A Systems Biology Approach to the Arabidopsis Circadian Clock

by

James Locke

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Biology and Theory Group, Department of Physics

March 2006

THE UNIVERSITY OF
WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

I would first like to thank Matthew Turner and Andrew Millar for their excellent co-supervision of my project.

I really appreciate all the help of Andrew's Lab, in particular Anthony Hall, László Kozma-Bognár, Vicky Hibberd, and Megan Southern in answering my incessant questions about the Arabidopsis clock, and their willingness to carry out experiments based on my modelling.

Thanks to Ozgur Akman for keeping me sane in our Tuesday meetings, and to Nathan Sircombe, Chris Westbrook and James Leake for all their typographical help. Thanks also to all my other friends through out my Ph.D. I would like to give special thanks to Sarah, Ali, Chris and Mateo for their efforts in keepie uppie, and to Mark and Sonya for our regular lunch time escapes.

I really appreciate the incredible amount of help my Mum and Brother have given me over the years, and for their patience as I attempt to stay a student as long as possible. Thanks especially to Elizabeth, whose company and support helped make my Ph.D so enjoyable.

I would like to dedicate this thesis in memory of my Dad, David Locke, who I know would have read every page.

Declarations

This Thesis contains work based on the following published papers.

- 'Modelling genetic networks with noisy and varied data: The circadian clock in *Arabidopsis thaliana*', J. C. W. Locke, A. J. Millar and M. S. Turner, *J. Theor. Biology*, **234**, 383-493, (2005).
- 'Extension of a genetic network model by iterative experimentation and mathematical analysis', J C W Locke, M M Southern, L Kozma-Bognar, V Hibberd, P E Brown, M S Turner and A J Millar, *Mol. Syst. Biol*, **1**:13, (2005).

I hereby declare that this Thesis is my own work except where explicitly stated.
No part of this Thesis has been previously submitted for a degree at the University of Warwick, or any other university.

James C. W. Locke
March, 2006.

Abstract

Circadian clocks involve feedback loops that generate rhythmic expression of key genes. Molecular genetic studies in the higher plant *Arabidopsis thaliana* have revealed a complex clock network. We begin by modelling the first part of the *Arabidopsis* clock network to be identified, a transcriptional feedback loop comprising *TIMING OF CAB EXPRESSION 1 (TOC1)*, *LATE ELONGATED HYPOCOTYL (LHY)* and *CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)*. As for many biological systems, there are no experimental values for the parameters in our model, and the data available for parameter fitting is noisy and varied. To tackle this we construct a cost function, which quantifies the agreement between our model and various key experimental features. We then undertake a global search of parameter space, to test whether the proposed circuit can fit the experimental data. Our optimized solution for the *Arabidopsis* clock model is unable to account for significant experimental data. Thanks to our search of parameter space, we are able to interpret this as a failure of the network architecture.

We develop an extended clock model that is based upon a wider range of data and accurately predicts additional experimental results. The model comprises two interlocking feedback loops comparable to those identified experimentally in other circadian systems. We propose that each loop receives input signals from light, and that each loop includes a hypothetical component that had not been explicitly identified. Analysis of the model predicts the properties of these components, including an acute light induction at dawn that is rapidly repressed by LHY and CCA1. We find this unexpected regulation in RNA levels of the evening-expressed gene GIGANTEA (GI), supporting our proposed network and making GI a strong candidate for this component.

We go on to develop reduced models of the *Arabidopsis* clock to aid conceptual understanding, and add a further proposed feedback loop to develop a 3-loop model of the circadian clock. This 3-loop model is able to reproduce further key experimental data.

Abbreviations

bHlH	basic-helix-loop-helix
CAB	CHLOROPHYL A/B BINDING PROTEIN
CCA1	CIRCADIAN CLOCK ASSOCIATED 1
Col	Colombia
CPU	Central Processing Unit
CRY	CRYPTOCHROME
CT	Circadian Time
CYC	CYCLE
dCLK	dCLOCK
DD	Darkness
DNA	Deoxyribonucleic Acid
ELF3	EARLY FLOWERING 3
ES	Enzyme Substrate
FRQ	FREQUENCY
GI	GIGANTEA
h	Hours
LD	Light-Dark
Ler	Landsberg erecta
LHY	LATE ELONGATED HYPOCOTYL

LUC	LUCIFERASE
mRNA	Messenger Ribonucleic Acid
nM	Nanomolar
ODE	Ordinary Differential Equation
PDP1	PAR Domain Protein 1
PER	PERIOD
PHY	PHYTOCHROME
PIF3	PHYTOCHROME INTERACTION FACTOR 3
PRC	Phase Response Curve
PRR	PSEUDO RESPONSE REGULATOR
PTC	Phase Transition Curve
RNA	Ribonucleic Acid
SMP	Symmetric Multi Processor
TIM	TIMELESS
TOC1	THE TIMING OF CAB EXPRESSION 1
UBQ	UBIQUITIN
VRI	VRILLE
WC	White Collar
WCC	White Collar Complex
Ws	Wassilewskija
WT	Wild Type
ZTL	ZEITLUPE

Chapter 1

Introduction

1.1 Circadian clocks

Almost all organisms have evolved to co-ordinate their activities with the many changes in the environment caused by the earth's rotation. A circadian clock that generates biological rhythms with a period of approximately 24 hour (h) is found in most eukaryotes, and some prokaryotes, such as cyanobacteria [1]. Circadian rhythms allow the timing of diverse biological processes throughout the day/night cycle, conferring a selective advantage to the organism [2]. It might be important, for example, that a particular process occurs in anticipation of a light/dark (LD) transition. The clock is capable of sustained oscillations under constant environmental conditions and maintains synchrony with the environment by entraining to rhythmic cues of the day/night cycle, especially input signals from light and temperature. Another key feature of the clock is that it is temperature compensated. When an organism is placed at different temperatures within its physiological range, the period of the rhythm does not change significantly [3].

Molecular genetic studies in several model organisms over the last thirty years have revealed a common regulatory mechanism for the circadian clock, as I will now review.

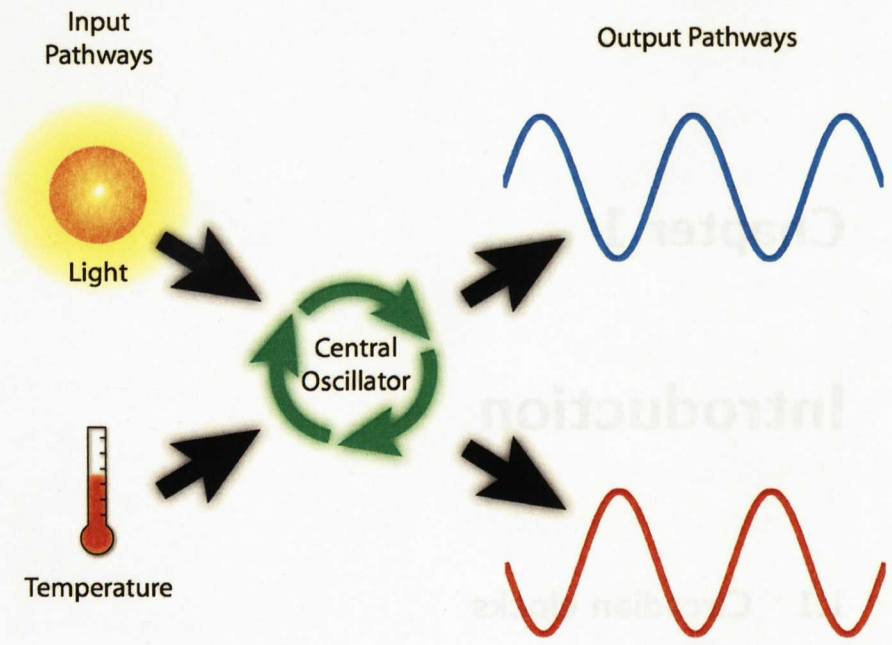


Figure 1.1: The circadian clock classically has been split into three basic components [4] - the input pathways, the central oscillator and the output pathways.

1.1.1 Molecular mechanism of circadian clocks

The circadian clock has classically been split into three basic components [4] - the input pathways, the central oscillator and the output pathways (Figure 1.1). The input pathways allow the entrainment of the oscillator to match the environmental cycle by providing information from stimuli such as light and temperature. For example, light can reset the clock either by inducing the transcription of a component of the core oscillator, or by promoting its degradation. Circadian clocks can be phase shifted by even brief light pulses, depending on the time of the pulse, which will be discussed later in terms of Phase Response Curve experiments (PRCs). Figure 1.2 shows a typical expression profile of a clock component whose transcription is activated by light, causing its rhythm to be entrained to the LD cycle.

A shared feature of the central clock mechanism is that the rhythms appear to

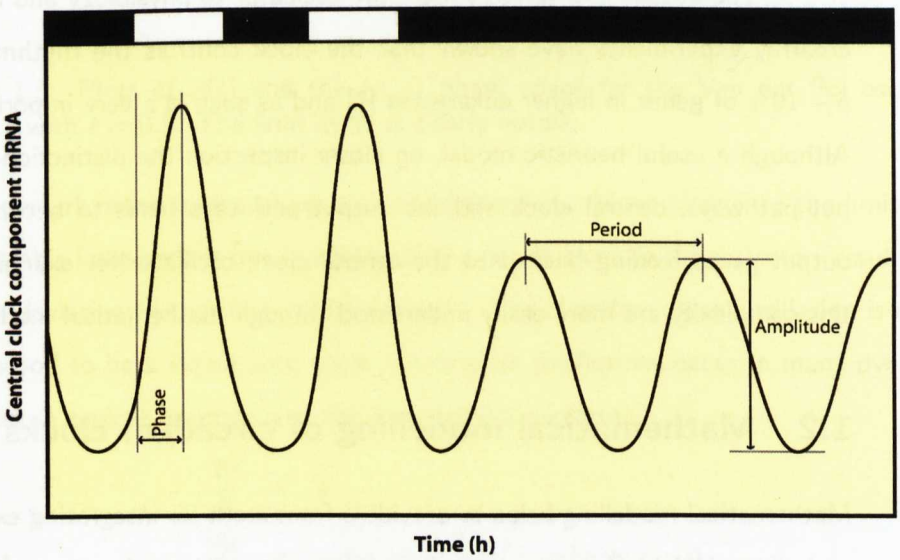


Figure 1.2: Expression profile of a typical clock component. White boxes represent periods of light, black boxes periods of darkness. Two light-dark (LD) cycles are shown before the oscillator goes into constant darkness. The oscillator is entrained to the LD cycle before free-running with a period slightly longer than 24h in constant darkness.

be generated by the interactions of a core set of around 6-12 rhythmically expressed genes that form positive and negative feedback loops [1]. In principle, one negative feedback loop in which a gene encodes a protein that, after several hours, turns off its own transcription is capable of creating a circadian rhythm. However, real circadian clocks have proven to be more complicated than this, with interlocked feedback loops [5, 6]. This network motif has evolved several times as there is little homology between the components of the fungal, plant, and mammalian clocks.

The output pathways connect the oscillator to physiology and metabolism. Microarray experiments have shown that the clock controls the rhythmic expression of $5-10\%$ of genes in higher eukaryotes [7] and as such is a very important mechanism. Although a useful heuristic model, on closer inspection the distinction between the input pathways, central clock and the output pathways tends to become blurred, with output genes feeding back onto the central clock oscillator for example. Networks of this complexity are more easily understood through mathematical modelling.

1.2 Mathematical modelling of circadian clocks

Mathematical modelling helps to provide a framework for integrating experimental data and allows insights into the behaviour of complex biological systems [8, 9]. Biological experiments often lead to conceptual models, which describe the major elements thought to be involved in a process. Converting these conceptual models in to mathematical models allows us to critically test our ideas, by giving our conceptual model more rigid detail. In studying these models, we can test whether they truly represent the observed phenomena, and propose new experiments.

Modelling has long been a part of circadian research, even before the molecular mechanism of the clock was known. By treating the clock as a limit cycle oscillator, mathematical models can be used to understand the clock's entrainment properties [10, 11]. A limit cycle is a periodic, cyclic trajectory around a fixed point in the phase-space of a system and may be stable or unstable. A stable limit cycle is one in which

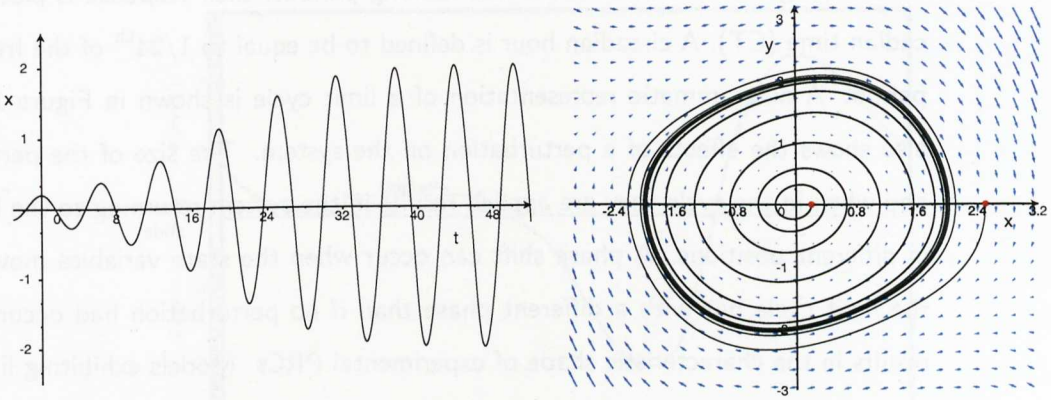


Figure 1.3: Plots of $x(t)$ and the (x, y) phase space for the Van der Pol oscillator, Eqn.1.1 with $\epsilon = 0.2$. The limit cycle is clearly visible.

slight perturbations to the phase-space trajectory are driven back into the limit cycle. An unstable limit cycle is one in which slight perturbations to the trajectory result in the system deviating further from the limit cycle. The circadian clock oscillation is usually understood to be a stable limit cycle. Limit cycle oscillations occur in many dynamical systems. A typical example is in the Van der Pol oscillator,

$$\begin{aligned} \frac{dx}{dt} &= y \\ \frac{dy}{dt} &= -x + \epsilon y\left(1 - x^2\right) \quad \text{where } \epsilon > 0 \end{aligned} \tag{1.1}$$

which is capable of generating a stable limit cycle (see Figure 1.3) that gives a reasonable imitation of a circadian oscillator in darkness [11, 12]. Here x and y can be treated as proxies for biological components of the clock. A modified version of the Van der Pol oscillator has been used to great effect to understand human circadian clock properties, in particular how human sleep/wake cycles are affected by light [13, 14].

Models exhibiting limit cycles can be used to analyse and understand phase response curves (PRCs) [15]. PRC experiments have been an invaluable tool in understanding entrainment in circadian rhythms. They produce a plot of phase shifts of a circadian rhythm against the circadian phase that a stimulus is given (see Figure 1.4 for a description of Type 0 and Type 1 PRCs). In order to be able to compare PRCs of

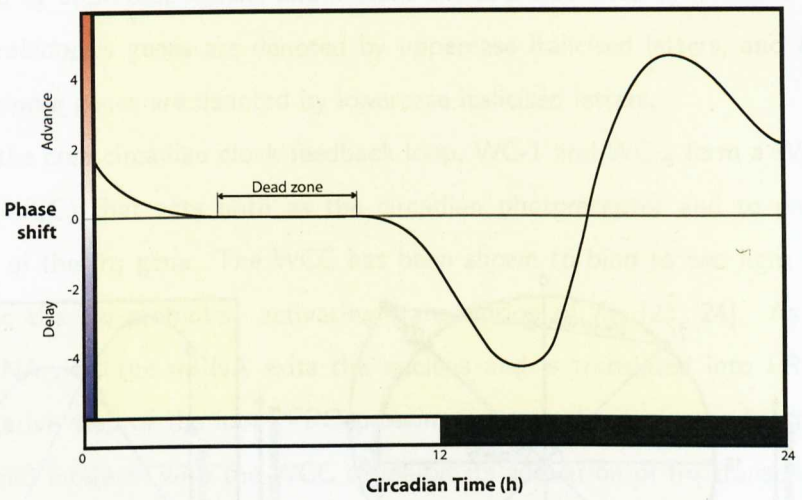
different organisms with different free-running periods, their response is plotted in circadian time (CT). A circadian hour is defined to be equal to $1/24^{\text{th}}$ of the free-running period. A diagrammatic representation of a limit cycle is shown in Figure 1.5, which also shows the effects of a perturbation on the system. The size of the perturbations and when in the cycle they are applied results in the system returning to the limit cycle in different positions. A phase shift can occur when the state variables move back to the limit cycle but with a different phase than if no perturbation had occurred. This results in the characteristic shape of experimental PRCs. Models exhibiting limit-cycles also suggest that stimuli applied at exactly the right phase and strength can push the oscillation towards its fixed point, or singularity, causing arrhythmia. Arthur Winfree was able to invoke arrhythmicity in fruit flies, by timing his stimuli from the predictions of his limit-cycle model [16].

The mathematical study of negative feedback in biological processes goes back to Goodwin's landmark work [17], where he shows that a relatively simple 3 variable system can be used to describe an enzymatic negative feedback loop. This work anticipated many experimental findings for circadian clocks, and slightly modified versions of the Goodwin oscillator have been used effectively to model the circadian clock in the fungus *Neurospora* [18, 19, 20, 21].

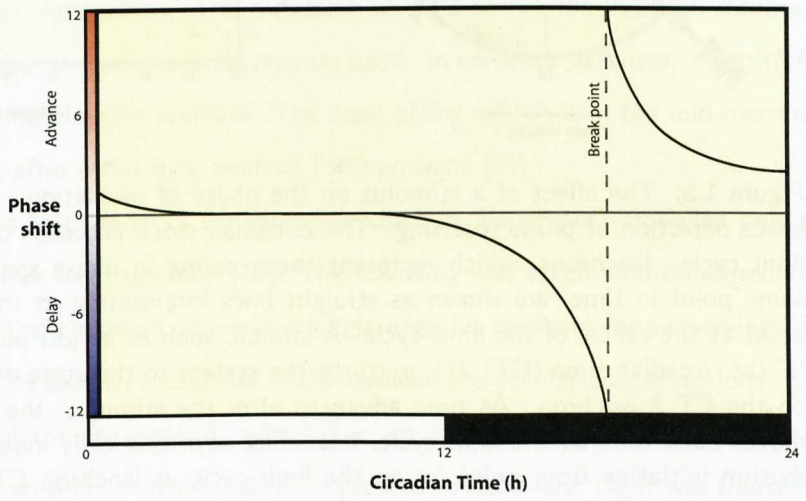
1.2.1 Modelling the Neurospora clock

Studies on the fungus *Neurospora crassa* have helped to understand many of the basic mechanisms that underlie circadian rhythms, including negative feedback and light and temperature entrainment, which are common to all clock systems. The *Neurospora* clock controls several rhythmic processes, the daily production of asexual conidiospores being the most frequently assayed.

The *Neurospora* circadian clock is based on an autoregulatory negative-feedback loop involving three proteins, the FREQUENCY protein, (FRQ), and the WHITE COLLAR proteins WC-1 and WC-2, (reviewed in Ref. [22]). Throughout this thesis, proteins



(a)



(b)

Figure 1.4: Typical phase response curves (PRCs). A PRC is a plot of phase shifts of a circadian rhythm as a function of the circadian phase that the stimulus is given. (a) A Type I PRC; in this case, a stimulus during the subjective day has little effect on phase (the dead zone). A stimulus applied early at night moves the phase back to the previous day (a phase delay) and a stimulus applied late at night moves the phase forwards into the next day (a phase advance). Responces are relatively small and there is a continuous transition between delays and advances. (b) A Type 0 PRC; in this case, the phase response is large and the transition between delays and advances is discontinuous. Often the strength of the stimulus determines which type of PRC is observed.

