
</tr>
<tr>
<td>a28</td>
<td>10.1701</td>
<td>a74</td>
<td>1.0786</td>
</tr>
<tr>
<td>a29</td>
<td>14.4303</td>
<td>a76</td>
<td>0.0397</td>
</tr>
<tr>
<td>a35</td>
<td>2.8801</td>
<td>a78</td>
<td>26.6245</td>
</tr>
<tr>
<td>a38</td>
<td>90.3621</td>
<td>a79</td>
<td>1.0000</td>
</tr>
<tr>
<td>a39</td>
<td>0.4405</td>
<td>a80</td>
<td>1.0000</td>
</tr>
<tr>
<td>a42</td>
<td>2.1452</td>
<td>a81</td>
<td>1.0000</td>
</tr>
<tr>
<td>a43</td>
<td>18.3830</td>
<td>a82</td>
<td>1.0000</td>
</tr>
<tr>
<td>a44</td>
<td>18.3830</td>
<td>a85</td>
<td>0.4879</td>
</tr>
<tr>
<td>a45</td>
<td>18.0265</td>
<td>a86</td>
<td>0.4879</td>
</tr>
<tr>
<td>a46</td>
<td>61.9968</td>
<td>a94</td>
<td>1.1951</td>
</tr>
<tr>
<td>a47</td>
<td>119.0685</td>
<td>a95</td>
<td>1.4286</td>
</tr>
<tr>
<td>a48</td>
<td>1.7859</td>
<td>a103</td>
<td>0.0193</td>
</tr>
<tr>
<td>a49</td>
<td>19.1582</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1-8 Estimated kinetic parameter values for model B.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1</td>
<td>0.8383</td>
<td>a51</td>
<td>14.4091</td>
</tr>
<tr>
<td>a2</td>
<td>0.8383</td>
<td>a52</td>
<td>24.9594</td>
</tr>
<tr>
<td>a3</td>
<td>0.0260</td>
<td>a53</td>
<td>0.0019</td>
</tr>
<tr>
<td>a4</td>
<td>0.0226</td>
<td>a54</td>
<td>38.0013</td>
</tr>
<tr>
<td>a5</td>
<td>0.1808</td>
<td>a55</td>
<td>0.0011</td>
</tr>
<tr>
<td>a6</td>
<td>24.4253</td>
<td>a56</td>
<td>0.0001</td>
</tr>
<tr>
<td>a7</td>
<td>0.0092</td>
<td>a57</td>
<td>34.4307</td>
</tr>
<tr>
<td>a8</td>
<td>0.1040</td>
<td>a58</td>
<td>34.4307</td>
</tr>
<tr>
<td>a9</td>
<td>0.4110</td>
<td>a59</td>
<td>1.9325</td>
</tr>
<tr>
<td>a10</td>
<td>0.4110</td>
<td>a60</td>
<td>11.9584</td>
</tr>
<tr>
<td>a11</td>
<td>0.9705</td>
<td>a61</td>
<td>36.9362</td>
</tr>
<tr>
<td>a12</td>
<td>0.0013</td>
<td>a62</td>
<td>14.3208</td>
</tr>
<tr>
<td>a18</td>
<td>0.7460</td>
<td>a63</td>
<td>0.0683</td>
</tr>
<tr>
<td>a19</td>
<td>0.7460</td>
<td>a64</td>
<td>0.0683</td>
</tr>
<tr>
<td>a20</td>
<td>0.0930</td>
<td>a65</td>
<td>0.0787</td>
</tr>
<tr>
<td>a21</td>
<td>0.1370</td>
<td>a66</td>
<td>0.0620</td>
</tr>
<tr>
<td>a22</td>
<td>0.2009</td>
<td>a67</td>
<td>0.6809</td>
</tr>
<tr>
<td>a23</td>
<td>0.6768</td>
<td>a68</td>
<td>0.1330</td>
</tr>
<tr>
<td>a24</td>
<td>0.0165</td>
<td>a71</td>
<td>0.1505</td>
</tr>
<tr>
<td>a25</td>
<td>3.7136</td>
<td>a72</td>
<td>0.9827</td>
</tr>
<tr>
<td>a26</td>
<td>12.2271</td>
<td>a73</td>
<td>0.0613</td>
</tr>
<tr>
<td>a27</td>
<td>12.2271</td>
<td>a74</td>
<td>1.0794</td>
</tr>
<tr>
<td>a28</td>
<td>3.5369</td>
<td>a76</td>
<td>0.0496</td>
</tr>
<tr>
<td>a29</td>
<td>28.8605</td>
<td>a78</td>
<td>0.0001</td>
</tr>
<tr>
<td>a35</td>
<td>2.8381</td>
<td>a79</td>
<td>1.0000</td>
</tr>
<tr>
<td>a38</td>
<td>43.5451</td>
<td>a80</td>
<td>1.0000</td>
</tr>
<tr>
<td>a39</td>
<td>0.3810</td>
<td>a81</td>
<td>1.0000</td>
</tr>
<tr>
<td>a42</td>
<td>0.0005</td>
<td>a82</td>
<td>1.0000</td>
</tr>
<tr>
<td>a43</td>
<td>19.0627</td>
<td>a85</td>
<td>0.5740</td>
</tr>
<tr>
<td>a44</td>
<td>19.0627</td>
<td>a86</td>
<td>0.5740</td>
</tr>
<tr>
<td>a45</td>
<td>17.8536</td>
<td>a94</td>
<td>1.2806</td>
</tr>
<tr>
<td>a46</td>
<td>66.0652</td>
<td>a95</td>
<td>1.5048</td>
</tr>
<tr>
<td>a47</td>
<td>116.8552</td>
<td>a103</td>
<td>0.0319</td>
</tr>
<tr>
<td>a48</td>
<td>0.0001</td>
<td>a37</td>
<td>1.2385</td>
</tr>
<tr>
<td>a49</td>
<td>21.0200</td>
<td>a41</td>
<td>0.4795</td>
</tr>
<tr>
<td>a50</td>
<td>21.0200</td>
<td>a77</td>
<td>0.4798</td>
</tr>
</tbody>
</table>
Table 1-9 Estimated kinetic parameter values for model C.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1</td>
<td>0.8115</td>
<td>a51</td>
<td>10.6784</td>
</tr>
<tr>
<td>a2</td>
<td>0.8115</td>
<td>a52</td>
<td>48.3504</td>
</tr>
<tr>
<td>a3</td>
<td>0.0471</td>
<td>a53</td>
<td>1.3285</td>
</tr>
<tr>
<td>a4</td>
<td>0.0211</td>
<td>a54</td>
<td>29.8500</td>
</tr>
<tr>
<td>a5</td>
<td>0.2476</td>
<td>a55</td>
<td>10.4627</td>
</tr>
<tr>
<td>a6</td>
<td>12.2126</td>
<td>a56</td>
<td>0.0100</td>
</tr>
<tr>
<td>a7</td>
<td>0.0072</td>
<td>a57</td>
<td>36.0983</td>
</tr>
<tr>
<td>a8</td>
<td>0.1250</td>
<td>a58</td>
<td>36.0983</td>
</tr>
<tr>
<td>a9</td>
<td>0.4799</td>
<td>a59</td>
<td>1.5655</td>
</tr>
<tr>
<td>a10</td>
<td>0.4799</td>
<td>a60</td>
<td>11.1768</td>
</tr>
<tr>
<td>a11</td>
<td>0.4852</td>
<td>a61</td>
<td>35.0620</td>
</tr>
<tr>
<td>a12</td>
<td>0.3383</td>
<td>a62</td>
<td>17.4692</td>
</tr>
<tr>
<td>a18</td>
<td>0.6586</td>
<td>a63</td>
<td>0.0466</td>
</tr>
<tr>
<td>a19</td>
<td>0.6586</td>
<td>a64</td>
<td>0.0466</td>
</tr>
<tr>
<td>a20</td>
<td>0.0963</td>
<td>a65</td>
<td>0.0488</td>
</tr>
<tr>
<td>a21</td>
<td>0.1408</td>
<td>a66</td>
<td>0.1106</td>
</tr>
<tr>
<td>a22</td>
<td>0.2074</td>
<td>a67</td>
<td>0.6815</td>
</tr>
<tr>
<td>a23</td>
<td>2.1580</td>
<td>a68</td>
<td>0.1100</td>
</tr>
<tr>
<td>a24</td>
<td>0.1250</td>
<td>a71</td>
<td>0.1219</td>
</tr>
<tr>
<td>a25</td>
<td>22.5000</td>
<td>a72</td>
<td>0.8968</td>
</tr>
<tr>
<td>a26</td>
<td>10.3028</td>
<td>a73</td>
<td>0.0260</td>
</tr>
<tr>
<td>a27</td>
<td>10.3028</td>
<td>a74</td>
<td>1.3428</td>
</tr>
<tr>
<td>a28</td>
<td>10.1701</td>
<td>a76</td>
<td>0.0113</td>
</tr>
<tr>
<td>a29</td>
<td>14.4303</td>
<td>a78</td>
<td>0.4014</td>
</tr>
<tr>
<td>a35</td>
<td>2.8801</td>
<td>a79</td>
<td>1.0000</td>
</tr>
<tr>
<td>a38</td>
<td>47.8645</td>
<td>a80</td>
<td>1.0000</td>
</tr>
<tr>
<td>a39</td>
<td>0.5973</td>
<td>a81</td>
<td>1.0000</td>
</tr>
<tr>
<td>a42</td>
<td>0.9468</td>
<td>a82</td>
<td>1.0000</td>
</tr>
<tr>
<td>a43</td>
<td>16.4783</td>
<td>a85</td>
<td>0.2006</td>
</tr>
<tr>
<td>a44</td>
<td>16.4783</td>
<td>a86</td>
<td>0.2006</td>
</tr>
<tr>
<td>a45</td>
<td>18.4941</td>
<td>a94</td>
<td>1.3727</td>
</tr>
<tr>
<td>a46</td>
<td>81.9441</td>
<td>a95</td>
<td>2.8600</td>
</tr>
<tr>
<td>a47</td>
<td>122.7670</td>
<td>a103</td>
<td>0.1171</td>
</tr>
<tr>
<td>a48</td>
<td>2.3343</td>
<td>a104</td>
<td>0.0194</td>
</tr>
<tr>
<td>a49</td>
<td>6.2160</td>
<td>a105</td>
<td>1.1585</td>
</tr>
<tr>
<td>a50</td>
<td>6.2160</td>
<td>a106</td>
<td>0.0100</td>
</tr>
</tbody>
</table>
Table 2-1 Initial conditions for all three models.