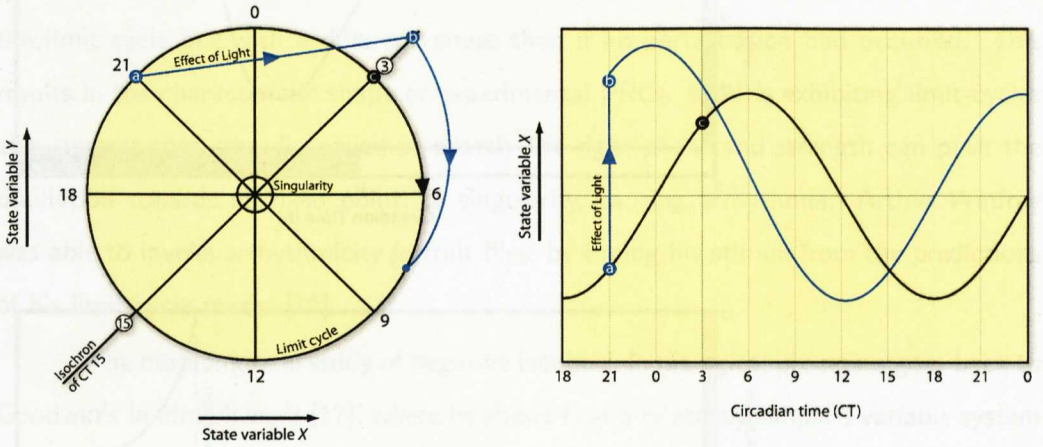


Figure 1.5: The effect of a stimulus on the phase of oscillation. Left Panel: (X, Y)-space depiction of phase resetting. The circadian clock proceeds clockwise around the limit cycle. Isochrons, which represent those points in phase space which are at the same point in time, are shown as straight lines intersecting at the fixed, or singular, point at the centre of the limit-cycle. A stimuli, such as a light pulse, applied at point 'a' (at circadian time (CT) 21), perturbs the system to the state described by point 'b' on the CT 3 isochron. As time advances after the stimulus, the perturbed oscillator moves back towards the limit-cycle, becoming asymptotically indistinguishable from a rhythm initiating from point 'c' on the limit-cycle at isochron CT 3. Because CT 3 may be considered as CT 27, the light pulse has advanced the phase of the perturbed oscillator by $27 - 21 = 6\text{h}$. Right Panel: Plot of state variable X against time. The blue line represents the phase shifted oscillation, the black line represents the unperturbed oscillator. (Figure modified from Ref. [1]).

are denoted by uppercase letters, and mutant alleles are denoted by italicised lower case letters. *Arabidopsis* genes are denoted by uppercase italicised letters, and *Drosophila* and *Neurospora* genes are denoted by lowercase italicised letters.

In the core circadian clock feedback loop, WC-1 and WC-2 form a White Collar Complex (WCC) that acts both as the circadian photoreceptor and to promote the expression of the *frq* gene. The WCC has been shown to bind to two light responsive elements in the *frq* promoter, activating transcription of *frq* [23, 24]. As the levels of *frq* mRNA rise, the mRNA exits the nucleus and is translated into FRQ protein. In the negative step of the loop, FRQ protein re-enters the nucleus a few hours after synthesis and interacts with the WCC to inhibit its activation of frq transcription [25]. As a result the amount of FRQ protein is reduced, which in turn results in less FRQ returning to the nucleus. The reduction in FRQ allows the WCC to resume activating *frq* transcription, and the cycle repeats itself. In constant darkness, *frq* mRNA and FRQ protein concentrations oscillate. The peak of *frq* mRNA is in the mid-morning [25] and is followed after 4-6 h by a peak of FRQ proteins [26].

The first model of the circadian clock in *Neurospora* modelled only the central FRQ negative feedback loop [18]. The following was taken as the mathematical model for the central circadian network: a single negative feedback loop involving FRQ protein repressing its own transcription. This involves the cellular concentrations $c_F^{(j)}(t)$ of the products of the *frq* gene F, where $j = m, c, n$ denotes that it is the corresponding mRNA, or protein in the cytoplasm or nucleus respectively. Light was treated as directly activating *frq* transcription, and the WC proteins were not included.

$$\frac{dc_F^{(m)}}{dt} = \frac{n_1 {g_1}^a}{{g_1}^a + {c_F^{(n)}}^a} - \frac{m_1 c_F^{(m)}}{k_1 + c_F^{(m)}} \tag{1.2}$$

$$\frac{dc_F^{(c)}}{dt} = p_1 c_F^{(m)} - r_1 c_F^{(c)} + r_2 c_F^{(n)} - \frac{m_2 c_F^{(c)}}{k_2 + c_F^{(c)}} \tag{1.3}$$

$$\frac{dc_F^{(n)}}{dt} = r_1 c_F^{(c)} - r_2 c_F^{(n)} \tag{1.4}$$

Here the various rate constants parameterise transcription (n_k, g_k), degradation (m_k,

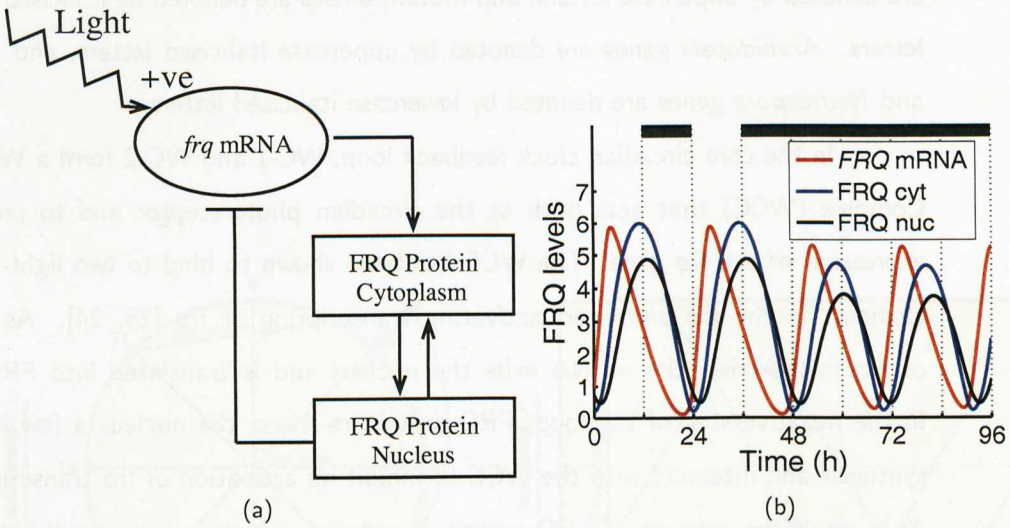


Figure 1.6: (a) Network diagram for *Neurospora* clock. Transcription of *frq* mRNA is light activated and is translated into protein in the cytoplasm. FRQ protein is then transported into the nucleus where it represses *frq* transcription. (b) Output of clock model simulated for 48h under light:dark cycles and 48h in constant dark. Parameter values are those used in [18]. Parameter values are $n_1 = 1.6\text{nMh}^{-1}$, $m_1 = 0.505\text{nMh}^{-1}$, $p_1 = 0.5\text{h}^{-1}$, $r_1 = 0.5\text{h}^{-1}$, $r_2 = 0.6\text{h}^{-1}$, $g_1 = 1\text{nM}$, $k_1 = 0.5\text{nM}$, $k_2 = 0.13\text{nM}$, $a = 4$. Parameter n_1 is increased by 25% under light conditions as *frq* transcription is light activated.

k_k), translation, (p_k), and the nuclear $\leftrightarrow$ cytoplasmic protein transport rate, (r_k). This model is closely related to the minimal model originally proposed by Goodwin [17]. Transcription is modelled by a saturated Hill function and degradation is given by Michaelis-Menten terms. For a full review of Michaelis-Menten kinetics and Hill functions please see Appendix A. In brief, Michaelis-Menten kinetics describe enzyme mediated reactions where the substrate is in much greater abundance than the enzyme, resulting in a saturated output. The Hill function represents a cooperative enzyme mediated reaction where the Hill coefficient (a) can be thought of as a measure of the cooperativity [27].

In order for the numerical solution of Eqns. 1.2 to 1.4 to give a stable limit cycle with a period of around 24h the equation parameters must be chosen to give enough non-linearity in the system to generate an oscillation, and a large enough delay

between each component [28]. The main difference between this model and that of the Goodwin model is that the degradation of the protein is described using Michaelis-Menten kinetics, rather than a linear term. In the Goodwin model for the system to exhibit sustained oscillation an unreasonably high Hill coefficient ($a > 9$) is required. This would represent a FRQ protein needing to form a complex with eight other FRQ proteins before being able to repress *frq* transcription - there is no experimental evidence for such high levels of cooperativity. Non-linear degradation of the sort set out in Eqns. 1.2 to 1.4 allows a more reasonable value for the Hill coefficient of 4. Solving numerically Eqns. 1.2 to 1.4 shows that the model consisting of a single negative feedback loop with a delay can generate circadian rhythms which look qualitatively similar to experimental data (Figure 1.6). The model successfully simulates light regulation, and entrainment of *frq* mRNA and protein levels. The model also predicts that oscillations of *frq* mRNA should continue in constant light and could be forced into chaos in LD cycles, but these phenomena contradict experimental observations [18].

Recent experimental studies have shown the *Neurospora* clock to be more complex than just one negative feedback loop. A second, positive feedback loop has been discovered involving the enhancement of WC-1 synthesis by FRQ. Following its induction, FRQ protein indirectly inhibits the WCC to shut down its own expression, but also has a positive post-transcriptional effect on WC-1 levels [5]. Mathematical modelling has been used to try and understand the properties of this network structure. This second feedback loop was shown not to be necessary for oscillations [29], but a more detailed model, using mass action kinetics without the assumption of Michaelis-Menten kinetics, has shown that the second feedback loop contributes significantly to the robustness of the oscillator period against parameter variation [30]. It is an open question why circadian systems have evolved a more complex structure. Recent mathematical studies propose that interlocked feedback loops increase the flexibility of regulation during evolution [31] and enhance precision [32].

1.2.2 Modelling the Drosophila Clock

Many models of the *Drosophila* clock are still based on the Goodwin model, using Hill functions to model the transcription and Michaelis-Menten kinetics to describe degradation. The basic mechanism for the *Drosophila* circadian clock is as follows; PERIOD (PER) and TIMELESS (TIM) proteins form a heterodimer (PER/TIM), which is then transported into the nucleus where it inhibits the transcription of PER and TIM [33, 34]. The repression is carried out by the PER/TIM heterodimer binding to the transcription factor, a heterodimer between dCLOCK (dCLK) and CYCLE (CYC), dCLK/CYC (Reviewed in [35]). Light entrains the clock by the rapid degradation of TIM [36]. PER protein and *per* mRNA fluctuate with a 24h period, with protein levels lagging by 4-6h, similar to the delay in FRQ expression seen in *Neurospora*. PER is modified by kinases and phosphatases, including the DOUBLE-TIME (DBT) protein [37], which appears to be expressed at a constant level, and post translational regulation plays a large role in the correct running of the clock.

This one loop mechanism has been investigated [18, 38], and has been shown to be capable of creating a 24h rhythm that can be entrained by light. The effect of phosphorylation of PER and TIM proteins was also investigated. Phosphorylation was proposed to be a method of creating a delay, by requiring the PER protein to be phosphorylated twice before entering the nucleus [18] and also as a method of introducing positive feedback [38]. In Ref [38] PER proteins were assumed to be heavily phosphorylated by DBT, whereas PER-TIM dimers were not. Therefore, as PER expression increases, an ever increasing amount of PER is dimerised and protected from DBT, giving a source of non-linearity as is required for sustained oscillations.

Recently, two additional feedback loops were identified as parts of the core circadian pacemaker in *Drosophila*. In one of the loops, VRILLE (VRI), a protein which is activated by dCLK/CYC was found to repress transcription of the *dClk* gene forming a negative feedback loop [6]. This second feedback loop has been modelled [39] and shown to produce more robust rhythms in relation to parameter changes than would be

the case with a one loop model. Another, positive, feedback loop has also been discovered in which the protein PAR Domain Protein 1 (PDP1) activates the transcription of *dClk*, which itself activates the transcription of *PDP1* [6]. Two recent studies have investigated the possible roles of these three positive and negative feedback loops in the generation and stabilisation of oscillations in *Drosophila* [40, 41]. The three loop model was able to explain the effect of altering gene dosage of *per* and *tim* on the period of the clock, and made several further predictions about the structure of the network that have yet to be confirmed experimentally.

1.2.3 Other Clock Models

It is possible to group processes, such as protein translation and transport in and out of the nucleus, into an explicit time delay [42, 43, 44]. Time delay models have the advantage of reducing the number of equations and the number of parameter values, and some people find them easier to understand conceptually. However, it is not possible to use these models to generate bifurcation diagrams and they are considered to be only an approximate representation of the underlying biochemical network. A further sub-set of models involve stochastic simulations of the clock. It has been suggested that noise in gene expression may severely affect the robustness of the circadian clock [45], although recent work suggests that noise is not crucial, and that the approach of using deterministic ordinary differential equations (ODEs) as opposed to stochastic DEs is often rather reliable [46].

The mouse circadian clock has also been studied [47, 48]. Ref [48] is of note, as an attempt to estimate the parameters in the model was made. Much molecular data aims to identify components of the network and to define connections between them. Generally there is little or no biochemical data for the numerous parameters that control the circadian network. Apart from Ref [48], these parameters have been chosen by hand in previous studies. Such an approach becomes more and more time consuming, and potentially unreliable, as the number of components in the model increases. This

opens up the very real possibility that apparent deficiencies in existing models may not be caused by an incomplete experimental understanding, or model, but rather by non-optimum choices of the parameters. The optimisation process used in [48] represents a step forward, as parameters were optimised by performing a least squares fit to experimental data. However, several key parameters such as nuclear export were fixed to be the same for all components of the model, and the starting choice of parameters before optimisation was not given. A major aim of this thesis is to develop a global parameter optimisation scheme that works well with noisy and varied experimental data, and to use this scheme to model the plant circadian clock.

1.3 Rhythmic Behaviour in Plants

One of the first experimental studies of endogenous biological rhythms was carried out in plants over two hundred and seventy years ago [49]. A plant (mimosa) that was already known to fold up its leaves at night and reopen them during the day was placed in constant darkness. It was found that the leaf opening and closing persisted just as if the plant was seeing the day and night. With this came the first experimental evidence for the persistence of an endogenous rhythmicity in the absence of environmental cues.

The circadian clock is now known to control many aspects of plant physiology, including leaf movement, flowering time, and photosynthesis. The clock is also used to measure photoperiod (day-length), allowing the detection of seasonal changes [50]. Microarray experiments in the model plant *Arabidopsis thaliana* [51] suggest that mRNA transcripts from at least 6 percent of the genome, over a thousand genes, are under the control of the clock. A recent study has suggested that as much as 15 percent of the genome is under the control of the clock [52].

There is strong evidence that the circadian clock in plants gives a competitive advantage. In wild type and long and short circadian period mutants of *Arabidopsis*, plants with a clock period matched to the environment contain more chlorophyll, fix more carbon, grow faster, and survive better than plants with circadian periods differing

from their environment [2]. It appears that the circadian clock allows the plant to be prepared for predictable environmental changes, and thus make the most of the resources available.

1.4 The Molecular Mechanism of the Arabidopsis Circadian Clock

Arabidopsis thaliana has many advantages for genome analysis, including a short generation time, small size and relatively small nuclear genome [53]. Its genetic structure is also representative of the world's major crop plants including soybeans, corn, and wheat. As such, it is an important model system for identifying genes and their functions. Several backgrounds of *Arabidopsis*, such as Landsberg *erecta* (Ler), Wassilewskija (Ws) and Colombia (Col) have been treated as wild type (WT), i.e. non-mutant, plants with which to experiment. Experimental work on *Arabidopsis* has revealed several aspects of the circadian clock in plants.

In *Arabidopsis*, most of the characterised input pathways to the clock involve light. Photoreceptors *PHYTOCHROME (PHY)* A, B, D and E and *CRYPTOCHROME (CRY)* 1 and 2 have all been shown to be inputs to the clock (Reviewed in [54]). Mutant studies have shown that PHY A, B, D,and E mediate signals from red light to the clock in an additive fashion, and CRY 1 and 2 from blue. The signalling pathways from the photoreceptors to the circadian clock are believed to be very complex, (reviewed in Ref [55]), and the connections between the photoreceptors and the clock are not well characterised. *PHYTOCHROME INTERACTION FACTOR 3 (PIF3)*, a basic-helix-loop-helix (bHlH) transcription factor, was proposed to play a role in the entrainment of the *Arabidopsis* clock. PIF3 was shown to interact with *PHYA* and *PHYB* and to bind to the promoter of two proposed central clock genes, *LATE ELONGATED HYPOCOTYL (LHY)* and *CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)* [56]. However, recent studies have shown that overexpression or lack of biologically functional PIF3 does not affect

period length of rhythmic gene expression or red-light-induced resetting of the circadian clock [57, 58]. Nonetheless, there are several homologues of PIF3, which may yet play a role in mediating light signals to the clock [59].

Molecular components in the central clock of *Arabidopsis* were first identified using plants expressing the promoter of the clock controlled output gene *CHLOROPHYLL A/B BINDING PROTEIN (CAB)* fused to the firefly *LUCIFERASE (LUC)* reporter gene. The *CAB* promoter drives a robust rhythm of bioluminescence that can be monitored with a photon-counting video camera [60]. Mutagenising these transgenic plants followed by bioluminescence analysis of abnormal *CAB* rhythms resulted in the isolation of a number of circadian period mutants. *THE TIMING OF CAB EXPRESSION 1 (TOC1)* was the first gene to be found using this method and now, using this and other techniques, several genes have been found to play a role in the clock (reviewed in [61]). The first proposed mechanism for the central clock in *Arabidopsis* is based on a single positive and negative feedback loop involving *TOC1*, and two redundant genes *LHY* and *CCA1*. We describe these genes, and the proposed network, below.

1.4.1 TOC1

A mutant with reduced *TOC1* function (caused by the semi-dominant allele *toc1-1*) displayed a period 3h shorter than wild-type, not only for *CAB::LUC* expression but also for circadian outputs such as leaf movement [60]. The *toc1-2* allele, which is fully recessive and is believed to cause a complete loss of function, caused arrhythmia in certain conditions including constant darkness and red light [62] (Figure 1.7 shows *LHY* and *CCA1* mRNA levels in a *toc1-2* background). Moderate over expression of TOC1 lengthens the free running period [62], and strong, constitutive overexpression abolishes rhythmic expression of all genes examined [59]. This suggests that *TOC1* is an important component of the core of the oscillator, rather than part of the light input pathway to the clock [63].

TOC1 mRNA levels have a broad peak at dusk, with slowly decreasing levels

through the night [64, 63]. The only evidence for TOC1 protein levels is in an overexpressing background [65]. TOC1 protein levels peak in the middle of the night. There is also evidence that TOC1 protein is degraded by *ZEITLUPE (ZTL)*, as in a *ztl-1* mutant there is constitutive levels of TOC1 protein expression [65].

1.4.2 LHY

LHY was first identified in a late-flowering mutant screen, and only later characterised as a clock component [66]. Overexpression of *LHY* causes complete arrhythmia in clock controlled genes (Figure 1.7), and a *lhy* loss of function mutation causes a 1-2 h period shortening [64, 67]. *LHY* mRNA levels oscillate, peaking in the morning, and have shown to be light activated. The LHY protein has been shown to bind DNA via MYB domains, and has significant similarity to CCA1. DE-ETIOLATED 1 (DET1) has been shown to be involved in the post-transcriptional modification of LHY, by regulating LHY's proteasomal degradation [68].

1.4.3 CCA1

CCA1 was first identified as a binding factor of the promoter of *CAB*, and was later shown to have very similar properties to that of *LHY*. Overexpression of *CCA1* causes complete arrhythmia in clock controlled genes, and a *cca1* loss of function mutation causes a 1-2 h period shortening [69, 67]. CCA1 also has a similar expression profile to that of *LHY*.

CCA1 has been shown to be phosphorylated *in vitro* by the protein kinase CASEIN KINASE 2 (CK2) [70]. Over expression of CK2 leads to short period oscillation demonstrating that CCA1 phosphorylation by CK2 is important for the normal functioning of the *Arabidopsis* clock [71].

1.4.4 The LHY/CCA1-TOC1 feedback loop

The first proposed mechanism for the central clock in *Arabidopsis* is based on a single positive and negative feedback loop involving *LHY/CCA1* and *TOC1* [72] (Figure 1.8). Microarray analysis identified a sequence in the promoters of many circadian regulated genes which peaked in the evening, the evening element [51]. LHY and CCA1 were shown to bind to the evening element in the *TOC1* promoter, and it is proposed to negatively regulate *TOC1*'s transcription, as *TOC1* mRNA levels are low and arrhythmic in an LHY or CCA1 overexpressor [72, 59]. TOC1 is then proposed to go on to positively activate LHY and CCA1, thus completing the loop. The evidence for TOC1 acting positively on *LHY* and *CCA1* comes from much reduced *LHY* and *CCA1* mRNA levels in the *toc1-2* mutant. Further supporting evidence for this model is that a double *lhy;cca1* mutant results in a very disrupted clock, which has been described as arrhythmic (Reviewed in [61]).

1.5 Other Clock related Genes

Although a useful model, there are several genes that have not been placed inside the *Arabidopsis* central clock that appear to play a role. We describe these genes below.

1.5.1 GI

GIGANTEA (GI) is an evening expressed gene, with a very similar expression pattern to that of *TOC1* [73]. *GI* mRNA is at a low, arrhythmic level in plants that over-express LHY [74] or TOC1 [59]. The sequence of the *GI* promoter includes several evening elements, the putative binding sites for LHY [51].

It had been suggested that GI encodes a novel, putative membrane protein [74], which made it unlikely that GI is involved directly in regulating transcription in other genes, but GFP imaging has shown GI in the nucleus [75, 76]. *gi* loss-of-function mutations result in low amplitude circadian rhythms, with low levels of *LHY* and *CCA1*

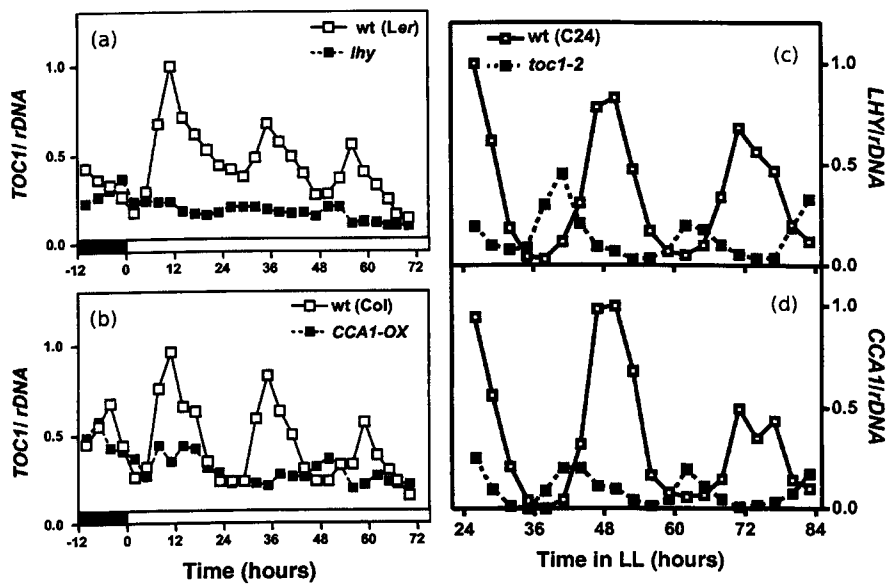


Figure 1.7: *TOC1* expression is strongly repressed in; (a) *LHY* over expressing plants (*lhy*); (b) *CCA1* over expressing plants (CCA1-OX). The dark and white bars at the bottom of each graph represent lights off and on, respectively; (c) *LHY* and; (d) *CCA1* expression is reduced in a strong toc1 mutant (toc1-2). Plants were germinated and entrained to cycles of 12h light, 12 h dark [LD (12, 12)] for 7 days and then released into constant light (LL). mRNA levels were analysed by Northern blotting (a technique for detecting specific RNAs separated by electrophoresis by hybridisation to a labelled DNA probe). Representative data are shown. Figure reproduced with permission from [72].

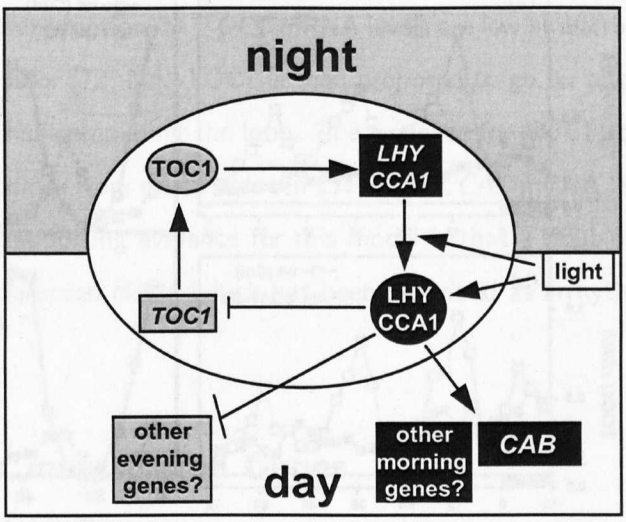


Figure 1.8: The LHY/CCA1-TOC1 feedback loop. Light activates *LHY* and *CCA1* expression at dawn. CCA1, and probably LHY, activate *CAB* expression; they may activate other genes that cycle with a phase similar to CAB. LHY and CCA1 simultaneously repress *TOC1* and potentially other evening genes. Progressive reduction of LHY and CCA1 expression levels during the day allows *TOC1* transcript levels to rise and reach maximum levels toward the end of the day, when LHY and CCA1 expression levels are lowest. TOC1, either directly or indirectly, appears to augment the expression of *LHY* and *CCA1*, which reach maximum levels at dawn, starting the cycle again. Figure reproduced with permission from [72].

RNA and either shorter or longer circadian periods [74, 73] or, in some conditions, in arrhythmia [77].

1.5.2 ELF4

ELF4 is an evening expressed gene, which is involved in photoperiod perception and circadian regulation [78]. Knockouts cause attenuated expression of *CCA1*, while output genes, such as *CCR2*, show output rhythms with highly variable period lengths. Mutations also cause early flowering, caused by elevated activation of CONSTANS, (CO), a gene that promotes floral induction. It appears that *ELF4* could carry out a very similar role to *TOC1*, activating *LHY* and *CCA1* [79], although more experiments are needed to place *ELF4* in the loop.

1.5.3 The PRR quintet

TOC1 is a member of a gene family comprising a quintet of genes called *PSEUDO RESPONSE REGULATORS, (PRR1/TOC1, PRR3, PRR5, PRR7, PPR9)*, that are under the control of the circadian clock, such that the *PRR* mRNAs start accumulating sequentially after dawn with 2 to 3h intervals, in the order of *PRR9-PRR7-PRR5-PRR3-PRR1/TOC1* [80]. It has been suggested that each gene activated each other in turn, although recent evidence suggests that this is not the case (Reviewed in Ref.[81]). The circadian phenotypes of the single *prr* mutants are small, with a period variation of 1 to 1.5h compared to the severe period shortening seen in the *toc1-2* mutant and the overexpression of PRR3, PRR5 or PRR9 only modestly affect the clock outputs whereas overexpression of TOC1 results in arrhythmicity [82, 81]. Hence, the other members of the *PRR* quintet do not seem to be as important to the clock as *TOC1* is. Recent investigations into double and triple mutants of the *PRR* genes have revealed a complex set of connections. It appears that PRR7 and PRR9 may be redundant genes involved in light and temperature input to the clock, and might also form another feedback loop with *LHY/CCA1* [83, 84]. A triple mutation of *PRR5*, *PRR7* and *PRR9* appears to be

arrhythmic in constant light and darkness [85]. Further characterisations of the PRR family are required before a full understanding of their role in the clock can be attained.

1.5.4 ELF3

ELF3 mediates the circadian gating of light responses and regulates light input to the clock [86, 87]. The *elf3* mutant abolishes circadian rhythms in *CAB* and *CCR2* gene expression in constant light, but *CCR2* is rhythmic in DD. This suggests that ELF3 only acts during light, and so is not a component of the core oscillator, but is in the pathway of the light input to the clock. *ELF3* appears to gate light response to the clock, the protein is found to peak at the middle of the night, and *elf3* mutants become arrhythmic in the middle of the night if pulsed with light [88, 89].

1.6 Spatial Organisation of the Plant Circadian System

In mammals a small cluster of cells in the supra chiasmatic nuclei in the brain, which are entrained by light signals from the eyes, are responsible for entraining the circadian clocks in cells throughout the body [90]. In plants, however, photoreceptors are present throughout the organism and there does not appear to be a central clock that entrains all the tissues. Circadian clocks in different cells in plants appear to be weakly coupled or independent oscillators. *CAB:LUC* expressing tobacco seedlings whose leaves are entrained to opposite phases maintain their different phases in constant conditions, and thus, are not entrained from a central pacemaker [91]. Also, output genes of the clock have been shown to have different periods under the same background and light conditions [92]. This suggests that the plant might contain several oscillators which are either not coupled, or not strongly coupled. Further evidence for different oscillators has been found in *Arabidopsis*, where different oscillators have been shown to control *CAB:LUC* and *CATALASE3:LUC* rhythms due to different temperature sensitivity [93].

1.7 Outline of thesis

By collaborating closely with experimentalists in Andrew Millar's lab, we have been able to take an iterative approach of mathematical modelling and experiment to develop an accurate mathematical model of the circadian clock in *Arabidopsis*. This approach is central to systems biology [94] which aims to integrate different levels of information in mathematical models to understand how biological systems function.

We first model the current understanding of the *Arabidopsis* circadian clock, i.e. the LHY/CCA1-TOC1 feedback loop, in order to see whether this network can explain the existing experimental data (Chapter 2). We develop an ODE model, using Michaelis-Menten kinetics and Hill functions, of the clock and optimise the parameters for this network using a novel optimisation scheme, outlined in Chapter 3.

In Chapter 4 we add extra components to the LHY/CCA1-TOC1 feedback loop to try and fit to more experimental data. For each new model we re-optimise all the parameters. This leads us to a two-loop model for the *Arabidopsis* clock involving two predicted genes, factor X and Y. In collaboration with experimentalists we find that *GI* is a candidate gene for Y.

In Chapter 5 we develop simplified versions of the *Arabidopsis* clock models. We do not model protein levels, replacing them with explicit delays. This allows a great reduction in the number of parameters and equations required, and we investigate these models accuracy compared to the biochemical reaction models described in Chapters 2 and 4.

In Chapter 6 we further develop the two loop model for the *Arabidopsis* clock to include the recently discovered feedback loop between LHY/CCA1 and PRR7/PRR9. Using this new triple-loop model we present evidence that *GI* acts as Y. Our results also suggest the plant clock comprises morning and evening oscillators, coupled intracellularly, which may be analogous to the morning and evening cells in the clocks of *Drosophila* and the mouse.

Chapter 2

Modelling the Arabidopsis clock

The first multi-gene loop identified in the *Arabidopsis* circadian clock comprises a negative feedback loop, in which two partially redundant genes *LATE ELONGATED HYPOCOTYL* (*LHY*) and *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) repress the expression of their activator, *TIMING OF CAB EXPRESSION 1* (*TOC1*) [72]. Several other clock genes have been discovered in *Arabidopsis* (reviewed in [61], Chapter 1), but these either have accessory functions [65] or have not yet been located relative to the one-loop LHY/CCA1 - TOC1 model. We therefore begin modelling the *Arabidopsis* clock by analysing the one-loop LHY/CCA1 - TOC1 model, using an optimisation method that will be widely applicable in tackling larger genetic networks (the optimisation process is described in detail in Chapter 3). By comparison with experimental data, the model suggests where additional components of the clock network may function.

This chapter is based on work published in 'Modelling genetic networks with noisy and varied data: The circadian clock in *Arabidopsis thaliana*', J. C. W. Locke, A. J. Millar and M. S. Turner, *J Theor. Biology*, **234**, 383-393, (2005).

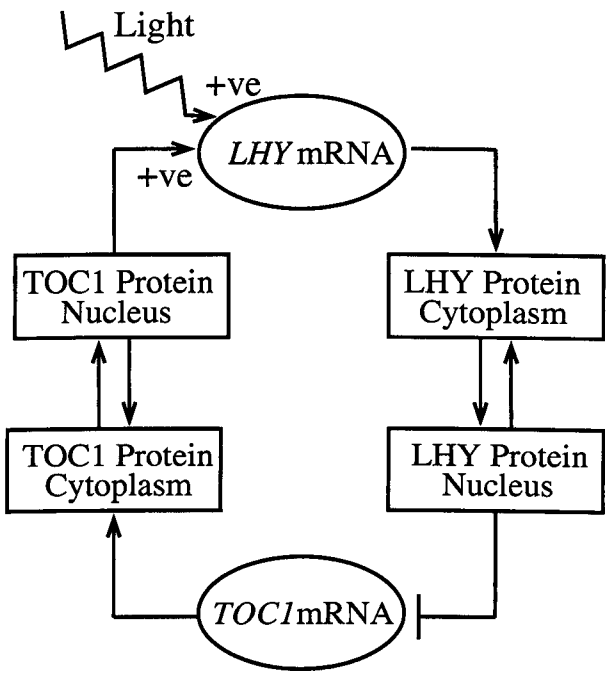


Figure 2.1: Model for the central feedback loop in the *Arabidopsis* clock. TOC1 protein in the nucleus activates transcription of *LHY* and *CCA1* mRNA, together with light mediated by protein P (not shown). When LHY and CCA1 proteins reach the nucleus they down regulate *TOC1* mRNA transcription.

2.1 Model Description

The initial description of the *Arabidopsis* clock network as outlined in Figure 2.1 required seven coupled differential equations to model the central loop, yielding a total of 29 parameters. As in previous clock models [18, 47, 39, 95, 28] Michaelis-Menten kinetics were used to describe enzyme mediated degradation of proteins, and Hill functions were used to describe the transcriptional activation term of the mRNA for *LHY* and *TOC1*. We use the cytosolic and nuclear pools of our model proteins to represent all the processes between the accumulation of an mRNA and the regulation of the next gene in the network by an active form of the protein. Other clock models include additional processes, such as mRNA export from the nucleus, but there is currently no data to

specify their dynamics in plants. The converse approach is to combine all intermediate steps as a time delay between the synthesis of RNA and active protein (See Chapter 4).

As LHY and $CCA1$ are indistinguishable for our purposes, we retain only one gene, LHY, in our model. Although quantitative differences in LHY and CCA1 regulation have sometimes been reported, [74], there is strong experimental evidence that LHY and CCA1 play redundant roles in the *Arabidopsis* circadian clock [64], and that their qualitative behaviour is very similar. Combining *LHY* and *CCA1* genes removes 16 parameters and 3 equations from the model. In principle these could later be included when better informed by further data.

We adopt the following equations to define our mathematical model for the central circadian network: an LHY-TOC1 feedback loop that involves the cellular concentrations $c_i^{(j)}(t)$ of the products of the i^{th} gene ($i = L$ labels LHY, $i = T$ labels TOC1) where $j = m, c, n$ denotes that it is the corresponding mRNA, or protein in the cytoplasm or nucleus respectively.

$$\frac{dc_L^{(m)}}{dt} = L(t) + \frac{n_1 {c_T^{(n)}}^a}{{g_1}^a + {c_T^{(n)}}^a} - \frac{m_1 c_L^{(m)}}{k_1 + c_L^{(m)}} \tag{2.1}$$

$$\frac{dc_L^{(c)}}{dt} = p_1 c_L^{(m)} - r_1 c_L^{(c)} + r_2 c_L^{(n)} - \frac{m_2 c_L^{(c)}}{k_2 + c_L^{(c)}} \tag{2.2}$$

$$\frac{dc_L^{(n)}}{dt} = r_1 c_L^{(c)} - r_2 c_L^{(n)} - \frac{m_3 c_L^{(n)}}{k_3 + c_L^{(n)}} \tag{2.3}$$

$$\frac{dc_T^{(m)}}{dt} = \frac{n_2 {g_2}^b}{{g_2}^b + {c_L^{(n)}}^b} - \frac{m_4 c_T^{(m)}}{k_4 + c_T^{(m)}} \tag{2.4}$$

$$\frac{dc_T^{(c)}}{dt} = p_2 c_T^{(m)} - r_3 c_T^{(c)} + r_4 c_T^{(n)} - \frac{m_5 c_T^{(c)}}{k_5 + c_T^{(c)}} \tag{2.5}$$

$$\frac{dc_T^{(n)}}{dt} = r_3 c_T^{(c)} - r_4 c_T^{(n)} - \frac{m_6 c_T^{(n)}}{k_6 + c_T^{(n)}} \tag{2.6}$$

Here the various rate constants n_k, g_k etc., parameterise transcription (n_k, g_k), degradation (m_k, k_k), translation (p_k), and the nuclear $\leftrightarrow$ cytoplasmic protein transport rate

(r_k). There is evidence that LHY and CCA1 proteins bind as a dimer to the promoter of $TOC1$ [71], and that there is only one active binding site on the $TOC1$ promoter [72], so the Hill coefficient of transcription, b, was set to 2. As there is no experimental evidence for the Hill coefficient a this was set to 1. The effect of light appears through the term $L(t)$. Light is known to give an acute, transient activation response for expression of LHY and $CCA1$ [96, 97, 98]. This was modelled through a simple mechanism involving an interaction of a protein P with the LHY gene promoter. P is a light sensitive protein similar to PIF3 in stability [99] that is present with concentration $c_P^{(n)}$.

$$L(t) = q_1 c_P^{(n)} \Theta_{\text{light}} \tag{2.7}$$

where $\Theta_{\text{light}} = 1$ when light is present, 0 otherwise. We propose that P is controlled by an equation of the form:

$$\frac{dc_P^{(n)}}{dt} = (1 - \Theta_{\text{light}})p_3 - \frac{m_7 c_P^{(n)}}{k_7 + c_P^{(n)}} - q_2 \Theta_{\text{light}} c_P^{(n)} \tag{2.8}$$

where the values of the four parameters that appear in this equation are chosen so as to give an acute light activation profile that is close to that observed in experiment. In principle these could also be optimised under our scheme, but since P is coupled into Eqns. 2.1 to 2.6 via an arbitrary coupling constant q_1, which is varied in our optimisation scheme, we consider this an adequate approach that captures the primary role of P in mediating light input. The essential features of Eqn. 2.8 are that P is produced only when light is absent and is degraded strongly when light is present. We modelled the acute effect of light at the level of transcription but note that a similar effect on translation would result in essentially identical network behaviour [96]. This left 23 parameters to be determined by our optimisation scheme (Chapter 3).