<table>
<thead>
<tr>
<th>Species</th>
<th>Initial value</th>
<th>Probability value</th>
</tr>
</thead>
<tbody>
<tr>
<td>perm</td>
<td>0.2395</td>
<td>Preper 0.0431</td>
</tr>
<tr>
<td>timm</td>
<td>0.2395</td>
<td>Prc 0.043</td>
</tr>
<tr>
<td>pdpm</td>
<td>0.3175</td>
<td>Prcpdp 0.08</td>
</tr>
<tr>
<td>vrim</td>
<td>0.2571</td>
<td>Prcv 0.0585</td>
</tr>
<tr>
<td>cwom</td>
<td>0.2156</td>
<td>Prccwo 0.043</td>
</tr>
<tr>
<td>clkm</td>
<td>0.2583</td>
<td>Prcwwo 0.043</td>
</tr>
<tr>
<td>PER</td>
<td>2.7527</td>
<td>Prcwper 0.0431</td>
</tr>
<tr>
<td>TIM</td>
<td>2.7527</td>
<td>Prcwt 0.043</td>
</tr>
<tr>
<td>PDP</td>
<td>4.1953</td>
<td>Prcwpdp 0.08</td>
</tr>
<tr>
<td>VRI</td>
<td>3.175</td>
<td>Prcwv 0.0585</td>
</tr>
<tr>
<td>CLK</td>
<td>3.6628</td>
<td>Prcv 0.489</td>
</tr>
<tr>
<td>CWO</td>
<td>2.4774</td>
<td>Prpc 0.426</td>
</tr>
<tr>
<td>PT</td>
<td>0.4014</td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td>0.5566</td>
<td></td>
</tr>
<tr>
<td>CCPT</td>
<td>0.4982</td>
<td></td>
</tr>
<tr>
<td>CYC</td>
<td>1</td>
<td>Pern 0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Timn 0.003</td>
</tr>
<tr>
<td>Species</td>
<td>Initial value</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>---------------</td>
<td>------------------</td>
</tr>
<tr>
<td>CWD</td>
<td>1.3416</td>
<td>Vrin 0.003</td>
</tr>
<tr>
<td>HP</td>
<td>1.5632</td>
<td>Cwom 0.003</td>
</tr>
<tr>
<td>CWPT**</td>
<td>1.4201</td>
<td></td>
</tr>
</tbody>
</table>

* Denotes additional species in model B only, and ** Denotes new species in model C only.
Chapter 7: Simulation results and testing

This chapter deals with the simulation results of our models including a variety of testing outputs which are compared with in vitro findings. We have done a local sensitivity analysis to understand the effects of the new CWO negative feedback loop in our models. We have simulated the models for known circadian mutants in DD (constant darkness). In addition, we have simulated the entrainment of light and tested the robustness. All model outputs for the above said settings are compared with results from wet-lab experiments. In the final section (7.5.3), we design a hypothetical cwo\textsuperscript{CWPT} mutant which hypothetically answers how CWO might functions both as an activator and repressor in the pacemaker neurons.

7.1 Simulations results of Models A, B and C

7.1.1 Oscillation of clock components in DD (constant darkness)

Our model output in DD produced indefinite oscillations of clock components at appropriate period and amplitude. Using the estimated parameters provided in tables 1-7, 1-8, and 1-9 for models A, B and C, all our three models showed rhythmic 24 hour oscillations for \textit{cwo}, \textit{pdp1}, \textit{vri}, \textit{per}, \textit{clk} and \textit{tim} mRNAs including their synthesised clock proteins CWO, PDP1, VRI, PER, CLK and TIM respectively.

According to wet-lab observations, \textit{tim}, \textit{per} and \textit{cwo} mRNAs amplitude peaks are achieved at “Circadian Time” (A standard of time based on period of a rhythm e.g. CT = 0 at the onset of oscillations) CT 12 – CT 16 i.e. early evening and their oscillations are in same phase (Lim et al. 2007; Kadener et al. 2007; Richier et al. 2008). It was reported that, concentration levels of \textit{clk} mRNA peak between CT 23 – CT 4, i.e. late night to early morning and its locomotor activity rhythms are in opposite phase with \textit{tim} and \textit{per} mRNA oscillations (Bae et al. 1998). Similarly, we know from research reports that, \textit{vri} and \textit{pdp1} mRNAs are in anti-phase with \textit{clk} mRNA oscillations, their phase is similar to the \textit{tim} and \textit{per} mRNA phase and they both peak at roughly CT 12 – CT 14 (Glossop et al. 2003; Cyran et al. 2003).
The simulated plots for mRNA oscillations are shown in Figure 7-1. From these we can see that *tim*, *per* and *cwo* mRNAs peaked at “Model Time” (A standard of time based on period of a rhythm in model simulation) approximately 9.2, 9.2 and 12.7 hours in model A, next 8.9, 8.9 and 12.4 hours in model B and finally 9.2, 9.2 and 13.1 hours in model C respectively. The *clk* mRNAs peaked at model time 3.8, 3.9 and 3.6 hours in models A, B and C respectively, and were observed to be in anti-phase with *tim*, *per* and *cwo* mRNAs in all three models. The *vri* and *pdp1* mRNAs were noticed with peaks at model time 8.6 and 13 in model A, 8.7 and 14 in model B, 11.2 and 11 hours in model C respectively. The phase and anti-phase relationship between different mRNA oscillations were similar to observed biological results.

Findings from experimental research showed that, after a 4 – 6 hours delay from their mRNA peaks, the peaks of PER and TIM proteins were observed; CLK protein levels peaked at CT 4.5 (Rosbash et al. 1996). Peaks of CWO oscillation were observed at CT 15 (Lim et al. 2007). VRI protein oscillations peaked as soon as its mRNA is synthesised i.e. CT 12 approximately; on the other hand PDP1 proteins exhibit a 3 – 6 hours’ time lag from their mRNA synthesis i.e. around CT 18 (Cyran et al. 2003). The reason for such time delay is not known.

Oscillations of circadian clock proteins simulated in our models are shown in figure 7-2. Figure 7-2 shows the DD oscillation peaks of the proteins. In our simulated output, PER, TIM, CLK, VRI, PDP1 and CWO proteins peaked at model time 15.3, 15.3, 4.5, 10.2, 17.8 and 14.8 hours respectively in model A. Similarly, the peaks were observed for the same proteins at model time 15, 15, 4.5, 10.6, 18 and 13.5 in model B and 15.9, 15.9, 5, 12, 17.2 and 14.5 hours in model C respectively. The majority of our simulated peaks in model time fall within the biologically observed circadian time. The phase and anti-phase relationships between clock components were maintained.

The objective of our parameter estimations was to produce limit cycle oscillations of all circadian components, each with appropriate rhythmic period and amplitude. In our results, we found that all oscillations were maintained with an approximately 24 hour period mimicking daily rhythms. In all three models, the amplitudes of mRNA
Figure 7-1 DD Oscillation of mRNAs: (a) In model A, (b) In model B, and (c) In model C.
Figure 7-2 DD Oscillation of proteins: (a) In model A, (b) In model B, and (c) In model C.
oscillations were maintained between 0.1 to 0.5 nM equivalent to approximately 7 to 34 copy numbers. Similarly, the amplitudes of protein concentrations were maintained between 2 to 5 nM in all three models which corresponds to 136 to 340 protein molecules in the whole cell. It has to be noted that we were able to reproduce rhythmic oscillations of all clock components indefinitely in all three models. These results are in itself is a great accomplishment considering the huge number of unknown parameters needed to be estimated for all three models.

While simulating clock oscillations in the models, the times taken for each component to achieve limit cycles slightly vary from each other. Due to this effect, the simulated model time differs from the observed circadian time by +/- 1 to 3 hours in total 9 out of 36 mRNA and protein oscillations as shown in table 2-2. The exact circadian time peaks are not necessary while testing the models through mutational analysis and to investigate our overall research question, since we only needed to simulate the relative change in concentrations, amplitude, period length and phase shifts of oscillations to test our models. These tests can be simulated by analysing the results in model time. Yet, in order to increase the biological resemblance of our models, model time was converted to circadian time by manually shifting the oscillations of these components at the start of the limit cycle +/- 1 to 3 hours. Lists of oscillatory peaks in circadian time for all three models are provided in table 2-2.

7.1.2 Robustness of the models to parameter variations

The Drosophila circadian components have the ability to regulate the phase relationships with change in different extrinsic noise. Experiments in in vivo and in vitro conditions show that even in viable mutant flies and wild type flies perturbation in environmental factors such as temperature and light only produced minor changes in molecular oscillation of clock components. Delaying or accelerating the exposure of Drosophila flies entrained in DD, to light by a few hours only resulted in CT 0.1 hour period variations at molecular level (Levine et al. 2002). Similarly, in a different experiment, flies were entrained at temperatures ranging from 29 °C to 20 °C. It was found that there was only a negligible period of variation in clock oscillation towards CT 0.1, CT 0.2 and CT 0.06 for corresponding temperatures of 20 °C, 25 °C and 29 °
### Table 2-2 Oscillation peaks of clock components in DD (Constant darkness)

<table>
<thead>
<tr>
<th>Component</th>
<th>Experimental Output in CT</th>
<th>Simulated Outputs CT</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Model A</td>
<td>Model B</td>
</tr>
<tr>
<td><em>per</em> mRNA</td>
<td>12 - 16</td>
<td>12.2</td>
<td>12.9</td>
</tr>
<tr>
<td><em>tim</em> mRNA</td>
<td>12 - 16</td>
<td>12.2</td>
<td>12.9</td>
</tr>
<tr>
<td><em>clk</em> mRNA</td>
<td>23 - 4</td>
<td>3.8</td>
<td>3.6</td>
</tr>
<tr>
<td><em>pdp1</em> mRNA</td>
<td>12 - 14</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td><em>vri</em> mRNA</td>
<td>12 - 14</td>
<td>12.6</td>
<td>12.7</td>
</tr>
<tr>
<td><em>cwo</em> mRNA</td>
<td>12 - 16</td>
<td>12.7</td>
<td>12.4</td>
</tr>
<tr>
<td>PER protein</td>
<td>~ 15</td>
<td>15.3</td>
<td>15</td>
</tr>
<tr>
<td>TIM protein</td>
<td>~ 15</td>
<td>15.3</td>
<td>15</td>
</tr>
<tr>
<td>CLK protein</td>
<td>~ 4.5</td>
<td>4.5</td>
<td>4.5</td>
</tr>
<tr>
<td>VRI protein</td>
<td>~ 12</td>
<td>12.2</td>
<td>11.6</td>
</tr>
<tr>
<td>PDP1 protein</td>
<td>~ 18</td>
<td>17.8</td>
<td>18</td>
</tr>
<tr>
<td>CWO protein</td>
<td>~ 15</td>
<td>14.8</td>
<td>14.5</td>
</tr>
</tbody>
</table>
C (Bao et al. 2001). Even though these temperature compensation period variations appear negligible, over a period of time roughly 20% - 30% flies entrained between 20°C - 25°C ended up arrhythmic. Therefore, a good mathematical model should be very robust with rhythmic oscillation and perturbation through parameters should produce less than CT 1 hour variations so that the flies can be rhythmic. A detailed analysis should investigate the robustness through parameter changes in a wide variety of space levels. But it is nearly impossible to do this globally in the three systems with 65, 68 and 69 transient parameters, hence we followed the local techniques proposed in earlier models (Leloup et al. 2003; Xie et al. 2007). Where, in a stable model with estimated parameters, only one is changed at any given time, usually by +/- 20% (Lema et al. 2000; Smolen et al. 2004).