2.1.1 Experimental Data

Various experimental data sets [80, 96, 64] give us approximate values for the phase and period of the oscillations in mRNA levels of the known central clock genes in *Arabidopsis*,

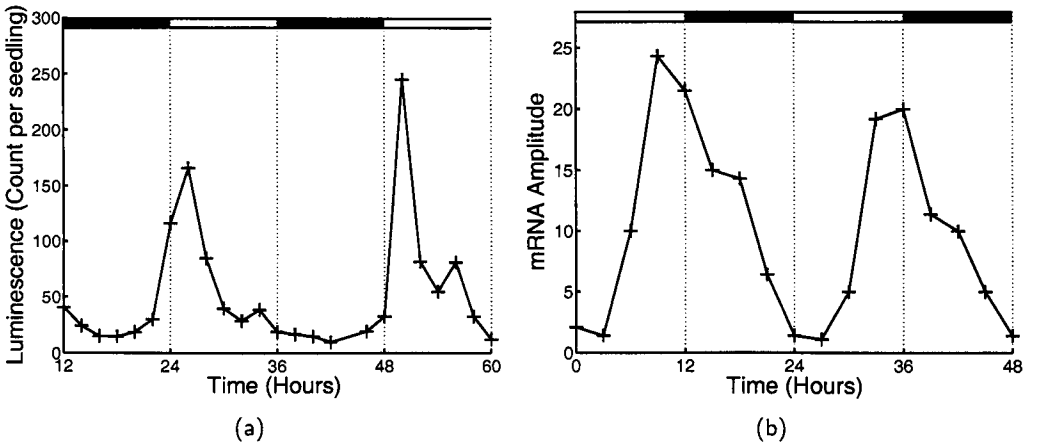


Figure 2.2: Typical experimental data sets for the time-variation of mRNA levels, for (a) *LHY*, inferred from the luminescence levels of a Luciferase reporter gene [96], and (b) *TOC1* [64]. Note that the phase of the peak of *LHY* is just after dawn while the phase of the peak in *TOC1* is much later, at approximately 12h after dawn.

see Figure 2.2. $TOC1$ and LHY mRNA expression levels have been shown to peak around dusk and dawn respectively. In contrast neither the absolute nor the relative levels of LHY and $TOC1$ mRNA are known. Furthermore, none of the parameters in our model, such as the transcription rate of $TOC1$, have been experimentally measured. We estimated these parameters using a global optimisation scheme described in detail in Chapter 3, and summarised here. Our optimisation approach involves first defining a qualitative cost function Δ, effectively a set of least-squared error terms representing the essential experimental phenomena that the model must reproduce. This cost function is then calculated across a sample of points in the high dimensional parameter space. By point we here mean a vector (position), the components of which represent each of the free parameters of the model. The 50 points with the lowest cost function score (best agreement with experiment) found from this search are then passed on to a further local simulated annealing scheme. Our scheme resulted in a set of 41 solutions with $\Delta < 100$, which we took as our set of candidate solutions to be examined further.

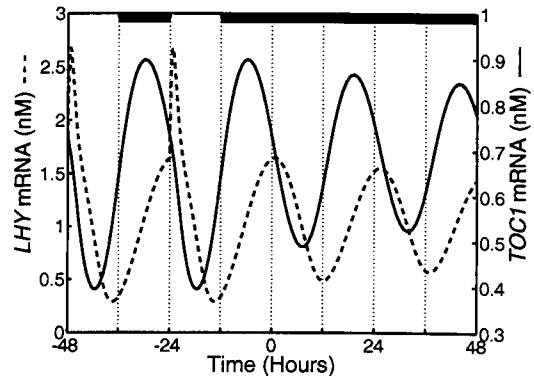


Figure 2.3: mRNA Levels for a typical annealed solution, $\Delta = 81$. $q_1 = 4.5703$ h^{-1}, q_2=1.0 h^{-1}, n_1=7.5038 nM/h, n_2=0.6801 nM/h, g_1=1.4992 nM, g_2=3.0412 nM, m_1=10.0982 nM/h, m_2=1.9685 nM/h, m_3=3.7511 nM/h, m_4=2.3422 nM/h, m_5=7.2482 nM/h,m_6=1.8981 nM/h, m_7=1.2 nM/h, p_1=2.1994 h^{-1}, p_2=9.4440 h^{-1},p_3=0.5 h^{-1}, r_1=0.2817 h^{-1},r_2=0.7676 h^{-1},r_3=0.4364 h^{-1}, r_4=7.3021 h^{-1}, k_1=3.8045 nM,k_2=5.3087 nM,k_3=4.1946 nM, k_4=2.5356 nM, k_5=1.4420 nM, k_6=4.8600 nM, k_7=1.2 nM, $a = 1$, $b = 2$. In this and other figures; filled box above panel, dark interval; open or no box, light interval. The time is redefined so $t = 0$ corresponds to the first dawn missing in constant conditions. In order to entrain the oscillations, the equations are solved for 300h under LD conditions before being solved for constant conditions.

2.2 Results of parameter search

The result of our extensive parameter search suggests that it is not possible to fit the single loop model to all the experimental data. Figure 2.3 shows a typical solution obtained after the simulated annealing, with a score of $\Delta = 81$. $TOC1$ mRNA expression peaks too late in the daily cycle, and LHY mRNA expression comes up too soon in the night. This solution oscillates on a limit cycle in DD with a period of 25h.

Figure 2.4 shows the results for the optimal solution, i.e. the one with the lowest cost function score, in this case $\Delta = 27$. In this solution, $TOC1$ and LHY mRNA levels both peak at roughly the correct time, and the oscillation slowly damps in darkness with a period of 25h, similar to experiment.

Using this model, it is possible to investigate how light entrains the clock, by

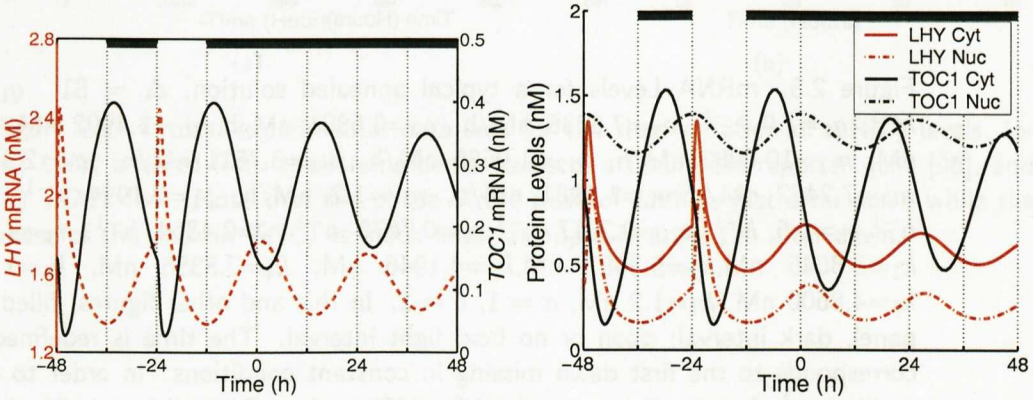


Figure 2.4: (a) mRNA levels for optimal parameter set for the LHY/CCA1-TOC1 network, $\Delta = 27$. $q_1 = 2.9741\ h^{-1}$, q_2=1.0 h^{-1}, n_1=16.9711 nM/h, n_2=1.3043 nM/h, g_1=3.0351 nM, g_2=0.386 nM, m_1=9.3383 nM/h, m_2=16.9058 nM/h, m_3=1.0214 nM/h, m_4=1.6859 nM/h, m_5=0.4212 nM/h, m_6=0.0484 nM/h, m_7=1.2 nM/h, p_1=4.9753 h^{-1}, p_2=1.2947 h^{-1}, p_3=0.5 h^{-1}, r_1=1.4563 h^{-1}, r_2=0.8421 h^{-1}, r_3=0.0451 h^{-1}, r_4=0.0018 h^{-1}, k_1=1.3294 nM, k_2=0.8085 nM, k_3=0.1445 nM, k_4=0.2089 nM, k_5=0.3187 nM, k_6=0.3505 nM, k_7=1.2 nM, $a = 1$, $b = 2$. (b) Protein levels for optimal model of the LHY/CCA1-TOC1 network.

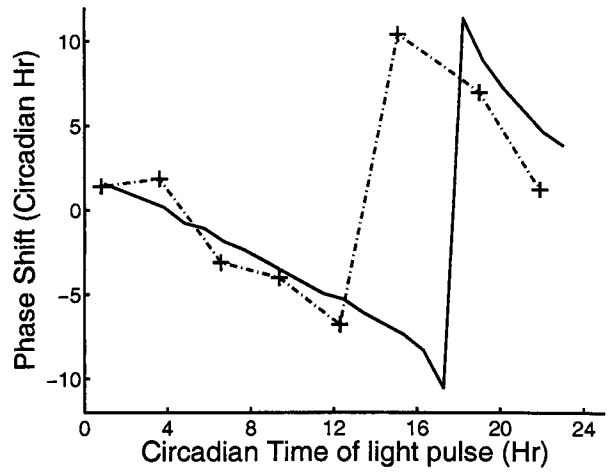


Figure 2.5: Simulated Phase Response Curve (PRC) for optimal solution of LHY/CCA1-TOC1 network (solid curve). Phase advances are plotted as positive values, and delays are plotted as negative values. The simulation followed the same experimental protocol as in the data extracted from a published red light PRC, [88] (dashed curve).

simulating a light treatment phase-response curve experiment (PRC). Circadian rhythms are entrained by effects of transitions - lights-on or lights-off - on the phase of the rhythm. Switching on the light in the morning, for instance, induces a phase shift: either a phase advance or a phase delay (See Chapter 1 for review). A light treatment phase-response curve experiment systematically explores the phase-shifting effects of light. Conventionally, the phase shifts in expression that are generated by one hour light treatments are plotted against the circadian time at which the light pulses were given. Our simulated PRC has a similar shape to experiment, see Figure 2.5.

The phases of the peak of LHY mRNA and $TOC1$ mRNA under LD cycles are not overly sensitive to light levels (both an 80% reduction in light levels and a 100% increase in light levels causes no significant change in phase). Although not imposed as a specific constraint in the cost function, LHY protein levels do indeed peak 1-2h after the peak of LHY mRNA expression, as suggested in [69]. However, LHY mRNA expression still over-anticipates dawn. We found that 5 of the annealed solutions had low cost function scores with the pathological feature that $TOC1$ mRNA expression

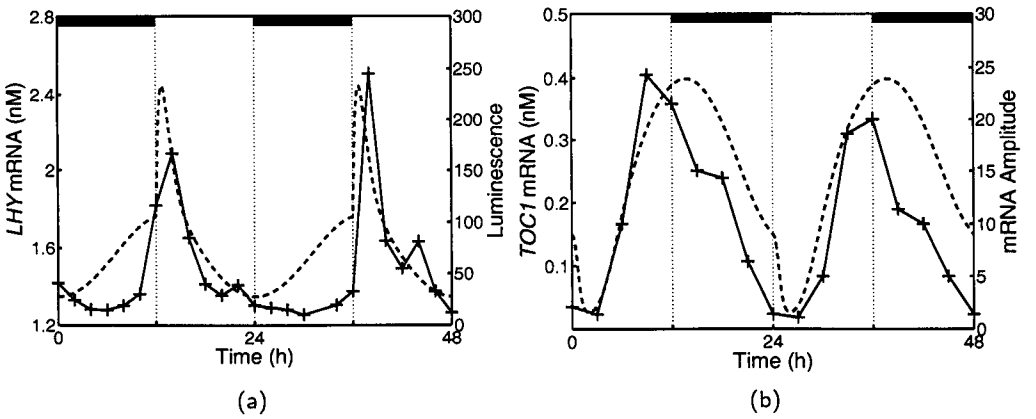


Figure 2.6: Comparison between experimental concentrations (solid curves, arbitrary conc. scale, right axis) with those obtained from our optimal model (dashed curves, left axis), as also shown in Figures 2.2 and 2.4 respectively. Our model reproduces the phases of LHY and $TOC1$ mRNA correctly. LHY anticipates dawn in the simulation to a greater degree than experiment. Experimental traces show $TOC1$ levels are falling during the night before LHY levels start to rise, suggesting that our model is missing some factor that would serve to bring $TOC1$ down at the end of the day.

had saturated through the night. This suggests that in future work an extra term in the cost function might be added for the profile shape of $TOC1$ mRNA.

We interpret the inadequacies in all the annealed solutions as reflecting failures in the structure of the network. This is because the global search shows that there is not a distribution of model parameters that gives a good agreement with experiments for this network structure. Further examination of the experimental data shows that $TOC1$ mRNA expression starts to fall before LHY mRNA expression has started to increase, meaning that experimental efforts should be focused on understanding what is missing from the network in order to bring down $TOC1$ mRNA levels at night, see Figure 2.6. Also, in order for LHY mRNA to be expressed for a shorter time interval compared to $TOC1$ mRNA, post-translational modification of the TOC1 protein may be required.

Previous papers [29] have carried out stability analyses based on the period and amplitude of solutions after a small percentage increase and decrease for each parameter value in turn. We have also carried out such an analysis which shows that, as for previous

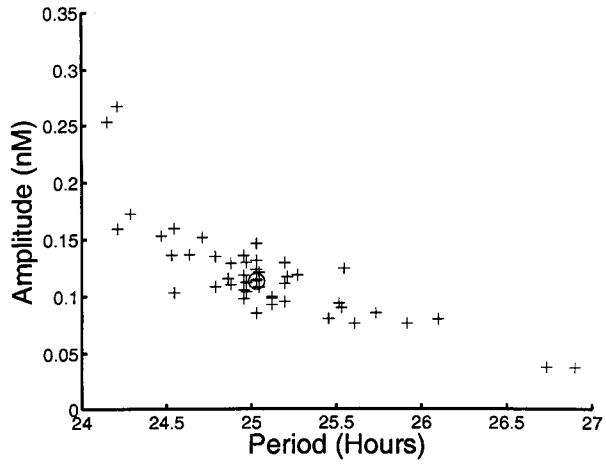


Figure 2.7: Stability analysis for optimal parameter set of LHY/CCA1-TOC1 network. The mean period and amplitude of *TOC1* mRNA over 300h in DD are calculated for a small perturbation, a 5 percent increase and decrease, for each parameter value in turn. The circle represents the period and amplitude of the original parameter values.

models, the solution is extremely sensitive to small changes of a few parameters, see Figure 2.7. For example, a reduction of five percent in n_2, the transcription rate of *TOC1* mRNA, causes the oscillations to damp to experimentally undetectable levels after 300h in darkness.

The LHY/CCA1-TOC1 network does not respond to changes in day length, as simulated gene expression profiles are identical in LD cycles with long and short photoperiods, Figure 2.8. It is clear experimentally that the clock has a later phase under longer photoperiods [101, 100]. The model is insensitive to light at the end of the day, because the effect of protein P effectively terminates within 3 h after lights on. Otherwise there should be a drop of *LHY* and *CCA1* transcription at dusk. *LHY* and *CCA1* expression fall to a low level before the end of a 12h photoperiod [96], which suggests there might be another mechanism to mediate light input at the end of the day.

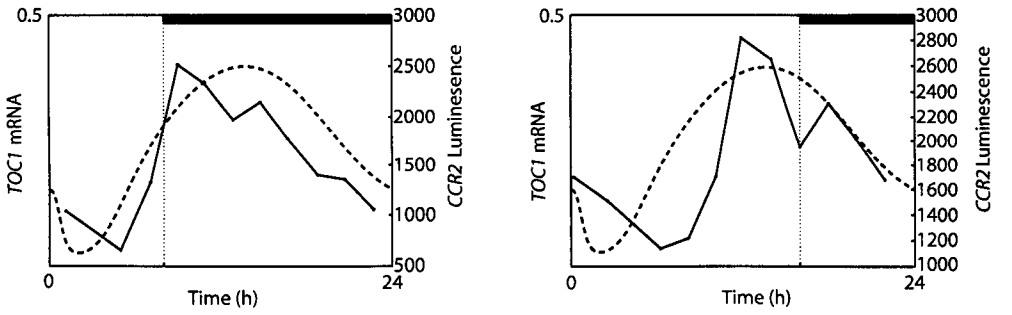


Figure 2.8: Effects of altered photoperiod on circadian rhythms. Simulations (dashed line, (nM)) are compared to data (solid line, arbitary concentrations) a) under light dark cycles (LD) consisting of 16h of darkness and 8h of light (LD16:8), b) under light dark cycles (LD) consisting of 8h of darkness and 16h of light (LD8:16). CCR2 luciferase imaging data data from [100] is used as a late evening marker to compare to simulations of *TOC1* mRNA, as *CCR2* and *TOC1* have been seen to have comparable phases of expression in vivo (for example ref [64]).

2.2.1 Simulated Mutant Analysis

We have further characterised the output of the 41 annealed parameter sets with $\Delta < 100$, by carrying out a simulated mutant analysis, see Figure 2.9. This involved the comparison of simulated *lhy* mutants to experimental data, and revealed further interesting information about the role of LHY in the network.

Single null mutations in LHY or $CCA1$ experimentally result in short periods of around 21h [72]. We simulated this experiment by reducing the translation rate of LHY protein to half of its original value, and for every single one of our annealed parameter sets this resulted in a long period mutant or arrhythmia. This is because in our model, a reduction in LHY level delays the repression of $TOC1$, so $TOC1$ mRNA peaks later in the night, a feature that is not observed experimentally. This discrepancy suggests that the main role of LHY in the plant clock is to stop $TOC1$ mRNA peaking too early in the day, rather than to bring $TOC1$ levels down during the night, which it does in the model. All of the 41 solutions with $\Delta < 100$ showed similar mutant analysis, suggesting

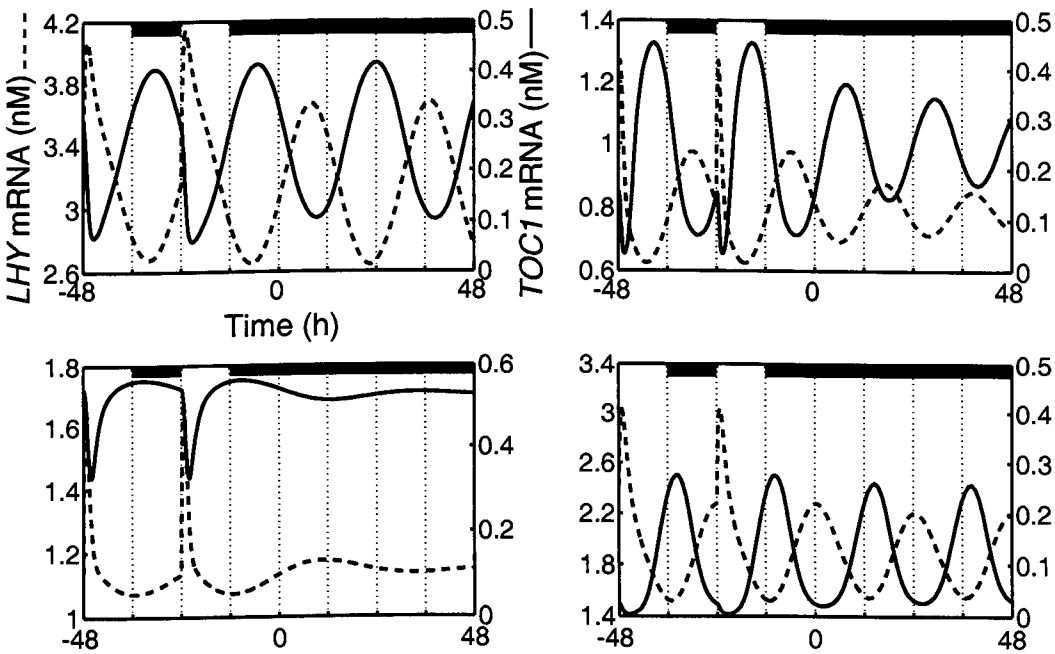


Figure 2.9: Simulation of gain and loss of function mutants of *LHY* and *TOC1* using optimal model of LHY/CCA1-TOC1 network. Upper left pane, Translation rate of LHY, (p_1) halved, the average period over the 300h in DD, $\tau_d = 29.2$h. Upper right pane, p_1 doubled, $\tau_d = 21.3$h. Lower left pane, translation rate of TOC1, (p_2), halved, $\tau_d = 44.7$h. Lower right pane, p_2 doubled, $\tau_d = 23.5$h.

that the failure in this regard is a generic feature of the network.