Accordingly, robustness of all three models was checked by introducing perturbations with 20% increase or decrease from estimated parameter values one at a time. Local perturbation simulations carried out were in the order of 130 (20% +/- in 65 parameters), 136 (20% +/- in 68 parameters) and 138 (20% +/- in 69 parameters) for models A, B and C respectively without including the control simulations. We found in all three systems, that oscillations of clock components were maintained indefinitely. For +/- 20% change in parameter values, the period variations compared to unperturbed values were less than 0.4 hours in a majority of transient parameter values in all three models (Figure 7-3). Specifically in models A, B and C, out of 65, 58 and 69 original values, 59, 62 and 63 maintained period variations to less than 0.8 hours. This shows that our model parameters are very robust. A biological interpretation of this result is that, all three of our systems will survive and can adapt to change in external noise.

The biggest difference in period of more than 0.4 hours was caused by the same six parameters in all three models (Figure 7-4). Perturbation in parameters a43, a47, a57, a61, a63 and a67 resulted in a period difference of more than 0.6 hours. These parameters correspond to transient rate of transcription of per gene, transient rate of transcription of clk gene, transient rate of translation of per mRNA, transient rate of translation of clk mRNA, degradation rate of synthesised per mRNA and degradation rate of synthesised clk mRNA respectively. Among these six parameters, the largest
Figure 7-3 Robustness to parameter variations: (a) In model A, (b) In model B, and (c) In model C.
Figure 7-4 A 3D chart of largest period difference effected by 6 parameters are shown: (a) In model A and B; (b) in model C.
period increase and decrease of more than 0.8 hours was caused by a57, which is related to PER protein synthesis.

In the earlier Smolen’s model, similar +/- 20% changes in parameter values were used to check the robustness. It was found in his model that greater than 3 hour period variations were observed for all parameter perturbations. A largest period difference of higher than 3.5 hours occurred with a parameter governing the PER nuclear accumulation, which exists since this model is designed with separate nuclear and cytoplasmic compartments (Smolen et al. 2004). Compared with Smolen’s circadian model governed by Michaelis-Menten kinetics, we can see that, our detailed models with individual rate equations governed by mass action kinetics are more robust and closer to biological observations with largest observable period difference of 0.8 hours with respect to a57. Our models are similar to Xie’s circadian model and robustness was checked with the same +/- 20% parameter perturbations. In their model all period variations were limited to 0.8 hours, and the largest difference of 0.75 hours was observed in three parameters related to transcription initiation state. They were the transient rate of PDP1 protein binding to the V/P box of the clk gene and, the transient rates of the CLK/CYC activator complex binding to E-boxes in per and pdp1 promoters (Xie et al. 2007). Four out of six large period differences in our models were shown in parameters related to translation and degradation rates. Hence, even though the architecture of our models is similar to Xie’s model in terms of using probability based mass action governed ODEs, all three of our models are more sensitive to the post-translational modification rate limiting steps. Our model parameters and their sensitivities are totally different from previous models implying that these models have considerable changes in conceptual facets, and are defined in a substantially different parameter space.

7.2 Light entrainment of the circadian clock

7.2.1 Effect of light

After the models were estimated and analysed as described in the previous section, the next step was to test whether they could predict the effects of external perturbations (zeitgebers). Although both temperature and light cycle every day, among them light
is known to be the more dominant zeitgeber (Aschoff et al. 1974). A major characteristic of the circadian clock is its capacity to respond to different light conditions in the environment. The nature of the circadian molecular clock is such that it should respond to different intensities in light in terms of exposure time. Light entrainment is observed through resetting the phase of circadian oscillators. Biologically, in the *Drosophila* circadian clock, exposure to light leads to degradation of TIM protein through its interaction with light activated photoreceptor CRY (cryptochrome). Mutational screenings have shown that, absence of TIM protein in a *tim*01 mutant leads to reduced levels of cytoplasmic PER repressor and exposing wild-type flies to constant light produced the same result (Zerr et al. 1990; Vosshall et al. 1994; Price et al. 1995). Degradation of TIM was shown to affect the oscillation phase of other clock proteins and a phase reset in rhythms takes place (Rosbash et al. 1996). This ability of internal clock to be in synch with external noise contradicts with the robustness of clock component oscillations, which is a very good example for the stochasticity of biological systems. In our opinion, a mathematical model with two compartments can be constructed in future. A first group of ODEs with robust parameters to depict endogenous core clock components and a second group of ODEs with sensitive parameters in order to depict environmental realism. This approach in a future model could reflect the observed environmental resetting mechanism more accurately.

The primary effect of TIM degradation is the destabilisation of PER concentrations in cytoplasm since the lack of TIM protein leaves PER permanently bound to phosphorylating kinase DBT resulting in hyper-phosphorylation and eventual degradation of PER protein (Shafer et al. 2002). In some previous models, this indirect effect of light, i.e. PER degradation was used to simulate light entrainment by increasing the parameter value governing the transient rate of the degradation of PER protein (Olde et al. 1999; Lema et al. 2000). For example in Smolen’s model, the parameter for PER degradation called PERtot was replaced with a new value “Klight”, which was 20% higher than the original value (Smolen et al. 2004). In some other earlier models, the parameter governing TIM protein degradation was increased to model light entrainment (Leloup et al. 1998; Leloup et al. 2003). Likewise in Xie’s model, TIM and PER protein degradation rates “dtim” and “dper” were replaced with higher “klight” values (Xie et al 2007). In a more recent model, microarray data of *cry*
mRNA cycling was used to describe a rate equation for CRY and the relative concentration of CRY protein was used as a regulatory weight in the rate equation describing TIM protein synthesis. Thus, light entrainment is simulated by “turning on” the CRY equation (Fathallah-Shaykh et al. 2009). We ignored the CRY component in our model architecture and also PER and TIM protein phosphorylation components. In order to simulate light entrainment, we followed the method used in previous models, where the transient rates of degradation of PER and TIM proteins are increased (Xie et al. 2007).

Accordingly, the parameter rates of ‘a85’ and ‘a86’ for PER and TIM degradation, as shown in tables 1-7, 1-8 and 1-9, are replaced with a new “klight” rate for all three models. We studied the response of the model to a 12-hour alternate zeitgeber time (ZT) where the time of light signal input is considered ZT 0. Thus in a light dark (LD) cycle ZT 0 – ZT 12 is considered the light phase and the following ZT 12 – ZT 24 is considered the dark phase. In order to simulate the light phase of the LD cycle, parameters ‘a85’ and ‘a86’ are replaced by an arbitrarily chosen ‘klight’ rate value of 1. Alternatively, since our model parameters are estimated and validated to satisfy oscillations in constant darkness, to simulate the dark phase of the LD cycle the ‘klight rates were reverted back to their original values.

The LD simulations of clock proteins are plotted as shown in Figure 7-5. In the LD cycle, with a klight rate value of 1, rhythmic protein oscillations were maintained for all proteins in all three models. Similar to DD conditions, protein and mRNA phase and anti-phase relationships were observed in LD cycles as well. To observe the phase resetting more easily, comparative CLK protein levels were plotted in Figure 7-6 with klight and kdark degradation rates. As we can see from the results in all our three models light entrainment can be simulated. To simulate the arrhythmic effect of constant light (LL) conditions, we used the same methods followed in previous models (Qiu et al. 1996, Xie et al. 2007). A high klight value 5 was used to replace ‘a85’ and ‘a86’, the PER and TIM degradation rates. The LL simulations are plotted in Figure 7-7, which shows the arrhythmic damped oscillations of all proteins similar to in vivo results (Price et al. 1995). Thus we can say with confidence, that all our three models satisfy the experimental clock oscillation in all light conditions (DD, LD and LL).
Figure 7-5 LD oscillation: (a) In model A, (b) In model B, and (c) In model C.
Figure 7-6 LD oscillation showing phase shift: (a) In model A, (b) In model B, and (c) In model C.
Figure 7-7 Damped protein oscillation in high light: (a) In model A, (b) In model B, and (c) In model C.
7.3 Testing mutations

Next, all three models were tested to predict accurately the molecular effects of mutations in the core clock components. As discussed in previous chapters, circadian rhythms in *Drosophila* are influenced by a variety of mutations. The common methodology to test mutations in mathematical models is by introducing parametric perturbations to mimic the functional properties of the mutants. In this section mutations related to *clk*, *tim* and *per* genes are discussed. Mutations with respect to the *cwo* gene are discussed in a detailed fashion in section 7.5.

7.3.1 Testing null mutants

First, null mutants were tested. Biologically, null mutants are viable but produce proteins with defective functional domains. In *Drosophila* there are three well known null mutants with mutations on crucial clock proteins *clk*, *tim* and *per*, which are known as *clk*\(^{Jrk}\), *tim*\(^{01}\) and *per*\(^{01}\) respectively. The transient rates of translation of these respective genes were set to zero in order to simulate these mutants.

It was found from the simulations that, rhythmic oscillations of all clock proteins were lost in *clk*\(^{Jrk}\), *tim*\(^{01}\) and *per*\(^{01}\) in silico mutants as shown in Figures 7-8 and 7-9. Our results are in concurrence with experimental findings that rhythmicity and oscillation of *clk*, *per*, *tim*, *pdp1* and *vri* gene transcripts were abolished in *clk*\(^{Jrk}\), *tim*\(^{01}\) and *per*\(^{01}\) mutants (Bae et al. 1998; Cyran et al. 2003). There were some divergence from research reports with respect to mRNA concentration levels. The mRNA levels of *tim* and *per* genes were observed to be low in *tim*\(^{01}\) and *per*\(^{01}\) mutants (So et al. 1997), but in all our three model simulations, *tim* and *per* mRNA levels were maintained at moderate levels as shown in Figure 7-9. High levels of *pdp1* mRNAs and intermediate levels of *vri* mRNA were biologically observed in *tim*\(^{01}\) and *per*\(^{01}\) mutants (Blau et al. 1999; Cyran et al. 2003. On the other hand in our models, *vri* mRNA concentrations were very high and *pdp1* mRNA levels were at intermediate levels. Yet in all three models, the *clk* mRNA levels in *tim*\(^{01}\) and *per*\(^{01}\) mutants were identically low, similar to wet-lab findings as shown in Figure 7-8 (Glossop et al. 1999).
Figure 7-8 Damped mRNA oscillation in $per^{01} \ tim^{01}$ mutants: (a) In model A, (b) In model B, and (c) In model C.
Figure 7-9 Damped mRNA oscillation in $clk^{jrk}$ mutants: (a) In model A, (b) In model B, and (c) In model C.
The fact that *tim* and *per* mRNA levels were not low might be due to the reaction structure of our models and the nature of perturbed parameters. Since we simulated the mutant, by setting the transient rate of translation of *tim* and *per* mRNAs to zero the effective transcription repression handled by the PER and TIM proteins has been virtually removed. In our in silico *tim* and *per* mutants, the transient concentrations of TIM and PER proteins are found to be zero. Thus, the difference in simulated and experimental results is that, due to the total lack of PER and TIM in the system, CLK protein is free to synthesise more *tim* and *per* mRNAs.

We found excellent agreement between simulated and experimental data with respect to clock mRNA levels in mutant *clk* flies. All clock mRNA components were arrhythmic with a very low level of vri, *tim*, *pdp1* and *per* mRNAs and high levels of *clk* mRNA concentrations near to wild-type peaks were observed in *in vivo* *clk* mutants (Allada et al. 1998; Glossop et al. 1999; Cyran et al. 2003). The *in silico* *clk* mutant result showed not only the abolition of mRNA oscillations, but it was accurate to the extent that the high levels of *clk* mRNA were near unperturbed simulation levels.

### 7.3.2 Testing *per* and *per* mutants

In *Drosophila*, apart from arrhythmic null mutants, other mutants such as *per* and *per* with viable free running clock oscillations have been found. The superscripts ‘L’ and ‘S’ in their names indicate the nature of the mutants. The *per* mutant produces a clock oscillation with a lengthened period (Huang et al. 1995). By contrast, the *per* mutant produces an activity rhythm with shortened period. Approximately 5 hours period lengthening (to 29 hours) and shortening (to 19 hours) was observed in *per* and *per* mutants respectively (Konopka et al. 1971). Compared to wild-type flies, it was found that in the *per* mutant there is a time delay in the nuclear accumulation of PER protein a reason for which is not known (Lee et al. 1996). Similarly in the case of *per* mutants the nuclear accumulation of phosphorylated PER protein is found to happen earlier than usual in the morning (Edery et al. 1994).

Even though the exact rate limiting step is not known biologically for these mutants, previous *Drosophila* circadian models were shown to replicate the biological results.
by altering the PER and TIM protein stability. In one of the earlier models published immediately after per\textsuperscript{L} discovery, the long activity rhythms were produced by adjusting the parameters governing stability of the PER/TIM dimer complex (Ruoff et al. 1996). Likewise in more recent models, per\textsuperscript{L} and per\textsuperscript{S} mutants were tested by increasing and decreasing the degradation rates of time evolving TIM and PER proteins. We followed the same procedure to test the per\textsuperscript{L} and per\textsuperscript{S} mutants in all three models, by setting the TIM and PER degradation rates as 0.1 for per\textsuperscript{L} and 1 for per\textsuperscript{S} in models A and B respectively. And for model C the degradation rates were set at 0.05 and 0.7 to test the per\textsuperscript{L} and per\textsuperscript{S} mutants respectively. We found that, similarly to previous models, our models can be trained to produce period changes of 19 hours and 29 hours by adjusting the transient rate of degradation of TIM and PER proteins. A comparative plot of these mutant results for all the models is provided in Figure 7-10.

### 7.4 Local sensitivity analysis

In order to maintain a robust circadian clock, experimental observations in various biological systems showed that interlocked feedback loops are the key (Cheng et al. 2001). As discussed in the previous sections, to maintain the rhythmic oscillation of circadian clock components, the effective translation of clk, per and tim genes resulting in the synthesis of their translated proteins is very important. Thus it is well known that CLK transcriptional activation loops with PER/TIM self-repression loops are vital. With the inclusion of a CWO negative feedback loop in our new models, we have the opportunity to test the role of the CWO loop in maintaining the robustness of the circadian clock. Wet-lab analysis has suggested that CWO acts as a behavioural rhythm amplifier and that it has a modulatory effect in repressing the CLK/CYC activation of E-boxes (Lim et al. 2007). Hence, it is imperative to check whether the robustness of the Drosophila circadian clock is increased by the CWO feedback loop. In the previous CWO model, to understand the effects of CWO on period oscillations, simulations where performed by increasing or decreasing (up to 10\%) of the individual repressive weights of CWO in DD conditions (Fathallah-Shaykh et al. 2009). In our models, to understand the behaviour of the CWO loop better, we performed a sensitivity analysis. In order to quantify the robustness occurring due to
Figure 7-10 Period changes in PER protein oscillation of \textit{per}^L and \textit{per}^S mutants: (a) In model A, (b) In model B, and (c) In model C.
CWO loops, we compared the estimated local sensitivities of all parameters with and without the CWO feedback loop. To perform the sensitivity analysis without the CWO feedback loop, the oscillation of CWO protein was stopped and its value was fixed. We found in resulting simulations that, all core clock proteins and mRNA oscillations were preserved except CWO in all three models. The plots of protein oscillations in all three models without CWO loop are provided in Figure 7-11. Next, the periodic sensitivity ratios were calculated from sensitivity outputs. In order to understand the sensitivity outputs, it is necessary to understand how local sensitivity is calculated in dynamical systems based on ODEs.

In oscillatory systems local sensitivity analysis is carried out by measuring the period difference in oscillation with respect to perturbation in individual parameter values. To generate local sensitivities, 5% parametric perturbations were assigned and the sensitivities were calculated based on the following function,

\[
S_{i,j}(t, \tau) = \frac{\partial \chi_i(t)}{\partial p_j(\tau)}
\]  

(8.8)

which is explained as follows,

\[
S_{i,j}(t, \tau) = \frac{\text{change in } (i-th) \text{ state at time } (t)}{\text{pert. on } (j-th) \text{ parameter at time } (\tau)}
\]  

(8.9)

where, \( p_j \) is the parameter with parameter index \( j \), \( t(p_j) \) is the period of the system. It is clear from the sensitivity function that the system robustness to parameter perturbations will increase with proportional decrease in the \( S \) value. The initial time \( t_0 \) is commonly considered to be the perturbation time \( \tau \) in local sensitivity analysis (Ihekwaba et al. 2005).

Thus, the differential equation for the sensitivity coefficients can be written without \( \tau \), as follows
Figure 7-11 Robust oscillation of proteins with CWO loop removed:
(a) In model A, (b) In model B, and (c) In model C.
\[
\frac{d}{dt} S(t) = \frac{\partial f}{\partial x} S(t) + \frac{\partial f}{\partial p};
\]

(9.0)

in which \( \frac{\partial f}{\partial x} \) is the two dimensional matrix of all the first order partial derivatives, \( \frac{\partial f}{\partial p} \) denotes the list of functions to be differentiated and \( \frac{\partial p}{\partial p} \) denotes the list of parameters.

In order to screen for more sensitive parameters, the general sensitivity rankings of parameters were generated in COPASI, directly from the sensitivity coefficients. To generate parameter rankings, as the parameter values might have huge ranges, standardised sensitivity values given by the following expression are used in COPASI:

\[
\bar{S}_{\text{ij}}(t) = \frac{\partial x_i(t)}{\partial p_j} \frac{p_j}{x_i(t)} = \frac{\partial \log x_i(t)}{\partial \log p_j}
\]

(9.1)

Thus, sensitive parameters were shortlisted from COPASI, sensitivities were simulated as shown in Eqn 9.3 and the calculated ratios are plotted in Figure 7-12. The simulation results showed that without the CWO negative feedback loop in all three models most of the parameters showed negligible sensitivity changes, but the sensitivities of the same six parameters (a43, a47, a57, a61, a63 and a67) related to \( \text{clk} \) and \( \text{per} \) gene components previously identified in robustness analysis yielded a maximum change of up to 80%.

In particular, the sensitivities of the transient rate of translation of \( \text{per} \) and \( \text{clk} \) mRNAs (a57 and a61) decreased more than 80% (for both \( \text{per} \) and \( \text{clk} \)) when removing the CWO feedback loop in model A and B. But the same parameter sensitivities increased by 22% and 36% (for \( \text{PER} \) and \( \text{CLK} \) respectively) in model C (Figure 7-13). Even though models A and B with respect to model C, gave opposite effects in their sensitivities to \( \text{PER} \) and \( \text{CLK} \) synthesis, one agreement they all had in common was that CWO loop removal affected the rhythmic oscillation of the two crucial clock components PER and CLK. From the sensitivity analysis results of all three models our personal judgement favours the results of the CWPT model, since CWO is known...
Figure 7-12 A 3D histogram of Local sensitivity results with CWO loop removed and normal: (a) In model A, (b) In model B, and (c) In model C.
Figure 7-13 Comparing Local sensitivity ratios with CWO loop removed.
to play a modulatory role in oscillation of clock proteins affecting their amplitude (Lim et al. 2007; Richier et al. 2008). This result of model C is in excellent agreement with the wet-lab arguments. From these results a probable function of the CWO feedback loop can be inferred. It is possible that CWO plays an important role in modulating CLK and PER protein feedback loops, thereby increasing the circadian clock robustness.