2.3 Discussion

We have modelled the LHY/CCA1 - TOC1 feedback loop of *Arabidopsis*, which was first proposed by Alabadi [72] as an important component of the plant circadian clock. We developed a cost function to score the performance of the model with a particular set of parameter values, together with a method to explore the space of possible parameter values and identify an optimal set. The resulting model reproduces many features of *Arabidopsis* circadian rhythms, and it points to a specific phase of the circadian cycle where further experimentation is required to identify one or more additional components of the circuit. A significant advantage of our approach is in determining that a gene network model is inconsistent with experimental data because its circuit is incorrect, rather than simply being due to a poor choice of parameter values.

The LHY/CCA1 - TOC1 loop was proposed based on the mRNA accumulation patterns of *Arabidopsis* plants in which these genes were mutated or mis-expressed [72, 69, 66]. Our model of this loop drives rhythms with the major circadian properties observed in *Arabidopsis* which were specified in the cost function used to select the parameter values. The model has a period close to but longer than 24h in constant darkness, it entrains to 24h light:dark cycles, and it shows peak expression of the $LHY/CCA1$ and $TOC1$ mRNAs close to the required phases. It also reproduces features that were not explicitly specified in the cost function, such as a phase-response curve (PRC) to light treatments with the correct characteristic shape, Figure 2.5. For some parameter values, the model has a sustained limit cycle in constant darkness, (see Figure 2.3), though the optimal solution is slightly damped, as is often observed in experimental data. The model supports rhythms in constant light, as it does in darkness. We do not address any temperature effect explicitly, due to the paucity of relevant molecular data. The model is likely to be capable of temperature compensation, as much simpler models can exhibit this behaviour [19].

The waveforms of *LHY* and *TOC1* expression during the day match the data closely, both for the optimal parameter choice, Figure 2.6, and with many of the optimised parameter sets selected from a million quasi-random choices, Figure 2.3, indicating that no further biochemical complexity is necessary to simulate this phase of the cycle. However, even the optimal parameter choice fails to reproduce mRNA accumulation patterns during the night. *LHY* mRNA levels start to rise early in the night in the optimal parameter choice under LD cycles, anticipating the observed rise by several hours. Simulated mutant analysis highlights this limitation more clearly. Deletion of either the *LHY/CCA1* or the *TOC1* function leads trivially to arrhythmia, so another timing loop must be present in the plant, because complete arrhythmia is observed in the relevant mutants only in specific conditions [65], if at all [67]. Such components might simply be paralogues of *LHY/CCA1* and *TOC1* with non-identical functions [102, 103] which are absent from the model. More significantly, partial reduction or increase in either gene function in the model causes period changes in the opposite direction to those observed, Figure 2.9. This discrepancy cannot be explained by partial gene redundancy in the plant, nor by a feature specific to the optimal parameter set in the model, as it was observed in all of the annealed solutions with $\Delta < 100$. It arises because LHY in the model represses *TOC1* expression continuously from the mid-night phase, so reduced LHY function leads to a later peak of *TOC1* expression at night and a longer period. Neither LHY nor CCA1 expression is observed until close to dawn in the plant [69, 96], so *LHY* and *CCA1* mutants allow an early rise in TOC1 expression in the day and have a short period.

The failure of the model to reproduce gene expression patterns during the night or the gross phenotypes of clock mutants, despite this global search of parameter space, indicates that the gene network in the model is incorrect. In Chapter 4, we attempt to extend our network model to include extra gene functions, in order to fit to more experimental data.

Chapter 3

Optimisation scheme

3.1 Parameter Optimisation

Parameter Optimisation is a very large field, touching on many areas of research, including the physical and mathematical sciences, and several branches of engineering. Essentially it involves minimising or maximising a function that depends on one or more independent variables to obtain some sort of "best fit". Our optimisation approach involves first defining a qualitative cost function, effectively a set of least-squared error terms representing the essential experimental data that the model must fit to. This cost function is then calculated for a sampling of parameter space, the coordinates being chosen using a quasi-random number generator on a uniform, bounded domain. The best points found from this search are then passed on to a further local simulated annealing scheme. Each section of this process is described in detail in this chapter (See Figure 3.1.

This chapter is based on work published in 'Modelling genetic networks with noisy and varied data: The circadian clock in *Arabidopsis thaliana*', J. C. W. Locke, A. J. Millar and M. S. Turner, *J. Theor. Biology*, **234**, 383-393, (2005).

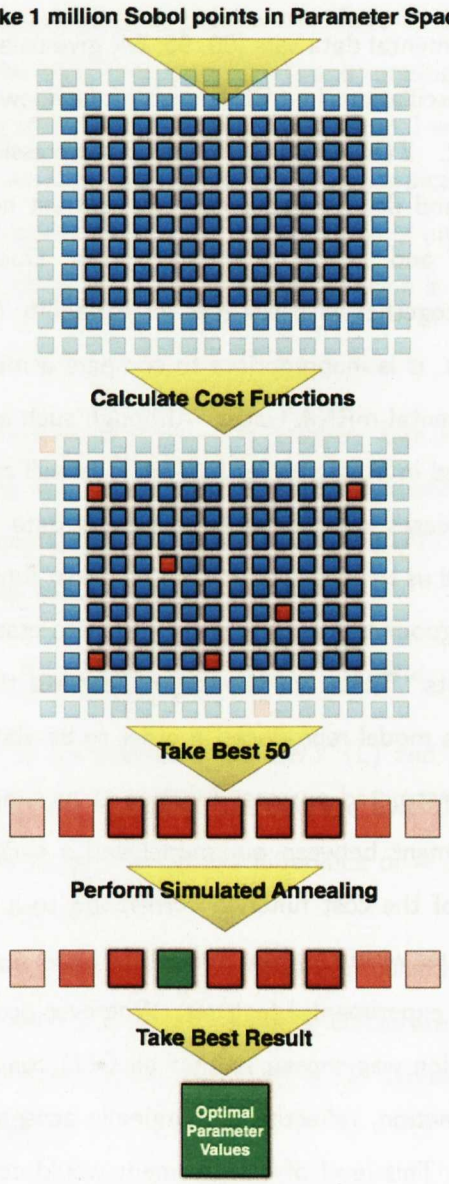


Figure 3.1: Flow diagram of optimisation scheme. The qualitative cost function Δ is solved for 10^6 points in parameter space chosen using a Sobol quasi-random number generator. The top fifty parameter sets with the lowest Δ scores are then passed to a further simulated annealing acheme. The parameter set which gives the lowest Δ score after 10^5 annealing points is taken as our optimal parameter set.

3.1.1 The Cost Function

Various experimental data sets [80, 96, 64] give us approximate values for the phase and period of the oscillations in mRNA levels of the known central clock genes in *Arabidopsis*, see Figure 2.2. $TOC1$ and LHY mRNA expression levels have been shown to peak around dusk and dawn respectively. In contrast neither the absolute nor the relative levels of LHY and $TOC1$ mRNA are known. Due to the lack of time points in RNA experiments, together with the level of variability (noise) currently ubiquitous in such biological data, it is inappropriate to compare a model quantitatively to any particular set of experimental mRNA traces. Although such a direct quantitative comparison has been attempted in other studies [104, 48], as well as in cell cycle models [105], it is our belief that in cases with sparse experimental data this simply will not be appropriate. This motivated us to construct an empirical cost function designed to give a quantitative value for the goodness of fit of our solution to essential qualitative features present in the experiments. These are the sort of features that plant circadian biologists would demand that a model reproduces in order to be viable.

We constructed our cost function Δ as simply a sum of terms that each quantify the agreement between our model and a single qualitative experimental feature. Small values of the cost function correspond to a model (or, more precisely, a set of parameter values within a given model) that give a good qualitative agreement with the corresponding experimental features. Wherever possible the weighting of each term in the cost function was chosen so that an O(1) contribution would arise from any term in the cost function, reflecting a marginally acceptable divergence between the model and the data. This level of disagreement would correspond to roughly the size of the experimental error bars. There is inevitably some arbitrariness in how we defined the level of acceptable error, but at all times we tried to base our assumptions on inferences from experimental data. In order to evaluate the terms in the cost function we solved numerically Eqns. 2.1 to 2.8 over 600h, 300h in 12h light:12h dark cycles (LD), followed by 300 h in darkness (DD) (the first 200 h of each solution are discarded as transitory).

The equations were solved using MATLAB, integrated using the inbuilt stiff equation solver ODE15s [106]. The optimisation process described in the following sections was carried out by compiling the MATLAB code into C and running the code on a task farm super computer consisting of 31 x 2.6 GHz Pentium4 Xeon 2-way SMP nodes (62 CPUs in total). In what follows we identify 1nM and 1h as the typical concentration and time scales, and measure all concentrations and rate constants in units where these are unity. We initialised our simulation at $c_i^{(j)} = 1$. The cost function is given by:

$$\Delta = \delta_{\tau_{ld}} + \delta_{\tau_d} + \delta_\phi + \delta_{c_L} + \delta_{\text{size}} \tag{3.1}$$

we now describe each term of the cost function, Eqn.3.1, in turn.

First, $\delta_{\tau_{ld}}$ measures the difference between the experimental target period and the mean period of the oscillation in mRNA levels of LHY and $TOC1$ in light:dark (LD) cycles as exhibited by the model;

$$\delta_{\tau_{ld}} = \sum_{i=L,T} \langle (24 - \tau_i^{(m)})^2 / 0.15 \rangle_{ld} \tag{3.2}$$

This is the summed error in the period, τ, for LHY (L) and $TOC1$ (T) mRNA levels in light:dark cycles (LD), where $\langle\rangle_{ld}$ gives the average over the cycles between $200 < t < 300$, and a marginally acceptable period difference of $\approx 25\text{mins}$ contributes O(1) to the cost function for each term.

Second, the term δ_{τ_d} gives a similar measure in constant darkness (DD). These two terms ensure that the entrained and free running clocks are near limit cycles with the experimentally observed period (stably entrained in LD cycles and with a free running period greater than 24h [107]),

$$\delta_{\tau_d} = \sum_{i=L,T} \langle (25 - \tau_i^{(m)})^2 / f \rangle_d \tag{3.3}$$

where the average of $\langle\rangle_d$ is now over $300 < t < 600$ (DD). The biological evidence strongly indicates that the free running period of the clock is greater than 24 [107], probably about 25, but we have less confidence in assigning a precise value hence we adopt values of $f = 0.05$ if $\tau_i^{(m)} \leq 25$ and $f = 2$ if $\tau_i^{(m)} > 25$.

Thirdly δ_ϕ measures the difference between the target phase and the average phase of the peaks of LHY and $TOC1$ mRNA expression in LD. It also ensures that the oscillations are entrained to the LD cycles,

$$\delta_\phi = \sum_{i=L,T} \left[\langle \Delta\Phi_i^2 \rangle_{ld} + \left(\frac{\sigma[c_i^{(m)}(t_{\rm p})]_{ld}}{0.05\langle c_i^{(m)}(t_{\rm p})\rangle_{ld}} \right)^2 + \left(\frac{\sigma[\Delta\Phi_i]}{5/60} \right)^2 \right] + \delta_{ent} \qquad (3.4)$$

The first term compares the mean difference in phase over the LD cycles, where $\Delta\Phi_i = \bar{\phi}_i - \phi_i$, ϕ_i is the phase (from dawn) of the RNA peak in the model and $\bar{\phi}_L = 1hr$, $\bar{\phi}_T = 11hr$ are the target phases of the peaks in $c_L^{(m)}$ and $c_T^{(m)}$ respectively. We assume a cost that is O(1) for solutions that differ by an hour. The next two terms ascribe a cost of O(1) for limit cycle solutions in LD cycles whose peak heights vary only within 5 percent of one another, and whose variations in peak phases are 5 minutes. $\sigma[]_{ld}$ is the standard deviation for the cycles in LD. The term δ_{ent} checks that the solution is truly entrained to the light/dark cycle, i.e is not oscillating with the correct phase simply because of the initial conditions chosen. This is achieved as follows: the solution is rerun for 75h, taking the solution at 202h and shifting it back 3h, i.e initialising the $t = 202$ solution as the t = 199 solution. The new phase of the second peak is compared to the original phase of the second peak. If the phase discrepancy is still near 3 h, then the solution is too weakly entrained, and the solution is pathological. The LD cycles have failed to phase shift the response. We assume that the rate of adjustment of the phase is linear in the discrepancy of the phase. This gives us a phase discrepancy that goes to 0 exponentially in time (like the radioactive decay equation). The characteristic time is then trivially related to the log of the phase discrepancy. It is this logarithmic variation that is reflected in our choice of δ_{ent}. Hence δ_{ent} takes the form of $\log(0.5)/\log(\delta\phi/3)$, where $\delta\phi$ is the phase discrepancy in hours between the shifted and original solution, and $\delta\phi/3$ is therefore the fraction of the imposed 3h phase shift remaining after 2 periods. The term $\log(0.5)$ gives the acceptable remaining phase difference of 1.5h for the second cycle, which results in an O(1) contribution to the cost function.

Next $\delta_{\rm size}$ checks that the oscillation sizes are large enough to be detectable

experimentally, and quantifies the degree to which the clock in the model is damped in darkness: we require that it is not strongly damped,

$$\delta_{\text{size}} = \sum_{i=L,T} \left(\frac{1}{\langle \Delta c_i^{(m)} \rangle_{ld}} \right)^2 + \left(\frac{\tau_o}{\tau_e} \right)^2 \tag{3.5}$$

The first term introduces a > 1 cost for solutions in LD cycle with oscillation sizes, $(\Delta c_i^{(m)} = {c_i^{(m)}}_{max} - {c_i^{(m)}}_{min})$, less than 1nm, and the second term penalises oscillations that decay too quickly when entering DD as follows: τ_o is a time characterising the decay in the oscillations over the 300h in DD, $\tau_o = -300/\log(({\Delta c_T^{(m)}}_{ld} - {\Delta c_T^{(m)}}_{d})/{\Delta c_T^{(m)}}_{ld})$, and τ_e gives the marginally acceptable decay time, corresponding to TOC oscillation amplitudes that have dropped by $1/4$ over 300h, $-300/\log(0.75)$.

Finally the term δ_{c_L} contains a measure of how broad the peak of LHY mRNA expression is in the proposed solution in LD cycles and is small only if the trace peaks sharply, as observed experimentally. This term is also only small if the peak heights of LHY mRNA expression drop when going from LD to DD,

$$\begin{aligned} \delta_{c_L} \quad = \quad & \sum_{i=2,-2} \langle \left(\frac{2/3 c_L^{(m)}(t_{\text{p}})}{c_L^{(m)}(t_{\text{p}}) - c_L^{(m)}(t_{\text{p}}+i)} \right)^2 \rangle_{ld} + \ldots \\ & \langle \left(\frac{0.05(c_L^{(m)}(t_{\text{p}}-2) - c_L^{(m)}(t_{\text{m}}))}{c_L^{(m)}(t_{\text{m}}) - c_L^{(m)}(t_{\text{m}}+i)} \right)^2 \rangle_{ld} + 10 \left(\frac{\langle c_L^{(m)}(t_p) \rangle_d}{\langle c_L^{(m)}(t_p) \rangle_{ld}} \right)^4 \end{aligned} \tag{3.6}$$

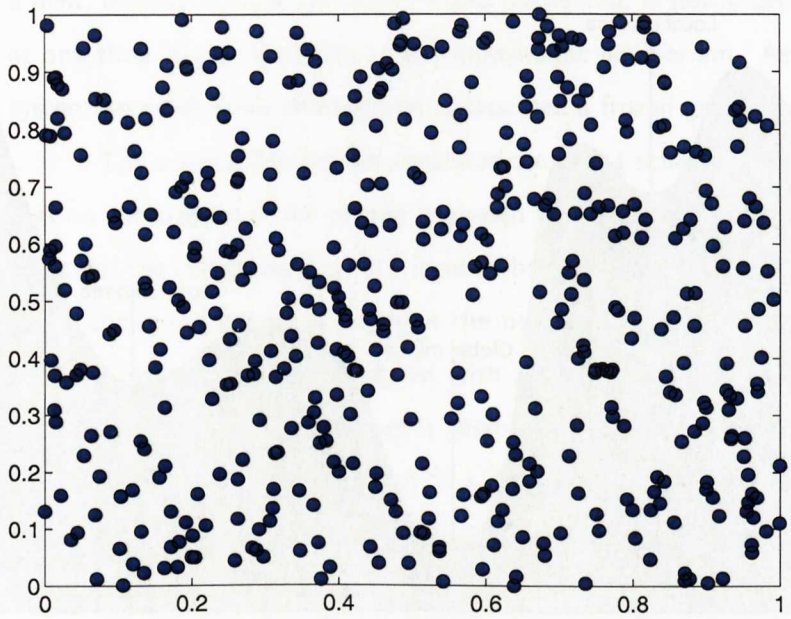
The first term penalises LHY mRNA expression profiles that do not have a sharp peak in LD cycles, with an O(1) contribution if LHY's expression level has dropped by 2/3 of its oscillation size within 2h before and after its peak of expression (at time t_p). The second term checks that LHY mRNA expression has a broad minimum, with an O(1) contribution if 2h before and after the minimum point (at time t_m) LHY's expression has only increased to 5 percent of the level 2 h before LHY's peak. The last term checks that the peak of LHY mRNA expression drops from LD into DD, as it loses its light activation.

Throughout the implementation the cost function was "capped" at $\Delta_{\max} = 10^4$,

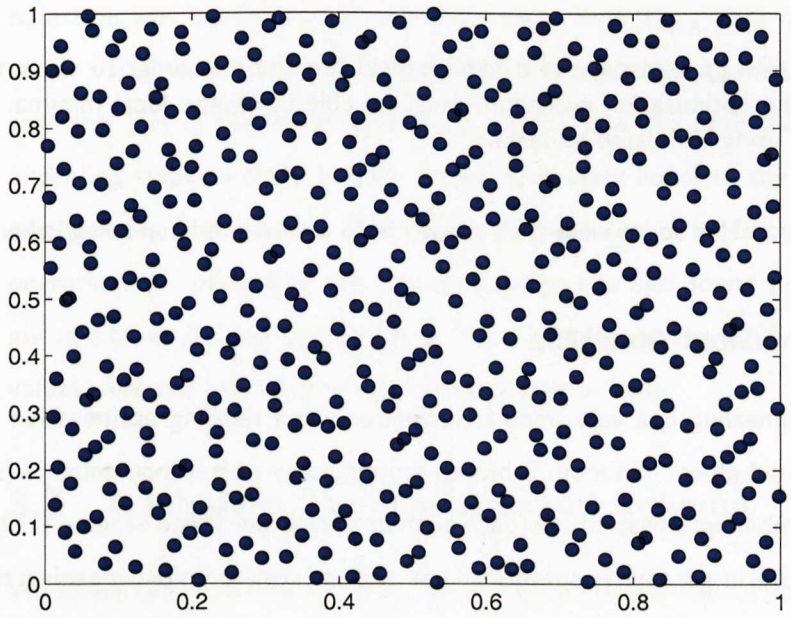
such that $\Delta \rightarrow Min(10^4, \Delta)$. This enabled us to discard solutions that grossly disagree with experimental observations.