7.5 CWO mutants

The RT–PCR (Reverse transcription polymerase chain reaction) results of two cwo-deficient fly strains e4027 and f5073, which contain unique transposon insertions at the beginning and end of the first intron, show that both insertions reduce cwo mature mRNA levels below the level of detection, probably as a consequence of deficient splicing (Kadenar et al. 2007). Moreover, no CWO protein is detectable in the f5073 strain as shown in vitro (Matsumoto et al. 2007). These strains have strong circadian locomotor activity phenotypes. During the first 4 days in free running conditions (constant darkness, DD), more than half of these flies were arrhythmic (56% and 51% for strains 4027 and 5073, respectively); the remaining flies had weaker and longer rhythms than control strains, which have nearly 0 arrhythmic flies (Matsumoto et al. 2007). After 4 d in DD, most mutant flies were arrhythmic (75% or 100% for 5073 and 4027, respectively).

In a more recent quantitative RT – PCR study, a UAS (Upstream Activation Sequence) insertion next to a cwo reporter gene was expressed with GAL4 activator protein in GS10340 Drosophila cell lines. Next by using EMS (Ethyl Methane Sulphonate) induced mutagenesis, viable mutant candidates were selected and were named cwoB9 (Richier et al. 2008). Immunoreactivity performed on cwoB9, even with cwo overexpression, showed that it is negative for CWO protein antibodies. The cwo gene sequence in cwoB9 mutants showed that the mutation was a result of a shift in mRNA splicing caused by an insertion and deletion of 496 and 7 base pair sequences respectively in the exon 2 site of the coding region. The resultant cwoB9 was considered a null allele, since the mutant was viable and its cDNA still produced a non-functional CWO peptide sequence of 36 amino acids long (Richier et al. 2008).
Wet-lab experiments of the $cwo^B9$ mutants were carried out in both DD and LD conditions. A delayed activity rhythm was observed in LD and in DD. A lengthened rhythm with 26.5 hours period was observed with robust oscillation for at least 10 days. The mRNA amplitude peaks of core clock elements, $vri$, $tim$, $pdp1$ and $per$ were reduced in the mutant. Although the mRNA trough level at the start of the day was similar to control wild-type strains, the consecutive mRNA peak concentrations of $pdp1$, $vri$ and $per$ with the exception of $tim$ were considerably decreased compared to the wild-type. Based on these results in $cwo^B9$ mutant pacemaker neurons it was proposed that, oscillations of clock proteins in these mutants might linger on for at least several days (Richier et al. 2008).

7.5.1 Testing $cwo^B9$ null mutants

In our simulations, the $cwo^B9$ null mutant was tested using the same procedure as $clk^{jer}$, $per^{01}$ and $tim^{01}$ mutants by keeping the translation rate of $cwo$ mRNA as zero in model A and model B. Resulting simulation results showed a strong effect of CWO loss of function on all four CLK target genes in model A and model B (Figure 7-14). In model A $per$ and $tim$ transcript oscillations were greatly affected. Time evolution of $per$ and $tim$ mRNAs showed that peak mRNA levels were reduced from 0.42 nM to 0.2 nM and their trough levels were reduced from 0.24 nM to 0.08 Nm. Likewise In model B $per$ and $tim$ mRNA transcript peaks were reduced from 0.43 nM in the wildtype to 0.27 nM in the mutant; similarly their troughs showed a drastic reduction from 0.24 nM to 0.11 nM.

Test results showed that the amplitude of oscillation is vastly affected resulting in reduced midnight peaks compared to the wild type in DD (Figure 7-14). This result also concurs with the argument that the $cwo^B9$ null mutants are not completely arrhythmic like $per^{01}$, $tim^{01}$, $clk^{jer}$, and $cyc^{01}$ null mutants. Our simulations showed that lack of CWO in the system leads to low amplitude oscillations (Figure 7-15) and this confirms the sensitivity ratios estimated without CWO loops see section 7.4, where it was found that parameters related to $clk$ and $per$ genes were vastly affected by the removal of CWO from the system. It was proposed from experimental studies
Figure 7-14 Time evolution (72 h) of mRNAs cwo\textsuperscript{B9} mutants: (a) in model A, and (b) in model B.
Figure 7-15 Time evolution (10 d) of mRNAs \textit{cwo}^{B9} mutants: (a) In model A and (b) in model B.
that CWO protein is important to drive high amplitude oscillations of other *Drosophila* clock components (Lim et al. 2007), which concurs with our low amplitude simulated results. Based on these effects we can say with confidence that CWO plays a modulatory role in driving clock oscillations. A longer simulation of the cwo\(^{B9}\) null mutant in model A and B showed sustained weak rhythms with low amplitude and rhythmic oscillations for more than 10 days in DD. Our mutant tests also observed a strong CLK targets mRNA peak decrease of vri, tim, pdp1 and per (Figure 7-15), from which it is inferred that CWO has at least a weak transcriptional activator function. All these results are in good agreement with (Richier et al. 2008).

Next, we simulated the same null mutant cwo\(^{B9}\) in our third model (model C), which includes a post-translational component CWPT, using the same method as used in model A and B. Surprisingly, the mutant test results showed a positive effect on CLK target genes (Figure 7-16). In model C per and tim transcript oscillations were greatly perturbed. Time evolution of per and tim mRNAs showed that peak mRNA levels were increased from 0.35 nM in the wildtype to 0.42nM in null mutants and their trough levels also increased from 0.22 nM to 0.25 Nm. This shows that in model C the CWPT complex acts as a repressor of all clock genes having E-boxes. This result is completely opposite to models A and B, but it concurs with the sensitivity results (section 7.4), as the removal of the CWO loop increased the sensitivity of parameter a61 (translation rate of clk mRNA) by 36%. Increase in CLK protein levels leads to corresponding mRNA increase in CLK activated gene transcripts. As in models A and B, the mutant simulations show sustained low amplitude weak rhythms (Figure 7-16), but not arrhythmic (oscillation) for more than 10 days in DD.

The main difference between experimental and simulation results in all the three models (A, B and C) was the DD level of cwo mRNA in cwo\(^{B9}\) mutants. It was observed experimentally that CWO is a self-repressor of its own expression. Hence the lack of functional CWO protein in cwo\(^{B9}\) mutants leads to increased cwo mRNA concentration levels. But in all our three models the results were opposite to experimental results, due to the fact that only CWO binding to E-boxes is included, but not a repressor or activator activity. As seen in the previous section, the complete lack of functional CWO leads to a 2 – 3 hours’ time-lag in accumulation of unphosphorylated PER protein and lengthening of activity rhythms in *Drosophila*.
Figure 7-16 Time evolution of mRNAs $cwo^{B9}$ mutants in model C: (a) 72 h and (b) 10 d.
cwo$^{B9}$ mutants. But simulation of the mutant in models A, B and C showed no such time delay. The reason for this result might be the fact that we do not have a time delay equation for CWO in our model. Even in wet-lab findings, there is no information on why this period lengthening happens. So, due to the lack of a proper CWO functional role, we cannot include time delay functions or weights. It has to be noted that this lack of relevant data was the prime motivation for our CWPT model. A hypothetical cwo$^{CWPT}$ mutant in the CWPT model provides a probable reason for such delayed accumulation of PER unphosphorylated protein (see section 7.5.3).

**7.5.2 Testing cwo double knockout mutant**

CWO belongs to the basic helix-loop-helix-ORANGE (bHLH-O) transcription factor family. It is well known that in addition to cwo, 13 more basic helix-loop-helix-ORANGE family coding genes are present in the *Drosophila* genome (Matsumoto et al. 2009). In order to perform their transcriptional regulatory function, bHLH-O TFs form a homo or hetero dimer (Davis et al. 2001). Wet-lab experiments were carried out to screen for other potential bHLH-O candidate genes which can play a role in *Drosophila* circadian rhythm (Matsumoto et al. 2011). Promoter assays of E-box having promoters in individual bHLH-O gene knockout flies showed that, apart from CWO, the SIDE, Mβ and Mγ gene knockouts affected *Drosophila* locomotor activity. Further, to check whether any of these new candidates interact with CWO activity, a double knockout mutant study was carried out, in which among the four target genes any one was knocked out along with CWO (Matsumoto et al. 2011). It was found that the double knockout mutants exhibited a damped circadian locomotor activity with a slightly longer period. Among which the cwo–side double knockdown mutant showed a period lengthening of 1.1 – 1.5 hours. Similarly, cwo–mβ and cwo–mγ double knockouts also showed period lengthening of locomotor activity rhythms (Matsumoto et al. 2011).

As discussed in section 6.1.2, at the conceptual level to check the possible dimer hypothesis, we included a CWD hetero dimer complex with the binding of CWO and a hypothetical protein (HP) in model B. In a double knockout mutant there will be the total loss of functional CWO and the candidate bHLH-O proteins. Since we did not include any equations to govern transcriptional control of the HP, to simulate the
double knockout effect the cycling of CWD was totally removed from the model by fixing its transient concentration to zero. We tested the possible effects of the double knockout by simulating the mutant model. By measuring the time evolution of CLK protein levels in the simulation results we found a period lengthening of approximately 1.46 hours in the double knockout mutant compared to wild-type conditions (Figure 7-17). In wild-type simulations the rhythmic oscillation of CLK protein level peaked at 4.46 and 28.5 hours with a period length of 24.04, whereas, in the mutant simulation CLK protein concentrations peaked at 4.25 and 29.83 hours with a period length of 25.5 hours. This mutant result of model B is in good agreement with the observed in vivo results (Matsumoto et al. 2011).

7.5.3 Testing our model assumption in a cwo\textsuperscript{CWPT} hypothetical mutant

In order to test our major hypothesis that CWO plays a post-translational role by its interaction with PER and thereby impacting the DBT mediated phosphorylation of both PER and CLK, we tested the phosphorylation component CWPT in our model. It is known that DBT mediated phosphorylation of PER, transport of DBT to CLK by PER and the eventual DBT induced phosphorylation of CLK are necessary for Drosophila circadian locomotor activity rhythms (Yu et al. 2006). Even though we did not include the phosphorylation and dephosphorylation in detail at the level of DBT, PP1 and GSK3, we wanted to check what role a new CWPT component with CWO, PER and TIM interaction can play. So we simulated a hypothetical cwo\textsuperscript{CWPT} mutant in which the functional domain needed for CWO/PER/TIM (CWPT) interaction is expressed but not the DBT binding domain, resulting in rapid association of the CWPT complex and its subsequent degradation resulting in accumulation of hypophosphorylated PER and TIM proteins. Now while the DBT career role of CWO is vastly diminished, yet the DNA binding domain of this mutant is still intact, so CWO is free to bind to E-boxes in promoters. We named this hypothetical mutant cwo \textsuperscript{CWPT}. The nature of this mutant is such that, the DNA binding helix-loop-helix domain of CWO is still functional, but the peptide sequence required for DBT/CWO interaction is not there. Due to the nature of this mutant, there
Figure 7-17 Time evolution of CLK protein concentration in wildtype and cwo Double knockout mutant.
will be a considerable increase of the CWPT complex in the system. This mutant was simulated by increasing the association rate of CWPT complex to 1 and by reducing the degradation rate of CWPT complex to 0.025.