3.1.2 Sampling parameter space in order to calculate Δ

Random numbers have many applications including in simulation, cryptography and statistical sampling. There are three main types of random number generators; truly random generators, pseudo-random generators and quasi-random generators [108]. Truly random generators are based on a genuinely random process, such as by tapping into electrical noise. Pseudo-random generators create a series of numbers generated by a deterministic process that mimics a random sequence. This has the advantage of being reproduceable and being computationally undemanding, both crucial features in numerical simulation. Quasi-random generators do not really produce truly random numbers. For some applications, such as evaluating integrals numerically, using random or pseudo-random sequences is not efficient. In our case, we are attempting to search an n-dimensional space for a point that minimises our cost function. One method of achieving this would be to choose points on a grid; the primary issue with this approach is that one has to decide in advance how fine a grid to use. Quasi-random generators pick sample points at random, yet spread out in a self-avoiding way, reducing the chance of clustering that occurs when using a pseudo-random generator. Figure 3.2 shows the advantages of points chosen using the Sobol algorithm, a standard quasi-random generator [109, 110], over random points. We implemented the Antoneev-Saleev variant of the Sobol quasi-random number generator to choose parameter values (vectors) in our $d = 23$ parameter space, adapted from [108]. The initial values as described in [111] were used, allowing number generation in up to 1111 dimensions. We carried out the following change of parameters $\hat{g_1} = {g_1}^a, \hat{g_2} = {g_2}^b, \hat{n_2} = n_2{g_2}^b$, retaining the 'hatted' variables, and then bounded the parameter space for the 23 parameters to be optimised to be bounded $\in [0, 10]$, where our typical parameter scale is unity in nM=hours=1. We calculated Δ for 1 million Sobol points, and the top 50 solutions found from this



(a)



(b)

Figure 3.2: Demonstration of the utility of using Sobol points for more "uniform" sampling. (a) 500 points distributed randomly in two dimensions. (b) 500 Sobol points in two dimensions. Sobol points are chosen in order to avoid the clustering and voids that occur with randomly generated points.

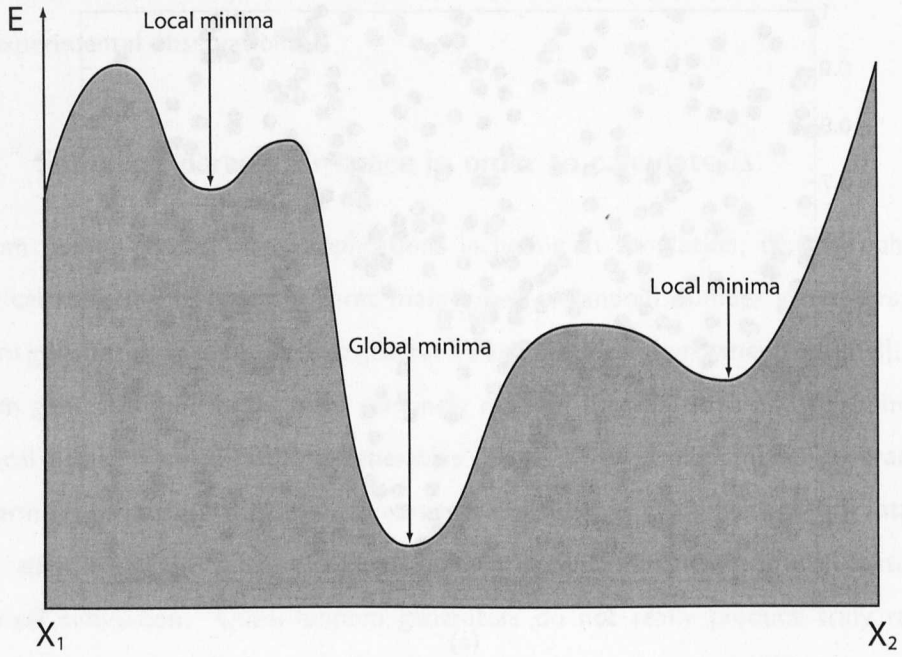


Figure 3.3: Energy landscape of a hypothetical function. In order to reach the global minimum, the optimisation procedure must be able to escape local minima. This can be achieved using simulated annealing.

search of parameter space were then passed onto a simulated annealing scheme.

3.1.3 Simulated Annealing

Simulated annealing is a very popular technique when tackling optimisation problems, where a desired global minimum is hidden among many poorer local minima (see Figure 3.3). A downhill optimisation approach such as the Nelder Mead scheme [112] will only accept a move in parameter space if it has a lower cost function, meaning that it will get stuck in local minima. Simulated annealing is a much more flexable and robust way of identifying minima.

The concept of simulated annealing [113, 114] is based on the manner in which liquids freeze or metals recrystallise in the process of annealing. In an annealing process

a melt, initially at high temperature and disordered, is slowly cooled so that the system at any time is approximately in thermodynamic equilibrium. As cooling proceeds, the system becomes more ordered and approaches a frozen ground state at $T = 0$.

The original Metropolis simulated annealing scheme for minimising a function is that an initial state of the system is chosen at energy E and temperature T. Holding T constant, the initial configuration is perturbed and the change in energy dE is computed. If the change in energy is negative the new configuration is accepted. If the change in energy is positive it is accepted with a probability given by the Boltzmann factor $\exp(-dE/T)$. The temperature is then decremented and the entire process repeated until a frozen state is achieved at $T = 0$. Hence at the beginning of the scheme it is easier to take uphill steps, just as a disordered material has more freedom to re-arrange at high temperature.

We annealed according to a standard Metropolis algorithm [114]. Temperature reduction was carried out linearly each step, from T_{max} to 0 over the 10^5 annealing steps. T_{max} was set as the average cost function value of the best 50 solutions. The step size $|\delta \mathbf{r}|$ in parameter space was set to allow the optimisation routine after 10^5 annealing steps to travel a distance approximately equal to the distance between two neighbouring Sobol points. This approach yields $|\delta \mathbf{r}| = 0.0431$. During the algorithms operation, the parameter values were reset to the best found point if a parameter set giving a lower Δ score had not been found over the previous 10^4 points. The parameter values were not bounded during the annealing process.

3.2 Analysis of the Optimisation Scheme

We solved our system of equations for 10^6 quasi-random Sobol points, each representing a list (vector) of all parameter values (see Figure 3.1 for flow diagram of the optimisation scheme). We then proceeded to calculate the qualitative cost function for these 10^6 random points in parameter space. Encouragingly, the best cost function obtained after N steps appears to converge, see Figure 3.4.

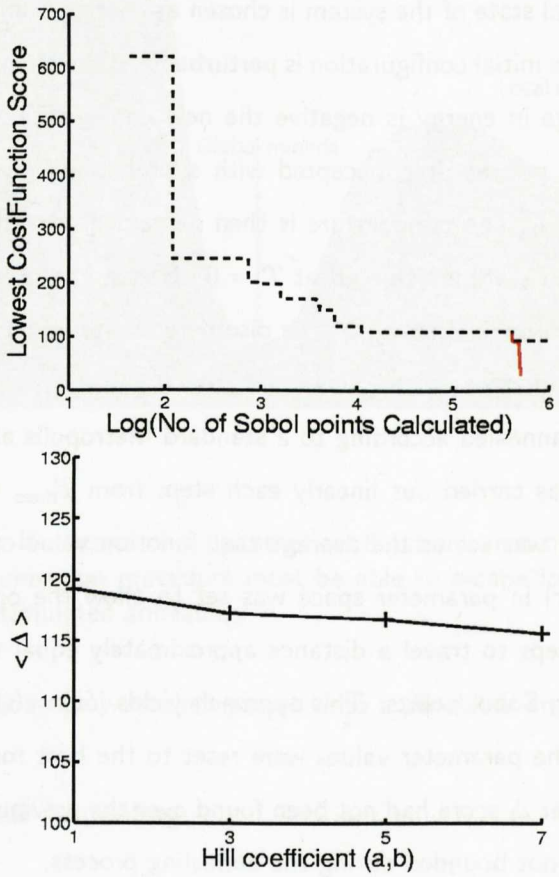


Figure 3.4: Convergence of best cost function with number of Sobol points (dotted line), (Upper pane). Also shown is the trace of the annealing steps leading to the optimum solution, with $\Delta = 27$ (solid red line). The lower pane shows $\langle\Delta\rangle$, the average value of Δ for the best 100 solutions, for varying values of the Hill coefficients, a = b. The average value remains relatively constant, suggesting the value of the Hill coefficients is not crucial for the optimisation of this model.

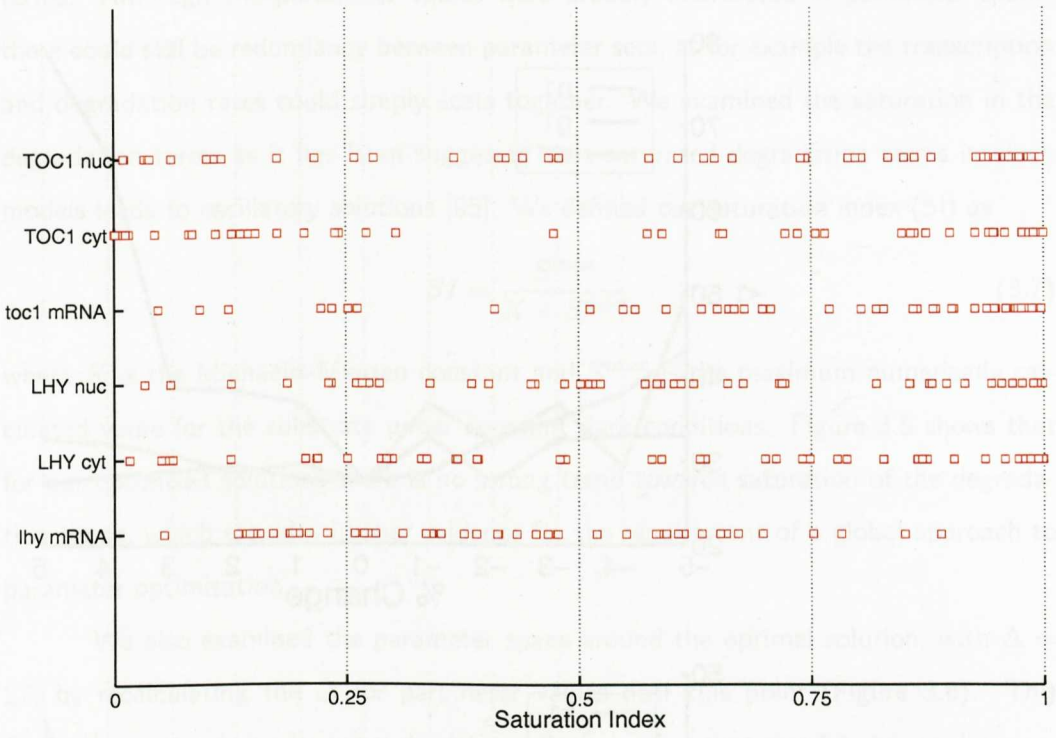


Figure 3.5: The Saturation index, as defined in the text, is shown for degradation terms in the model. A value of 1 shows the term is saturated, 0 shows the term is linear.

Also shown is how the best 100 values from the Sobol sampling vary with increasing Hill coefficients. Interestingly we found that the high Hill coefficients adopted in many previous models are not required to obtain a near optimal solution. The 50 top solutions all have $\Delta < 122$ and were broadly distributed in parameter space (mean of parameter values range from 3.53-7.08, standard deviations from 2.14-3.19). In our scheme the 50 points with the lowest cost function score obtained from the Sobol sampling were then passed to a simulated annealing routine. We then used the 41 annealed solutions with $\Delta < 100$ as diverse but reasonable annealed parameter sets for further analysis in Chapter 2.

We tested whether there was any similarity between the annealed parameter sets by examining the saturation of the Michaelis-Menten kinetics of the degradation

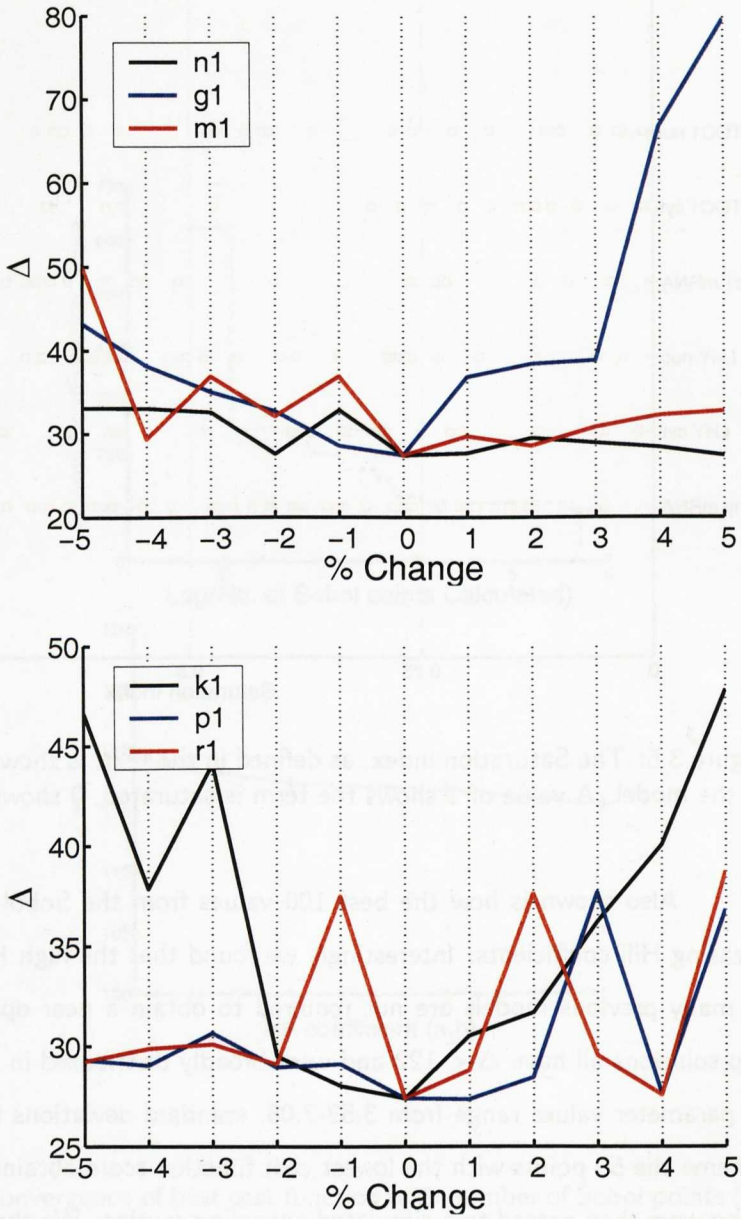


Figure 3.6: Variation of the cost function near the optimal solution of the LHY/CCA1-TOC1 network. Parameters n_1,g_1 and m_1 (Top panel) and parameters k_1, p_1, and r_1 (Bottom panel) are varied by +/- 5% and the resulting variation in Δ is shown. Parameter space can be seen to contain many local minima, supporting our choice of simulated annealing for the parameter optimisation.

terms. Although the parameter values were broadly distributed in parameter space, there could still be redundancy between parameter sets, as for example the transcription and degradation rates could simply scale together. We examined the saturation in the degradation terms as it has been suggested that saturated degradation terms in clock models leads to oscillatory solutions [95]. We defined our saturation index (SI) as

$$SI = \frac{S^{max}}{K + S^{max}} \tag{3.7}$$

where K is the Michaelis-Menten constant and S^{max} is the maximum numerically calculated value for the substrate under constant dark conditions. Figure 3.5 shows that for our optimised solutions there is no strong trend towards saturation of the degradation terms, which provides further evidence for the requirement of a global approach to parameter optimisation.

We also examined the parameter space around the optimal solution, with $\Delta = 27$, by recalculating the Δ for parameter values near this point (Figure 3.6). This figure shows the Δ landscape to be 'choppy', suggesting that simulated annealing was a reasonable choice for our optimisation scheme.

3.3 Discussion

We have developed a robust optimisation scheme that appears to work well with noisy and varied experimental data. The construction of the qualitative cost function allowed us to constrain the behaviour of the model to the consistent, qualitative features of the experimental data. This approach is effective even for noisy or underdetermined systems, when data are scarce and variable (for example, among different laboratories) so direct fitting to quantitative data is undesirable or impossible. It is easy to understand intuitively and can be expanded along with the available data. Terms can be added to the cost function to fit the phenotypes of mutants, for example. The terms in the cost function developed here are specifically based on the experimental data known for the *Arabidopsis* network, and these terms would probably have to be re-formulated to be

applied to a different network. However, a cost function containing terms similar to those as described here should be useful for analysing circadian networks in other organisms. The cost function can potentially combine the qualitative terms, used here, with direct fitting to data. A clock model was recently fitted to a particular set of molecular data for the first time [48], though without specifying how the starting parameter values were chosen. The qualitative terms will no longer be required when sufficient, reliable time-series data are available for fitting. The need for fitting will in turn be reduced as parameter values are measured experimentally. However, a common feature in scientific research is that data tends to be sparse at first, and only later measured extensively under all relevant conditions.

We solved the cost function for a large range of parameter values. A million starting parameter sets were generated quasi-randomly. This approach leaves fewer voids in parameter space than a random distribution. Each parameter was bounded at a maximum value. The 50 parameter sets (points, in parameter space) that gave the lowest cost function scores were taken as starting points for optimisation by simulated annealing. Our method is computationally intensive but achieves a far greater coverage of parameter space than the most extensive manual search, finding good solutions with a maximum Hill coefficient of 2. Optimisation may require fewer parameter steps if these can be directed [115], but our method clearly converges upon parameter values that give an excellent match to the data (Chapter 2).

Chapter 4

Extending the Arabidopsis clock model

Here, we extend the LHY/CCA1-TOC1 network beyond the structures easily inferred from data. This involves the development of new mathematical models that we use to direct further experimentation. To investigate the effect of variations of the network, the outputs of the extended models are compared to published data and to new experiments undertaken by the Millar lab. The biochemical parameter values required in the model are optimised using the method described in Chapter 3. This reduces the possibility that problems with the model are due to a particular set of parameter values, allowing us to examine the properties and failures of the network structure directly.

This chapter is based on work published in 'Extension of a genetic network model by iterative experimentation and mathematical analysis', J C W Locke, M M Southern, L Kozma-Bognar, V Hibberd, P E Brown, M S Turner and A J Millar, *Mol. Syst. Biol.*, **1**:13, (2005).

4.1 Limitations of the LHY/CCA1-TOC1 network

Our previous model of the single-loop LHY/CCA1-TOC1 network (Chapter 2) was able to correctly reproduce the phases of *TOC1* and *LHY* RNA accumulation in WT under light-dark cycle (LD) 12:12. However, simulated *TOC1* mRNA levels remained high until LHY protein accumulated, rather than falling after dusk as observed [64]. This was exaggerated by halving the *LHY* mRNA translation rate in the simulation (representing *lhy* or *cca1* loss-of function mutants), which incorrectly predicted a long-period phenotype. This led us to deduce that there must be another factor responsible for reducing *TOC1* expression, which is not modelled by this network.

Studies of a fusion of a fluorescent protein and TOC1, which allows measurement of the levels of TOC1 by imaging the flourescent protein, suggest an additional limitation [65]. The TOC1 fusion was shown to be close to its minimum abundance before dawn under LD12:12, whereas according to the single-loop LHY/CCA1-TOC1 network, TOC1 should be activating *LHY* transcription maximally at that time [116]. This suggests that either the active form of TOC1 is present at a far lower concentration than bulk TOC1 protein, perhaps in a complex, or that an additional, TOC1-dependent component is the direct activator of *LHY* and *CCA1*.

A third problem is that the LHY/CCA1-TOC1 network did not respond to day length (simulated gene expression profiles were identical in LD cycles with long and short photoperiods, Figure 2.8), whereas it is clear experimentally that the clock has a later phase under longer photoperiods [101, 100]. This limitation occurs because light input to this network is modelled only by the activation of *LHY* expression at dawn, so the model is insensitive to light at the end of the photoperiod. Indeed, *LHY* and *CCA1* expression fall to a low level before the end of a 12h photoperiod [96], so another mechanism is required to mediate light input at the end of the day.

4.2 The LHY/CCA1-TOC1-X network

We extended the single-loop LHY/CCA1-TOC1 network by adding components that would address these limitations, as directed by the experimental data. After each addition, we tested network parameters to establish whether the new network could account for further experimental data. We identified optimal parameters for the most promising of the extended, single-loop models, which we term the LHY/CCA1-TOC1-X network. Firstly, light activation of *TOC1* transcription was included to provide light input at the end of the day and, conversely, to reduce *TOC1* activation immediately after lights-off. Secondly, an additional gene *X* was added to the network after *TOC1*, with nuclear X protein as the immediate activator of *LHY* instead of nuclear TOC1. Thirdly, as the F-box protein ZEITLUPE (ZTL) has been shown to degrade TOC1 protein more effectively during the night [62], we added this factor into our network equations.