The resultant \( cwo^{CWPT} \) mutant was able to generate sustained low amplitude oscillations for up to 10 days (Figure 7-18) as observed by Richier et al. 2008. Simulation results (48 h in DD) of the mutant compared to wild-type showed high increase in \( cwo \) mRNA troughs from 0.19 nM to 0.23 nM (Figure 7-18) and the concentration peaks increased from 0.31 nM to 0.33 nM. However, mRNA transcripts of other CLK target genes were greatly reduced, and by the end of the fourth cycle after \( cwo^{CWPT} \) perturbation the oscillation can be seen dampening off considerably (figure 7-18).

In day 2 of DD the \( cwo \) mRNA transcript peaked at 0.37 nM. In the same cycle peak CLK target \( per \) and \( tim \) mRNA transcripts were considerably lower at 0.3 nM (Figure 7-19). \( cwo^{CWPT} \) mutant test results show a role reversal for CWO in the form of the CWPT component from that of \( cwo^{B9} \) mutant results (Figure 7-16). Based on \( cwo^{B9} \) mutant simulation in model C it was thought that the CWPT complex is a transcriptional repressor. In the \( cwo^{CWPT} \) mutant CWO still binds to all E-box containing promoters, but the resultant loss of CWPT in the system makes the model behave differently in two ways: 1. Increase in \( cwo \) mRNA levels, suggesting CWPT repression of \( cwo \) transcription 2. Decrease in \( pdp1, tim, per \) and \( vri \) mRNA levels, suggesting an activator role of CWPT in \( pdp1, tim, per \) and \( vri \) gene transcription.

It was observed \textit{in vivo} that unphosphorylated TIM and PER proteins are accumulated in cytoplasm with a 2 – 3 hours delay in \( cwo^{B9} \) mutants. PER, TIM and CLK proteins can oscillate in a dampened manner for up to 10 days. Our results concur in that PER and TIM proteins accumulate with a 2.6 h delay (figure 7-20) Furthermore, we were able to show that loss of CWPT in the system leads to considerable increase in \( cwo \) mRNA (figure 7-19). The test results showed damped protein oscillations for up to 10 d (figure 7-20) as mentioned in literature (Richier et al. 2008). Our test results of the hypothetical \( cwo^{CWPT} \) mutant lead to the conclusion that CWO can play the dual role of self-repressing its own expression and activating other CLK target clock gene promoters by reducing the amount of transcriptional
Figure 7-18 Oscillation (10 d) of mRNAs in cwo\textsuperscript{CWPT} mutant in model C
Figure 7-19 Time evolution (48 h) of mRNAs cwo\textsuperscript{CWPT} mutants in model C:
(a) cwo mRNA and (b) per mRNA.
Figure 7-20 Time evolution transcripts in $cwo^{CWPT}$ mutants in model C:
(a) PER protein oscillation and (b) Protein oscillations for up to 10 d.
repressor and DBT carrier PER. This result is fascinating, since there is no discovery of any such molecular mechanism existing in *Drosophila* clock neurons, confirms a modulatory role of CWO in the form of CWPT, and answers the ambiguity in molecular mechanism by which *cwo* mutants can exhibit both activator and repressor functions. We hypothesize that, CWO could directly repress its own E-box and indirectly activate CLK direct target genes by reducing the amount of phosphorylated PER.
Chapter 8: Conclusion

8.1 Contributions

This thesis contributes to the area of mathematical modelling of circadian rhythm in *Drosophila melanogaster*. Specifically, we introduce novel thinking and techniques to model the regulatory mechanism of circadian rhythm in *Drosophila melanogaster*. The primary objective of this thesis is to answer a probable role for CWO in order to explain the unique nature of this regulatory protein i.e., “how can CWO act both as repressor of its own gene expression, and yet act as an activator of other clock components especially when all these above mentioned gene promoters have the same E-boxes?”.

Our major contributions of the thesis are as follows,

1) We conceptualised three models A, B and C with 28, 30 and 29 ordinary differential equations having 69, 72 and 73 parameter values.

2) We developed three circadian models (A, B and C) with three different hypotheses for CWO activity.

3) We successfully estimated all the parameter values and proceeded to simulate the models.

4) We have successfully shown detailed modelling in the form of using implicit binding and unbinding probabilities and individual rates.

5) All three models satisfy the biological observations in terms of

   i) Indefinite oscillation of circadian protein and mRNA regulatory rhythms.
   ii) Model shows observed biology both in DD and LD conditions.
   iii) Entrainment by light
iv) *per*, *tim* and *clk* null mutants in all three models.

v) *cwo*\(^{B9}\) null mutant effects can be partially replicated in all three models.

6) We have shown that the *cwo* double knockout mutants can be successfully replicated using a CWD heterodimer in model B.

7) In model C we simulate a hypothetical mutant *cwo*\(^{CWPT}\) which not only replicates the wet-lab data but also explains the conflicting role of CWO protein unequivocally.

8) Overall, we show the usefulness of detailed modelling using probability-based mass action governed ODE’s not only in replicating observed effects, but also simulating unknown effects of the clock protein CWO.

The unique contribution of this research is that we have convincingly proved that CWO interacts with PER in a post-translational fashion resulting in the observed phenotypes biologically (Richier et al. 2008). Our model C shows that CWO in fact exists as an activator and a repressor at the same time. We established this fact without using any special function, but using the established biological facts and theories. This is a contribution towards establishing a whole new gamut of unique features in the cooperativity of CWO, and will be of a great significance to *Drosophila* circadian rhythm biologists. It is our wish that this hypothetical mutant could be tested *in vivo* in the near future thus, confirming our prediction.

### 8.2 Summary

We used the data findings from published research reports to build three probable models of the *Drosophila* circadian clock system.

All oscillatory core circadian clock components, including the new *cwo* gene and its products, were used to make three conceptual models. Three conceptual models A, B and C were converted into a set of 28, 30 and 29 ordinary differential equations
governed by mass action kinetics, in total with 69, 72 and 73 transient parameters for models A, B and C respectively. We simulated and analysed the models using the modelling tool COPASI (Complex PAthway Simulator). Simulated testing of our detailed models showed that, all major experimentally observed molecular oscillations can be replicated by all our three models (A, B and C). Rhythmic oscillation of clock components in both DD and LD settings, entrainment by light and all major mutations like \( \text{per}^S \) and \( \text{per}^L \) mutants, as well as \( \text{per} \), \( \text{tim} \), \( \text{clk} \) and \( \text{cwo} \) null mutants can be reproduced in models A, B and C. The sensitivity analysis of the new CWO feedback loop showed that, it plays a major role in maintaining the amplitude of clock component oscillations by affecting the parameters associated with \( \text{per} \) and \( \text{clk} \) mRNA and protein transcripts. In addition a hypothetical mutant \( \text{cwo}^{\text{CWPT}} \) was created in model C, and the mutant analysis showed valuable insights into how CWO can function as both activator and repressor.

8.3 Discussion and future directions

Among the three models designed in this study, models A and B were based on the assumption that regulation of gene expression primarily takes place by influencing the regulation of transcription. But in model C we showed that, the regulation of transcription can be achieved by incorporating a protein interaction complex (CWPT). We could reproduce and analyse a large number of phenotypes in our models. However, in order to understand the circadian system better, a more detailed model is needed with explicit E-box numbers and post-translational modifications involving individual phosphorylation and dephosphorylation compartments with kinase components such as CK2, DBT, PP1, PP2A, SGG (Sathyanarayanan et al. 2004). To build such a model, experimental data of kinases and stochasticity of E-box bindings should be reliable with some quantitative elements. Such a detailed model will be computationally complex but hypothetically sound and more biologically relevant (Martinek et al. 2001).

However, minimal models, such as those used in cyanobacterial clock research, are shown to be generically useful in decoding complex systems properties. For example, Elowitz et al. 2000 built a ‘repressilator’ in *Escherichia coli* to study oscillatory
dynamics. This study, and the accompanying models, examined the system for principles that enable robust oscillations. These included aspects of transcription, as well as translation, and protein and mRNA decay rates. A recent study revealed the existence of repressilator and ‘delayed negative feedback’ motifs in the transcriptional circuit of the mammalian circadian clock. Hence, we believe that simple models based on conservative assumptions can help in understanding any system better.

Even with various assumptions and simplifications, a major time consuming step while developing our models is parameter estimation. Usual practice in dealing with mathematical models is to estimate the parameters by trial and error until stable limit cycles are achieved (Hynne et al. 2001). Yet, we cannot guarantee whether our parameters are in biologically optimal parameter spaces without performing a global sensitivity analysis (Forger et al. 2003). For now, our major claim is that all our three models can simulate the observed biology to a maximum extent and that in particular our model C answers the ambiguity of CWO function meaningfully. It is our wish to develop a more elaborate model in the future, but without more sophisticated global algorithms for parameter estimation it will be a challenging task (Moles et al. 2003).

Circadian clocks are one of nature’s great mysteries. They are evolutionarily conserved and almost all organisms share a great deal of similarity. Drosophila falls among one of the well-researched model organisms with respect to circadian rhythms. Since circadian rhythms are evolutionarily conserved, it will be interesting to compare the clock in various systems like mammals, Drosophila and Arabidopsis. Due to this evolutionary similarity, it will be interesting to develop hypothetical models in well researched organisms and apply the model findings to similar homologous counterparts in other organisms. E.g. cwo mammalian homologues dec1 and dec2 act differently to the proposed cwo role in Drosophila. A good start for such an approach was provided in an earlier study, where a new circadian clock component was found through mathematical modelling in Arabidopsis, but the hypothesis was taken from mammal and Drosophila systems (Locke et al. 2005). Similarly, if our model hypothesis is proved to be right, it will expand the knowledge of mammalian circadian control as well, since the exact transcriptional regulation influenced by DEC1 and DEC2 transcriptional repressors is not known (Honma et al. 2002).
In conclusion, the mechanism of circadian rhythms in *Drosophila melanogaster* is controlled by a complex molecular mechanism consisting of both positive and negative feedback loops. There are only five known publications which have experimental data with respect to the new CWO feedback loop (Lim et al. 2007; Matsumoto et al. 2007; Kadener et al. 2007; Richer et al. 2008; Matsumoto et al. 2011). Research related to the CWO transcriptional repressor appears to have come to a dead end with no further studies published by any groups since 2011. With respect to *cwo* \( B9 \) mutants, more *in vitro* and *in vivo* research should happen in order to understand exactly how a same basic helix-loop-helix transcriptional factor can act as both activator and repressor if it is just an E-box binding repressor. This proposed role of CWO is unlike any other Hairy-Orange TFs. There is an enormous number of research papers discussing the important role post-translational effects play in the mechanism of circadian rhythms (Hardin et al. 2006). But even with obvious signs like accumulation of unphosphorylated PER in *cwo* \( B9 \) mutants, no study has been carried out to investigate any post translational interaction between CWO and PER, if it indeed happens. Our results strongly suggest towards that direction and we encourage further wet-lab experiments to investigate such interaction.

### 8.4 Concluding remarks

“*Sometimes the mathematics can be very simple. Useful mathematical biology research is not judged by mathematical standards but by different and no less demanding ones*”.

- Jim Murray (Murray, J.D 1993. Mathematical Biology)

In our thoughts, a good mathematical model of biological systems should be based on the following,

(i) A great appreciation for biology and a sound understanding of the research problem.

(ii) A mathematical representation which is biologically realistic.

(iii) Finding crucial qualitative solutions for the unknown biological phenomena.
Finally, a good model not only generates insights into the dynamic behaviour of a system, but also helps to make predictions about network behaviours in similar systems (Endy et al. 2001). Indeed, it is reasonable to assume that control mechanisms employed by the *Drosophila* circadian clock might also be effective in regulating other types of intracellular networks, such as the mammalian circadian clock. With this in mind, we use the insights gained from our analysis of the CWO anomaly to develop a new network regulatory rule that can be applied generally to any network system exhibiting control mechanisms similar to that of the *Drosophila* clock.

What these studies show convincingly is the power of systems approaches combining theoretical and quantitative models with experimental science. Although many of the biochemical and modelling details remain unresolved, systems approaches are providing information on mechanisms. Classical circadian principles such as synchronisation and periodicity are being addressed in a more meaningful way in this model than in others. That being said, it is important to point out some of the limitations. E.g. in computational models we can change their assumptions to fit the data. Moreover, most of these early models focus on *in vitro* dynamics and ignore the contributions from system level noise which is omnipresent during *in vivo* conditions.

Time evolution of circadian models (Figure 8-1) show that traditionally *Drosophila* circadian feedback loop models have been developed following biological discoveries, and almost all known feedback loops and endogenous rhythms are modelled in great fashion using various concepts in mathematics, but the contribution of these models to advancing an understanding and/or predicting actual biological phenomenon of *Drosophila* circadian system is not directly apparent. However, these biophysical modelling research will contribute to developing theoretical bases for systems biology in long term.
Figure 8-1 Time evolution of circadian models in *Drosophila melanogaster*.
Purely limit cycle models have been traditionally used to understand system level dynamics of circadian rhythms. But *in vivo*, the dynamics of the system is immensely complex. One of the biggest challenges in limit cycle models is parameter estimation. A major bottleneck of such models is the excessive computational and man hours needed to shift the unstable limit cycles to stable ones and produce oscillatory outputs. This step is usually by trial and error and takes months or years to complete and by the time a computational model is estimated to produce results based on observed biological findings, it becomes old and irrelevant due to the rapid pace of contemporary *in vitro* and *in vivo* research. It is necessary to accelerate the modelling studies at par with the speed of biological studies. So instead of estimating huge numbers of parameters for more and more complex models to produce limit cycles, at some point we should think about recycling earlier models which are robust and can produce rhythmic oscillations of all core clock components. Stable models like ours, can withstand the perturbations caused by adding any hypothetical equations and functions. Since, we used detailed regulatory steps, even by adding components, we believe new parameter sets can be estimated in relatively shorter time in our robust models. We intend to test more *in vivo* and *in vitro* hypotheses in our models and produce valuable suggestions to experimental biologists.

Hence, our suggestion is to take circadian modelling to the level of predicting new components and functions. Interdisciplinary studies are the need of the hour. It is entirely possible to predict new clock components and their roles by using a combination of bioinformatics and computational modelling. After all bioinformatics software and servers are developed from sophisticated algorithms and almost all are based on basic chemical kinetics and established mathematical theories. We have made an attempt in integrating bioinformatics screening and mathematical modelling to predict a molecular role for CWO in an attempt to answer *in vivo* research questions. We feel that this is a major achievement towards modelling *Drosophila* circadian rhythms in this fashion.
References


APPENDICES

APPENDIX A. Model B described by system of ODEs

A.1 Probabilities of TFs binding to an E box or a V/P box in promoters

\[
\frac{d \left( Pr_{\text{per}} \right)}{dt} = \left( 1 - Pr_{\text{per}} - Pr_{\text{cper}} \right) \times a_1 \times CC - Pr_{\text{per}} \times a_{18} \tag{9.2}
\]

\[
\frac{d \left( Pr_{\text{ct}} \right)}{dt} = \left( 1 - Pr_{\text{ct}} - Pr_{\text{ct}} \right) \times a_2 \times CC - Pr_{\text{ct}} \times a_{19} \tag{9.3}
\]

\[
\frac{d \left( Pr_{\text{dp}} \right)}{dt} = \left( 1 - Pr_{\text{dp}} - Pr_{\text{dp}} \right) \times a_3 \times CC - Pr_{\text{dp}} \times a_{20} \tag{9.4}
\]

\[
\frac{d \left( Pr_{\text{cv}} \right)}{dt} = \left( 1 - Pr_{\text{cv}} - Pr_{\text{cv}} \right) \times a_4 \times CC - Pr_{\text{cv}} \times a_{21} \tag{9.5}
\]

\[
\frac{d \left( Pr_{\text{cw}} \right)}{dt} = \left( 1 - Pr_{\text{cw}} - Pr_{\text{cw}} \right) \times a_5 \times CC - Pr_{\text{cw}} \times a_{22} \tag{9.6}
\]

\[
\frac{d \left( Pr_{\text{cwo}} \right)}{dt} = \left( 1 - Pr_{\text{cwo}} - Pr_{\text{cwo}} \right) \times a_6 \times CWD - Pr_{\text{cwo}} \times a_{25} \tag{9.7}
\]

\[
\frac{d \left( Pr_{\text{cuper}} \right)}{dt} = \left( 1 - Pr_{\text{cuper}} - Pr_{\text{cuper}} \right) \times a_7 \times CWD - Pr_{\text{cuper}} \times a_{26} \tag{9.8}
\]

\[
\frac{d \left( Pr_{\text{cwt}} \right)}{dt} = \left( 1 - Pr_{\text{cwt}} - Pr_{\text{cwt}} \right) \times a_8 \times CWD - Pr_{\text{cwt}} \times a_{27} \tag{9.9}
\]

\[
\frac{d \left( Pr_{\text{cwpdp}} \right)}{dt} = \left( 1 - Pr_{\text{cwpdp}} - Pr_{\text{cwpdp}} \right) \times a_9 \times CWD - Pr_{\text{cwpdp}} \times a_{28} \tag{10.0}
\]
\[
\frac{d(Pr_{cv})}{dt} = (1 - Pr_{cv} - Pr_{cvv}) \times a_{12} \times CWD - Pr_{cv} \times a_{29}
\]  \hspace{1cm} (10.1)

\[
\frac{d(Pr_{vc})}{dt} = (1 - Pr_{vc} - Pr_{pc}) \times a_7 \times VRI - Pr_{vc} \times a_{24}
\]  \hspace{1cm} (10.2)

\[
\frac{d(Pr_{pc})}{dt} = (1 - Pr_{pc} \times Pr_{pc}) \times a_6 \times PDP - Pr_{pc} \times a_{23}
\]  \hspace{1cm} (10.3)

A.2 Time evolution of \textit{per, tim, clk, vri, cwo} and \textit{pdp1} mRNA’s

\[
\frac{d(per_m)}{dt} = \left\{ \left[ \frac{a_{45}(CC) + a_{46}(CWD)}{(CC) + (CWD)} \right] \left[ 1 - (1 - Pr_{ba} - Pr_{bc})^{a_{39}} \right] + a_{56} \left( 1 - Pr_{ba} - Pr_{bc} \right)^{a_{30}} \right\} \times per_m - (a_{45} \times per_m)
\]  \hspace{1cm} (10.4)

\[
\frac{d(tim_m)}{dt} = \left\{ \left[ \frac{a_{44}(CC) + a_{45}(CWD)}{(CC) + (CWD)} \right] \left[ 1 - (1 - Pr_{ba} - Pr_{bc})^{a_{30}} \right] + a_{56} \left( 1 - Pr_{ba} - Pr_{bc} \right)^{a_{39}} \right\} \times tim_m - (a_{44} \times tim_m)
\]  \hspace{1cm} (10.5)

\[
\frac{d(pdp_m)}{dt} = \left\{ \left[ \frac{a_{43}(CC) + a_{44}(CWD)}{(CC) + (CWD)} \right] \left[ 1 - (1 - Pr_{ba} - Pr_{bc})^{a_{39}} \right] + a_{56} \left( 1 - Pr_{ba} - Pr_{bc} \right)^{a_{30}} \right\} \times pdp_m - (a_{43} \times pdp_m)
\]  \hspace{1cm} (10.6)

\[
\frac{d(vri_m)}{dt} = \left\{ \left[ \frac{a_{46}(CC) + a_{45}(CWD)}{(CC) + (CWD)} \right] \left[ 1 - (1 - Pr_{ba} - Pr_{bc})^{a_{30}} \right] + a_{56} \left( 1 - Pr_{ba} - Pr_{bc} \right)^{a_{39}} \right\} \times vri_m - (a_{46} \times vri_m)
\]  \hspace{1cm} (10.7)
\[ \frac{d(cwo_m)}{dt} = \left\{ \left\lbrack \frac{a_{46}(CC) + a_{54}(CWD)}{(CC) + (CWD)} \right\rbrack \left[ 1 - \left(Pr_{bc} - Pr_{hc}\right)^{a_{92}} \right] + a_{56} \left(Pr_{ba} - Pr_{hc}\right)^{a_{92}} \right\} \times cwo_m - (a_{68} \times cwo_m) \]  
(10.8)

\[ \frac{d(clk_m)}{dt} = \left\{ \left\lbrack \frac{a_{47}(PDP) + a_{55}(VRI)}{(PDP) + (VRI)} \right\rbrack \left[ 1 - \left(Pr_u - Pr_r\right) \right] + a_{55} \left(Pr_u - Pr_r\right) \right\} \times clk_m - (a_{67} \times clk_m) \]  
(10.9)