We extended the model described in chapter 2, again using Michaelis Menten kinetics and Hill functions to describe degradation and transcription respectively. We used the same form of equation for the acute light response and again grouped *LHY* and *CCA1* as a single gene, *LHY*.

We adopt the following as our mathematical model for the central circadian network: an LHY-TOC1-X feedback loop which involves the cellular concentrations $c_i^{(j)}(t)$ of the products of the i^{th} gene ($i = L$ labels LHY, $i = T$ labels TOC1, $i = X$ labels X) where $j = m, c, n$ denotes that it is the corresponding mRNA, or protein in the cytoplasm or nucleus respectively.

$$\frac{dc_L^{(m)}}{dt} = q_1 c_P^{(n)} \Theta_{\text{light}}(t) + \frac{n_1 {c_X^{(n)}}^a}{{g_1}^a + {c_X^{(n)}}^a} - \frac{m_1 c_L^{(m)}}{k_1 + c_L^{(m)}} \tag{4.1}$$

$$\frac{dc_L^{(c)}}{dt} = p_1 c_L^{(m)} - r_1 c_L^{(c)} + r_2 c_L^{(n)} - \frac{m_2 c_L^{(c)}}{k_2 + c_L^{(c)}} \tag{4.2}$$

$$\frac{dc_L^{(n)}}{dt} = r_1 c_L^{(c)} - r_2 c_L^{(n)} - \frac{m_3 c_L^{(n)}}{k_3 + c_L^{(n)}} \tag{4.3}$$

$$\frac{dc_T^{(m)}}{dt} = \frac{(n_2 + \Theta_{\text{light}}(t)\, n_3)\, {g_2}^b}{{g_2}^b + {c_L^{(n)}}^b} - \frac{m_4 c_T^{(m)}}{k_4 + c_T^{(m)}} \tag{4.4}$$

$$\frac{dc_T^{(c)}}{dt} = p_2 c_T^{(m)} - r_3 c_T^{(c)} + r_4 c_T^{(n)} - ((1 - \Theta_{\text{light}}(t))m_5 + m_6)\frac{c_T^{(c)}}{k_5 + c_T^{(c)}} \tag{4.5}$$

$$\frac{dc_T^{(n)}}{dt} = r_3 c_T^{(c)} - r_4 c_T^{(n)} - ((1 - \Theta_{\text{light}}(t))m_7 + m_8)\frac{c_T^{(n)}}{k_6 + c_T^{(n)}} \tag{4.6}$$

$$\frac{dc_X^{(m)}}{dt} = \frac{n_4 {c_T^{(n)}}^c}{{g_3}^c + {c_T^{(n)}}^c} - \frac{m_9 c_X^{(m)}}{k_7 + c_X^{(m)}} \tag{4.7}$$

$$\frac{dc_X^{(c)}}{dt} = p_3 c_X^{(m)} - r_5 c_X^{(c)} + r_6 c_X^{(n)} - \frac{m_{10} c_X^{(c)}}{k_8 + c_X^{(c)}} \tag{4.8}$$

$$\frac{dc_X^{(n)}}{dt} = r_5 c_X^{(c)} - r_6 c_X^{(n)} - \frac{m_{11} c_X^{(n)}}{k_9 + c_X^{(n)}} \tag{4.9}$$

$$\frac{dc_P^{(n)}}{dt} = (1 - \Theta_{\text{light}})p_4 - \frac{m_{12} c_P^{(n)}}{k_{10} + c_P^{(n)}} - q_2 \Theta_{\text{light}} c_P^{(n)} \tag{4.10}$$

Here the various rate constants n_j, g_j etc parameterise transcription (n_j, g_j), degradation (m_j, k_j), translation (p_j), and the nuclear $\leftrightarrow$ cytoplasmic protein transport (r_j). There is evidence that LHY and CCA1 proteins bind as a dimer to the promoter of *TOC1* [71], and that there is only one active binding site on the *TOC1* promoter [72], so the Hill coefficient for *TOC1* inhibition by LHY protein, b, was set to 2. As there is no experimental evidence to support different values for the Hill coefficients a and c these were also set to 2.

A constant light activation term, $\Theta_{\text{light}}(t)\, n_3$, was added to *TOC1* mRNA production and the effect of ZTL is modelled by the degradation terms for TOC1 in the cytoplasm and the nucleus, which are dark activated as suggested in [65]. An optimal parameter set was found using the methods described in Chapter 3.

Figure 4.1 shows the simulated expression profiles for the LHY/CCA1-TOC1-X network using the optimal set of parameter values (Appendix B). *TOC1* RNA levels peak at dusk in WT under LD12:12, and *LHY* RNA levels at dawn. The model allows *TOC1* mRNA levels to drop before *LHY* levels rise, as observed in experiment. Including gene

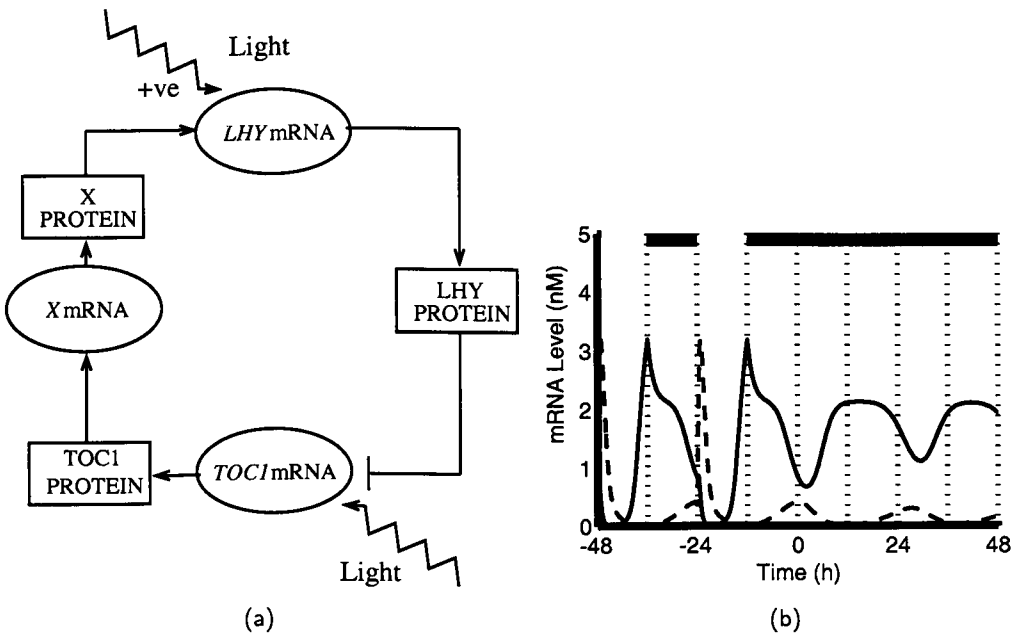


Figure 4.1: The single-loop LHY/CCA1-TOC1-X network. (a) Network diagram. *LHY* and *CCA1* are modelled as a single gene, *LHY* (genes are shown encircled). Nuclear and cytoplasmic protein levels are grouped for clarity (shown boxed) and degradation is not shown. Light acutely activates *LHY* transcription at dawn and activates *TOC1* transcription throughout the day. TOC1 activates a putative gene *X*, which in turn activates *LHY*. Nuclear LHY protein represses *TOC1* transcription. TOC1 degradation in the nucleus and cytoplasm is dark activated, due to data showing ZTL degrades TOC1 more efficiently at night (not shown) (b) Time variation of mRNA levels for the optimal parameter set. In all figures, filled box above panel, dark interval; open or no box, light interval. *LHY* mRNA (dotted line) peaks at dawn in LD12:12 and *TOC1* (solid line) falls after dusk, due to the loss of light activation.

Table 4.1: Comparison of models: period estimates of the LHY/CCA1-TOC1, LHY/CCA1-TOC1-X, and interlocked loop models for the *Arabidopsis* circadian clock network are compared to experimental data. The simulation estimates are the average period over 300h in DD of TOC1 mRNA for WT, *cca1*, and *cca1;lhy* backgrounds. Experimental data for WT and *cca1* mutant periods from [67], *cca1;lhy* data from Figure 4.3.

	WT period (h)	*cca1* Single mutant (h)	*cca1/lhy* mutant (h)
Experimental data (±1 S.E.M)	26.6 (±0.2)	25.4 (±0.2)	18.5 (±0.3)
Interlocked Feedback Model	25.9	25.5	17.0
LHY/CCA1-TOC1-X single loop model	25.9	29.5	Arrythmic
LHY/CCA1-TOC1 single loop model	25.0	29.2	Arrythmic

X within the model permits simulated TOC1 protein levels to fit well with the published data (Figure 4.2). *ztl* mutants were modelled by reducing the degradation rates of TOC1 protein in the cytoplasm and the nucleus by 50%. This results in a long period phenotype, with a period of 32h, similar to or longer than the period of ztl mutants [62]. A prediction of *X* mRNA and protein levels is also possible (Figure 4.2): *X* mRNA peaks in the middle of the night under LD12:12 and nuclear X protein levels peak at dawn. Strong *X* mutants have the same predicted phenotype as the strongest phenotype of *toc1* loss-of-function mutants, causing arrhythmia due to the lack of *LHY* activation. The pattern of *X* mRNA accumulation, and its mutant phenotype, are similar to those of characterised genes such as *ELF4* [78]. However, this model still incorrectly predicts a long period in the simulated *cca1* single mutant (Table 4.1), and the strong, constant light activation of *TOC1* transcription causes several problems; for example, the model becomes arrhythmic under LD cycles with long photoperiods.

4.2.1 Experimental characterisation of the cca1;lhy double mutant

The response of circadian phase to day length [101, 100] strongly suggested that the circadian system receives at least one light input in addition to the activation of *LHY* and *CCA1* expression, yet simulations with the *LHY*/CCA1-TOC1-X network indicated

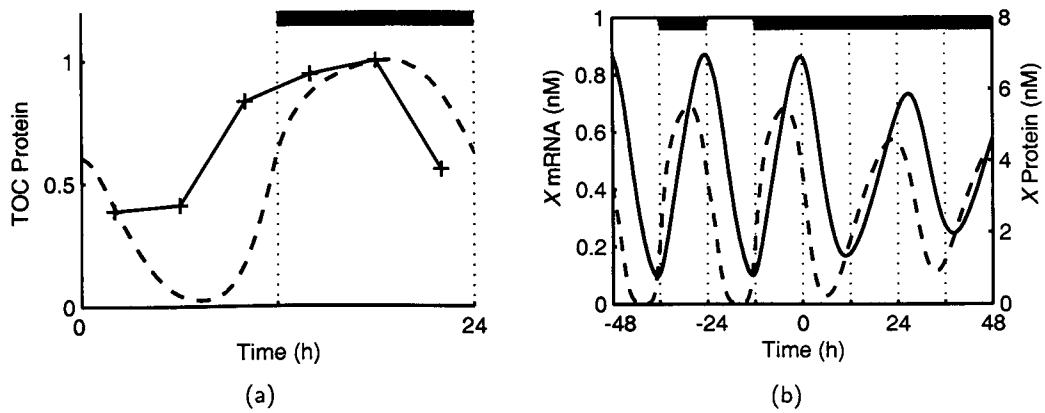
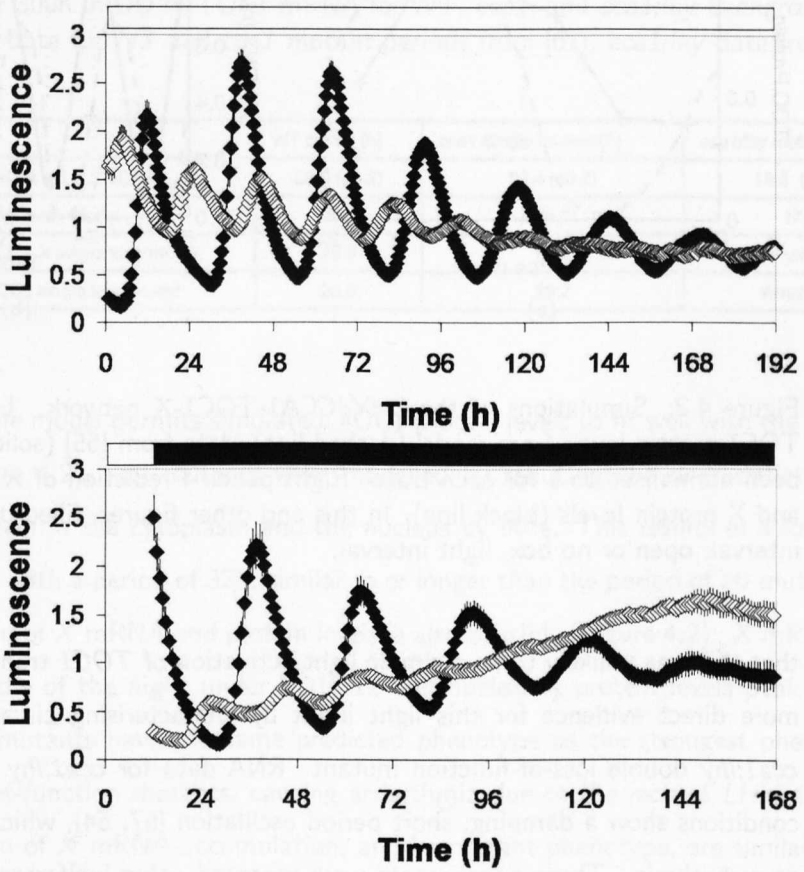


Figure 4.2: Simulations of the LHY/CCA1-TOC1-X network. Left panel: simulated TOC1 protein levels from model (dashed line), data from [65] (solid line). Maxima have been normalised to 1 for each trace. Right panel: Prediction of *X* mRNA (dotted line) and X protein levels (black line). In this and other figures, filled box above panel, dark interval; open or no box, light interval.

that this was unlikely to be a simple light activation of *TOC1* transcription. We sought more direct evidence for this light input by characterising circadian rhythms in the *cca1;lhy* double loss-of-function mutant. RNA data for *cca1;lhy* mutants in constant conditions show a damping, short period oscillation [67, 64], which has been described as arrhythmia. These experiments were repeated using luciferase imaging by Victoria Hibberd (Figure 4.3). In the *cca1;lhy* mutant, promoter activity of *CCA1* and of the clock output genes CCR2 and CAB2 showed an 18h rhythm for at least 3 cycles in constant light (LL), which subsequently lost amplitude. The rhythm is more robust in LL but is also apparent in constant dark (DD) (Figure 4.3). The double mutant retains a regulatory network capable of supporting rhythmic gene expression.

Reproducible entrainment of the double mutant by LD cycles was implicit in previous reports, suggesting that entrainment by light is still possible in the residual network [67, 64], (Figure 4.3). To test this more stringently, László Kozma-Bognár generated a Phase Transition Curve (PTC) for the WT and double mutant (Figure 4.4).



(a)

Figure 4.3: Experimental data for the expression of CCR2:LUC+ in WT (filled diamonds) and *cca1-11;lhy-21* double mutant (open diamonds) plants in LL (top) and DD (bottom). Luminescence of each seedling was normalised to its mean value over the entire timecourse. Data are averages of normalised luminescence from WT seedlings in LL, n=16, in DD n=18, *cca1;lhy* seedlings in LL n=13, in DDn=15. Error bars represent one SEM, often within symbols. This experiment was performed by Victoria Hibberd, see Ref. [117].

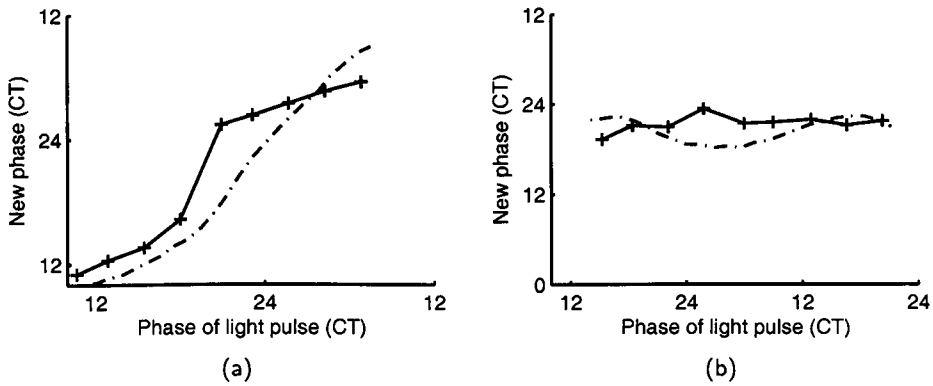


Figure 4.4: Phase Transition Curve for WT (panel a) and *cca1;lhy* double mutant (panel b). Seedlings expressing the CCR2:LUC reporter were grown and entrained for 12 days in LD cycles, and then transferred to DD at the predicted time of lights-off. Luminescence signals were monitored for 5 days in DD. After 24 h in DD, red light pulses (15 μmol $m^{-2}s^{-1}$ for 1h) were administered at 3h intervals. The new phase of the rhythm induced by the light pulses was converted to circadian time (CT (h), 24ths of the period) and plotted against the circadian time of light treatment (Solid lines). Experiment performed by László Kozma-Bognár, see Ref. [117]. Simulated phase responses are represented by dashed lines, and show simulated response of the interlocked feedback loop model to a 1h light pulse. Phase marker for simulation was *TOC1* mRNA peak, compared to CCR2:LUC+ peak in data.

The PTC shows the response of an oscillator to a resetting stimulus and is plotted as the phase to which the oscillator is set ("new phase"), for each phase at which the resetting stimulus is applied ("old phase"). In WT, light pulses induced phase delays during the early subjective night and phase advances during the late subjective night, whereas relatively small phase shifts were elicited during the subjective day. The WT showed a Type 1 (weak) resetting pattern with less than 6h maximal phase shifts, in contrast to the Type 0 (strong) resetting observed in a previous report [88], probably due to the lower fluence of our light stimulus. In contrast, the double mutant showed Type 0 resetting: irrespective of the phase of the light stimulus, the clock was reset to a narrow phase range (circadian time (CT) 20-23). Light input to a residual, rhythmic network remained without LHY and CCA1 function, leading us to add a second, light-responsive feedback loop to produce our next model.