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

\[ \frac{d(PER)}{dt} = (a_{57} \times per_m) - (a_{94} \times PER \times TIM) + (a_{95} \times PT) - (a_{85} \times PER) \]  
(11.0)

\[ \frac{d(TIM)}{dt} = (a_{58} \times tim_m) - (a_{94} \times PER \times TIM) + (a_{95} \times PT) - (a_{86} \times TIM) \]  
(11.1)

\[ \frac{d(PDP)}{dt} = (a_{59} \times pdp_m) - (a_{71} \times PDP) \]  
(11.2)

\[ \frac{d(VRI)}{dt} = (a_{60} \times vri_m) - (a_{72} \times VRI) \]  
(11.3)

\[ \frac{d(CLK)}{dt} = (a_{61} \times clk_m) - (a_{35} \times CLK \times CYC) + (a_{39} \times CC) - (a_{73} \times CLK) \]  
(11.4)

\[ \frac{d(CWO)}{dt} = (a_{62} \times cwo_m) - (a_{37} \times CWO \times HP) + (a_{41} \times CWD) - (a_{74} \times CWO) \]  
(11.5)
A.4 Time evolution of PER/TIM, CWO/HP, CLK/CYC and PER/TIM/CLK/CYC complex

\[
\frac{d(PT)}{dt} = (a_{64} \times \text{PER} \times \text{TIM}) - (a_{65} \times PT) - (a_{38} \times PT \times CC) + (a_{42} \times CCPT) - (a_{103} \times PT) \quad (11.6)
\]

\[
\frac{d(CC)}{dt} = (a_{55} \times \text{CLK} \times \text{CYC}) - (a_{39} \times CC) - (a_{38} \times PT \times CC) + (a_{42} \times CCPT) - (a_{76} \times CC) \quad (11.7)
\]

\[
\frac{d(CCPT)}{dt} = (a_{38} \times PT \times CC) - (a_{42} \times CCPT) - (a_{78} \times CCPT) \quad (11.8)
\]

\[
\frac{d(CWD)}{dt} = (a_{37} \times \text{CWO} \times \text{HP}) - (a_{41} \times CWD) - (a_{77} \times CWD) \quad (11.9)
\]
APPENDIX B. Model C described by system of ODEs

B.1 Probabilities of TFs binding to an E box or a V/P box in promoters

\[
\frac{d (Pr_{cper})}{dt} = (1 - Pr_{cper} - Pr_{cwper}) \times a_1 \times CC - Pr_{cper} \times a_{18} \tag{12.0}
\]

\[
\frac{d (Pr_{ct})}{dt} = (1 - Pr_{ct} - Pr_{cw}) \times a_2 \times CC - Pr_{ct} \times a_{19} \tag{12.1}
\]

\[
\frac{d (Pr_{cpdp})}{dt} = (1 - Pr_{cpdp} - Pr_{cwcpdp}) \times a_3 \times CC - Pr_{cpdp} \times a_{20} \tag{12.2}
\]

\[
\frac{d (Pr_{cv})}{dt} = (1 - Pr_{cv} - Pr_{cw}) \times a_4 \times CC - Pr_{cv} \times a_{21} \tag{12.3}
\]

\[
\frac{d (Pr_{ccwo})}{dt} = (1 - Pr_{ccwo} - Pr_{cwccwo}) \times a_5 \times CC - Pr_{ccwo} \times a_{22} \tag{12.4}
\]

\[
\frac{d (Pr_{cw})}{dt} = (1 - Pr_{cw} - Pr_{cw}) \times a_6 \times CWO - Pr_{cw} \times a_{25} \tag{12.5}
\]

\[
\frac{d (Pr_{cwper})}{dt} = (1 - Pr_{cwper} - Pr_{cwper}) \times a_7 \times CWO - Pr_{cwper} \times a_{26} \tag{12.6}
\]

\[
\frac{d (Pr_{ct})}{dt} = (1 - Pr_{ct} - Pr_{cw}) \times a_8 \times CWO - Pr_{ct} \times a_{27} \tag{12.7}
\]

\[
\frac{d (Pr_{cwcpdp})}{dt} = (1 - Pr_{cwcpdp} - Pr_{cwcpdp}) \times a_9 \times CWO - Pr_{cwcpdp} \times a_{28} \tag{12.8}
\]
\[
\frac{d \left( Pr_{cwv} \right)}{dt} = \left(1 - Pr_{vc} - Pr_{cwv}\right) \times a_{12} \times CWO - Pr_{cwv} \times a_{29} \quad (12.9)
\]

\[
\frac{d \left( Pr_{pc} \right)}{dt} = \left(1 - Pr_{vc} - Pr_{pc}\right) \times a_{25} \times VRI - Pr_{pc} \times a_{24} \quad (13.0)
\]

\[
\frac{d \left( Pr_{pc} \right)}{dt} = \left(1 - Pr_{vc} - Pr_{pc}\right) \times a_{6} \times PDP - Pr_{pc} \times a_{23} \quad (13.1)
\]

**B.2 Time evolution of per, tim, clk, vri, cwo and pdp1 mRNA’s**

\[
\frac{d \left( per_a \right)}{dt} = \left(\left[ a_{43} (CC) + a_{49} (CWO) \right] \frac{1 - \left(1 - Pr_{nc} - Pr_{nc}\right)^{a_{49}}}{(CC) + (CWO)} \right) \times per_a \left( a_{49} \times per_a \right) \quad (13.2)
\]

\[
\frac{d \left( tim_a \right)}{dt} = \left(\left[ a_{45} (CC) + a_{50} (CWO) \right] \frac{1 - \left(1 - Pr_{nc} - Pr_{nc}\right)^{a_{50}}}{(CC) + (CWO)} \right) \times tim_a \left( a_{50} \times tim_a \right) \quad (13.3)
\]

\[
\frac{d \left( pdp_a \right)}{dt} = \left(\left[ a_{45} (CC) + a_{53} (CWO) \right] \frac{1 - \left(1 - Pr_{nc} - Pr_{nc}\right)^{a_{53}}}{(CC) + (CWO)} \right) \times pdp_a \left( a_{53} \times pdp_a \right) \quad (13.4)
\]

\[
\frac{d \left( vri_a \right)}{dt} = \left(\left[ a_{46} (CC) + a_{52} (CWO) \right] \frac{1 - \left(1 - Pr_{nc} - Pr_{nc}\right)^{a_{52}}}{(CC) + (CWO)} \right) \times vri_a \left( a_{52} \times vri_a \right) \quad (13.5)
\]
\[
\frac{d(cwo_m)}{dt} = \left( \left[ a_{48}(CC) + a_{54}(CWO) \right] + a_{56} \left( 1 - Pr_{ba} - Pr_{bc} \right)^{a_{52}} \right) \times cwo_m - \left( a_{69} \times cwo_m \right)
\] (13.6)

\[
\frac{d(clk_m)}{dt} = \left( \left[ a_{47}(PDP) + a_{53}(VRI) \right] + a_{55} \left( 1 - Pr_{ba} - Pr_{bc} \right) \right) \times clk_m - \left( a_{67} \times clk_m \right)
\] (13.7)

**B.3 Time evolution of PER, TIM, CLK, VRI, PDP1 and CWO proteins**

\[
\frac{d(PER)}{dt} = (a_{57} \times per_m) - (a_{64} \times PER \times TIM) + (a_{85} \times PT)
\]
\[-(a_{104} \times PER \times TIM \times CWO) + (a_{103} \times CWPT) - (a_{85} \times PER)
\] (13.8)

\[
\frac{d(TIM)}{dt} = (a_{58} \times tim_m) - (a_{64} \times PER \times TIM) + (a_{85} \times PT)
\]
\[-(a_{104} \times PER \times TIM \times CWO) + (a_{103} \times CWPT) - (a_{86} \times TIM)
\] (13.9)

\[
\frac{d(PDP)}{dt} = (a_{59} \times pdp_m) - (a_{71} \times PDP)
\] (14.0)

\[
\frac{d(VRI)}{dt} = (a_{60} \times vri_m) - (a_{72} \times VRI)
\] (14.1)

\[
\frac{d(CLK)}{dt} = (a_{61} \times clk_m) - (a_{35} \times CLK \times CYC) + (a_{39} \times CC) - (a_{73} \times CLK)
\] (14.2)
\[
\frac{d(CWO)}{dt} = (a_{62} \times cwo_m) - (a_{104} \times PER \times TIM \times CWO) + (a_{105} \times CWPT) - (a_{34} \times CWO)
\]  

(14.3)

**B.4 Time evolution of PER/TIM, CLK/CYC and PER/TIM/CLK/CYC complex**

\[
\frac{d(PT)}{dt} = (a_{44} \times PER \times TIM) - (a_{95} \times PT) - (a_{38} \times PT \times CC) + (a_{42} \times CCPT) - (a_{103} \times PT)
\]  

(14.4)

\[
\frac{d(CC)}{dt} = (a_{55} \times CLK \times CYC) - (a_{99} \times CC) - (a_{38} \times PT \times CC) + (a_{42} \times CCPT) - (a_{76} \times CC)
\]  

(14.5)

\[
\frac{d(CCPT)}{dt} = (a_{38} \times PT \times CC) - (a_{42} \times CCPT) - (a_{78} \times CCPT)
\]  

(14.6)

\[
\frac{d(CWPT)}{dt} = (a_{104} \times CWO \times PER \times TIM) - (a_{105} \times CWPT) - (a_{106} \times CWPT)
\]  

(14.7)
SUPPLEMENTARY MATERIALS

APPENDIX C, APPENDIX D, and APPENDIX E are provided as supplementary materials in a CD-ROM, which is attached to the back cover of this thesis. Content of the CD include.

Folder (APPENDIX C):

Files: Model A.cps
       Model A.xml
       Model A_Parameters.csv
       Model A_Species.csv

Folder (APPENDIX D):

Files: Model B.cps
       Model B.xml
       Model B_Parameters.csv
       Model B_Species.csv

Folder (APPENDIX E):

Files: Model C.cps
       Model C.xml
       Model C_Parameters.csv
       Model C_Species.csv

where,

.cps files are copasi files.
.xml codes are in SBML Level2 Version4 format.
.csv files are parameter and species values.

Please read the ‘README.txt’ for information about how to execute the .cps and .xml codes.