4.2.2 The interlocked feedback loop network

Removing LHY function from the single loop models prevents any oscillation, so none of these models can reproduce the entrainable, damped rhythms observed in *cca1;lhy* plants. We therefore developed an interlocked feedback loop network that is capable of oscillation in simulated *cca1;lhy* double mutants (Figure 4.5). A hypothetical gene Y activates *TOC1* transcription and TOC1 protein represses *Y* transcription, forming a feedback loop. Light input into this loop is possible via transcriptional activation of *Y* rather than of *TOC1*; there is as yet no evidence of direct light activation of *TOC1* [118]. Y was allowed to be both acutely light activated, in the manner of LHY, so to explain the extremely light sensitive response seen experimentally in the PTC of the *cca1;lhy* mutant (Figure 4.3), and to have a constant light activation term, to allow the clock to sense photoperiod. *Y* is also repressed by LHY, as this allowed the network to fit the WT as well as the *cca1;lhy* experimental data. *LHY* therefore acts as a powerful delaying factor in the early day, when it inhibits expression of both *TOC1* and of *Y*. As for the LHY/CCA1-TOC1-X network, The TOC1 degradation term in the cytoplasm

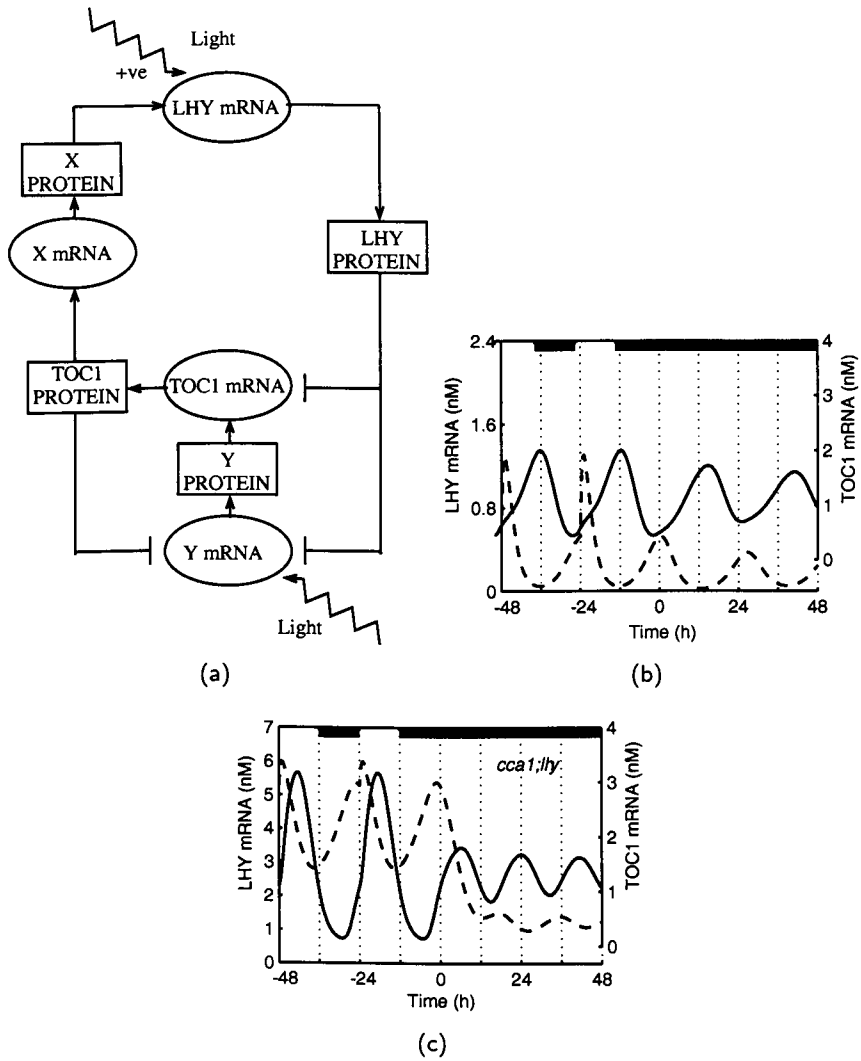


Figure 4.5: The interlocked feedback loop network. (a) Network diagram. This differs from that shown in Figure 4.1 in the following way: here *TOC1* is activated by light indirectly via hypothetical gene *Y*. *Y* activates *TOC1* transcription and both LHY and TOC1 repress *Y* transcription, forming a second feedback loop. (b) Simulation of *LHY* (dashed line) and *TOC1* (solid line) mRNA levels for the optimal parameter set, representing WT (top) and *cca1;lhy* double mutant (bottom) in DD. Translation rate of *LHY* mRNA in simulated mutant is 1/1000 WT value (Reducing the translation rate of *LHY* MRNA to 0 gives the same result). Period of WT in DD is 26h, period of mutant is 17h.

and nucleus was dark activated, to fit with data showing ZTL degrades TOC1 more efficiently at night .

We used a non-cooperative binding term for the activation and repression of $TOC1$ by Y and LHY respectively, and for the repression of Y by TOC1 and LHY. This means that LHY represses $TOC1$ transcription irrespective of the levels of Y. The Hill coefficients a,b,c,d,e in the equations were allowed to vary between 1 and 4 in the optimisation procedure. In order to obtain a compromise between flexibility and overall number of free parameters the Hill coefficients of activation and repression of $TOC1$ were set to the same value, i.e. b = c.

$$\frac{dc_L^{(m)}}{dt} = \Theta_{\text{light}}(t)\, q_1 c_P^{(n)} + \frac{n_1 {c_X^{(n)}}^a}{g_1^a + {c_X^{(n)}}^a} - \frac{m_1 c_L^{(m)}}{k_1 + c_L^{(m)}} \tag{4.11}$$

$$\frac{dc_L^{(c)}}{dt} = p_1 c_L^{(m)} - r_1 c_L^{(c)} + r_2 c_L^{(n)} - \frac{m_2 c_L^{(c)}}{k_2 + c_L^{(c)}} \tag{4.12}$$

$$\frac{dc_L^{(n)}}{dt} = r_1 c_L^{(c)} - r_2 c_L^{(n)} - \frac{m_3 c_L^{(n)}}{k_3 + c_L^{(n)}} \tag{4.13}$$

$$\frac{dc_T^{(m)}}{dt} = \left(\frac{n_2 {c_Y^{(n)}}^b}{g_2^b + {c_Y^{(n)}}^b}\right)\left(\frac{g_3^c}{g_3^c + {c_L^{(n)}}^c}\right) - \frac{m_4 c_T^{(m)}}{k_4 + c_T^{(m)}} \tag{4.14}$$

$$\frac{dc_T^{(c)}}{dt} = p_2 c_T^{(m)} - r_3 c_T^{(c)} + r_4 c_T^{(n)} - ((1 - \Theta_{\text{light}}(t))m_5 + m_6)\frac{c_T^{(c)}}{k_5 + c_T^{(c)}} \tag{4.15}$$

$$\frac{dc_T^{(n)}}{dt} = r_3 c_T^{(c)} - r_4 c_T^{(n)} - ((1 - \Theta_{\text{light}}(t))m_7 + m_8)\frac{c_T^{(n)}}{k_6 + c_T^{(n)}} \tag{4.16}$$

$$\frac{dc_X^{(m)}}{dt} = \frac{n_3 {c_T^{(n)}}^d}{g_4^d + {c_T^{(n)}}^d} - \frac{m_9 c_X^{(m)}}{k_7 + c_X^{(m)}} \tag{4.17}$$

$$\frac{dc_X^{(c)}}{dt} = p_3 c_X^{(m)} - r_5 c_X^{(c)} + r_6 c_X^{(n)} - \frac{m_{10} c_X^{(c)}}{k_8 + c_X^{(c)}} \tag{4.18}$$

$$\frac{dc_X^{(n)}}{dt} = r_5 c_X^{(c)} - r_6 c_X^{(n)} - \frac{m_{11} c_X^{(n)}}{k_9 + c_X^{(n)}} \tag{4.19}$$

$$\frac{dc_Y^{(m)}}{dt} = \left(\Theta_{\text{light}}(t)\, q_2 c_P^{(n)} + \frac{(\Theta_{\text{light}}(t)\, n_4 + n_5) g_5^e}{g_5^e + c_T^{(n)^e}}\right) \times \left(\frac{g_6^f}{g_6^f + c_L^{(n)^f}}\right) - \frac{m_{12} c_Y^{(m)}}{k_{10} + c_Y^{(m)}} \tag{4.20}$$

$$\frac{dc_Y^{(c)}}{dt} = p_4 c_Y^{(m)} - r_7 c_Y^{(c)} + r_8 c_Y^{(n)} - \frac{m_{13} c_Y^{(c)}}{k_{11} + c_Y^{(c)}} \tag{4.21}$$

$$\frac{dc_Y^{(n)}}{dt} = r_7 c_Y^{(c)} - r_8 c_Y^{(n)} - \frac{m_{14} c_Y^{(n)}}{k_{12} + c_Y^{(n)}} \tag{4.22}$$

$$\frac{dc_P^{(n)}}{dt} = (1 - \Theta_{\text{light}}(t))\, p_5 - \frac{m_{15} c_P^{(n)}}{k_{13} + c_P^{(n)}} - q_3 \Theta_{\text{light}}(t)\, c_P^{(n)} \tag{4.23}$$

Optimal parameters for the interlocked feedback loop network (Appendix D) were identified (see Appendix C for details). The optimised model achieved a good fit to experimental results that were specifically encoded into the cost function and hence the optimisation process. This demonstrates that the proposed network is sufficient to explain these data, in contrast to previous models. Simulations of the WT and *cca1;lhy* mutant using the optimal model fit well to experimental RNA expression profiles in DD and in LD12:12 (Figure 4.6(a) and 4.6(b)). For the WT simulation (Figure 4.5), *LHY* mRNA peaks at dawn, *TOC1* at dusk, and the oscillations follow a stable limit cycle with a period of 26h in DD. The simulation of *cca1;lhy* gives a low amplitude oscillation in DD with a 17h period (Figure 4.5), as observed experimentally (Figure 4.3). Under LD12:12, *TOC1* mRNA oscillates with an early peak phase in the double mutant, 5̃h after dawn, as contained within the optimisation cost function. The rhythm of *TOC1* expression in the double mutant also shows a higher amplitude than WT (Figure 4.5), which is observed experimentally [64], but this was not specified explicitly in the cost function during optimisation. Figures 4.6(c) and 4.6(d) show similar expression profiles for simulated and observed [64] *TOC1* mRNA in the WT and *cca1;lhy* mutant under LD16:8 (note that normalisation of data and simulated values obscures the change in amplitude in this figure). *TOC1* mRNA anticipates dawn in the simulation of the *cca1;lhy* double mutant, which has not been so clearly observed in published experimental data and

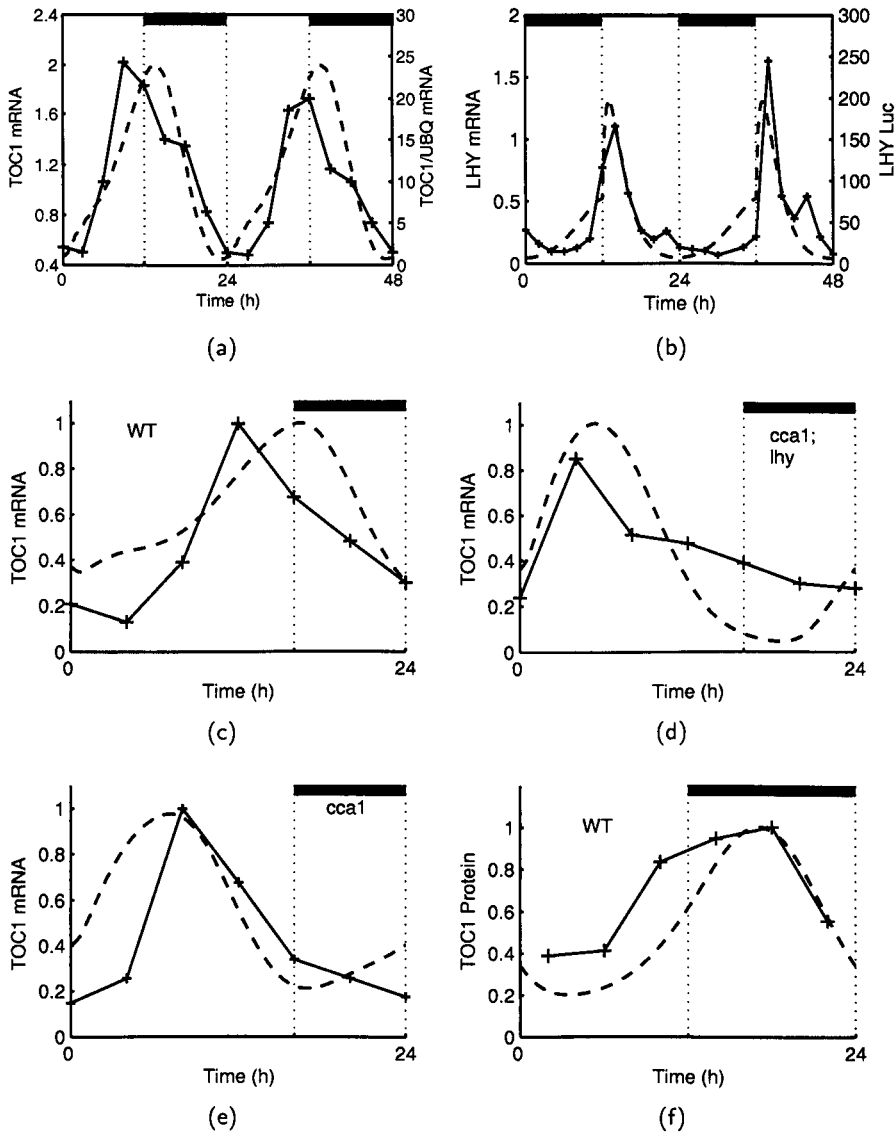


Figure 4.6: Comparison of interlocked feedback loop simulations (dashed line) under LD to data (solid line). (a) *TOC1* mRNA levels in WT plants in LD12:12, left axis; data from [59], right axis. (b) *LHY* mRNA levels in WT plants in LD12:12, left axis; data from [96], right axis. (c, d, e) *TOC1* mRNA levels in WT (c), *cca1;lhy* mutant (*LHY* mRNA translation rate 1/1000th WT value in simulation), (d), *cca1* mutant (*LHY* mRNA translation rate half WT value in simulation) (e) entrained to LD16:8, data from [64]. Highest value of data and simulation is set to 1, for each panel. (f) TOC1 protein levels for WT simulation entrained to LD12:12, TOC1 fusion protein data from [62].

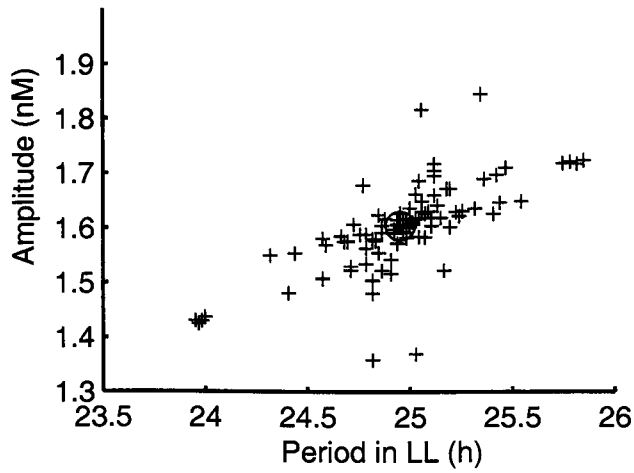


Figure 4.7: Stability analysis of optimal parameter set in the interlocked feedback loop model. The period and amplitude of *TOC1* mRNA oscillations over 300h in LL are calculated for a 5% increase and decrease to each parameter value in turn. The circle represents the period and amplitude of the optimal parameter values.

points to an area for future experimentation.

4.2.3 Analysis and validation of the interlocked feedback loop network

This model with the optimal parameters not only fits these data but its behaviour is also robust to parameter changes. This is widely thought to be a realistic requirement for models of biological regulation, because effective parameter values may be poorly buffered in biology. Changes in the period and amplitude of *TOC1* RNA oscillation under LL were examined after a 5% increase or decrease of each parameter value in turn (Figure 4.7). The resulting change in period varied from 0 to 4%. As for previous clock models [116, 44], some parameters are more sensitive to change than others. The most sensitive parameters are those involved in TOC1 degradation, and X translation and nuclear transport. The period and amplitude of this model are much less sensitive to parameter changes than the single-loop LHY/CCA1-TOC1 model (Fig 2.7), suggesting

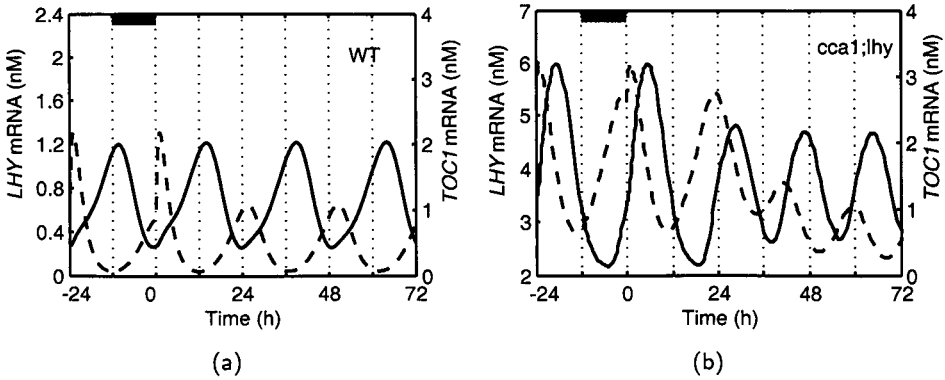


Figure 4.8: Simulations of the interlocked feedback loop network in LD12:12 and LL. (a) mRNA levels for WT in LD12:12 then LL. *LHY* mRNA level peaks at dawn (dotted line) and *TOC1* mRNA level (solid line) at dusk in LD cycles. Period is 25h in LL. (b) mRNA levels for *cca1;lhy* double mutant in LD12:12 then LL. *TOC1* mRNA peaks in the middle of the day and oscillates with an 18h period in LL, as seen experimentally in Figure 4.3.

that some of the weaknesses of this model have been overcome.

Simulations using the optimal parameter set also fit well to several experimental results that were not specified in the optimisation, giving additional support for the proposed network structure. This is the first model that fits well to LL data for *LHY* and *TOC1* mRNA levels. The WT period in LL is correctly shorter (25h) than the period in DD (26h, Figure 4.8) although this effect is less than that observed experimentally. The rhythms in LL generally have a higher amplitude than in DD, as observed. The model correctly predicts the short-period phenotype of *cca1* and lhy single mutants in LL and DD (Table 4.1), and the early phase of *TOC1* RNA expression in the single mutant under LD12:12 (Figure 4.6(e)). The single mutants were simulated by halving the *LHY* mRNA translation rate. Simulated overexpression of LHY produced arrhythmia with low levels of *TOC1* mRNA, as observed in plants that overexpress LHY or CCA1 [72, 66, 69]. Protein levels are also well fitted: simulated LHY protein levels peak 1-2h after *LHY* mRNA levels, as observed [96]. Figure 4.6(f) compares simulated and

measured [62] TOC1 protein levels in WT, showing low levels at dawn in both cases. The optimal parameter set has minimised the light regulation of TOC1 degradation (<1% of total TOC1 degradation), indicating that light-regulated degradation [62] is not required to fit these data. Simulation of a *ztl* mutant by halving the total TOC1 degradation rate results in a 28h period phenotype, again similar to that observed in ztl mutants [62]. The value of a 50% reduction was chosen arbitrarily, but we note that the correct qualitative behaviour of period lengthening with reduced TOC1 degradation is oberved in simulations. A simulated *toc1* mutant results in lower levels of *LHY* mRNA as expected from experiment [72], and simulated TOC1 overexpression is predicted to increase *LHY* mRNA levels. The observed decrease in *LHY* mRNA where TOC1 is overexpressed [59, 65, 119] remains paradoxical.

Simulations of the WT and *cca1;lhy* double mutant PTCs were performed (Dotted trace, Figure 4.4). Both simulations are similar to our experimental data, with a type 1 PTC in the WT, and a type 0 PTC in the double mutant. As expected, the entrained phase of the interlocked feedback model is photoperiod-responsive (Figure 4.10), with simulated mRNA levels peaking later under longer photoperiods, as observed [100, 120]. Light input to *Y* allows the network to respond to light throughout the day. This network will therefore be a good starting point for models of the photoperiod sensor involved in flowering time. The balance of light input to *LHY*, *Y*, and ZTL should now be examined in greater detail to determine how their contributions affect circadian entrainment.

4.2.4 GIGANTEA is a candidate for Y

The interlocked feedback model predicts a distinctive pattern of *Y* mRNA accumulation in the WT and *cca1;lhy* double mutant (Figure 4.11). *Y* mRNA levels peak at the end of the day, but also increase transiently at dawn due to the acute light response of Y transcription. This early expression is quickly repressed by rising LHY protein levels, delaying the peak in *Y* mRNA level until after LHY protein is degraded at the end of the day. *Y* transcription is then repressed as TOC1 protein levels begin to rise during

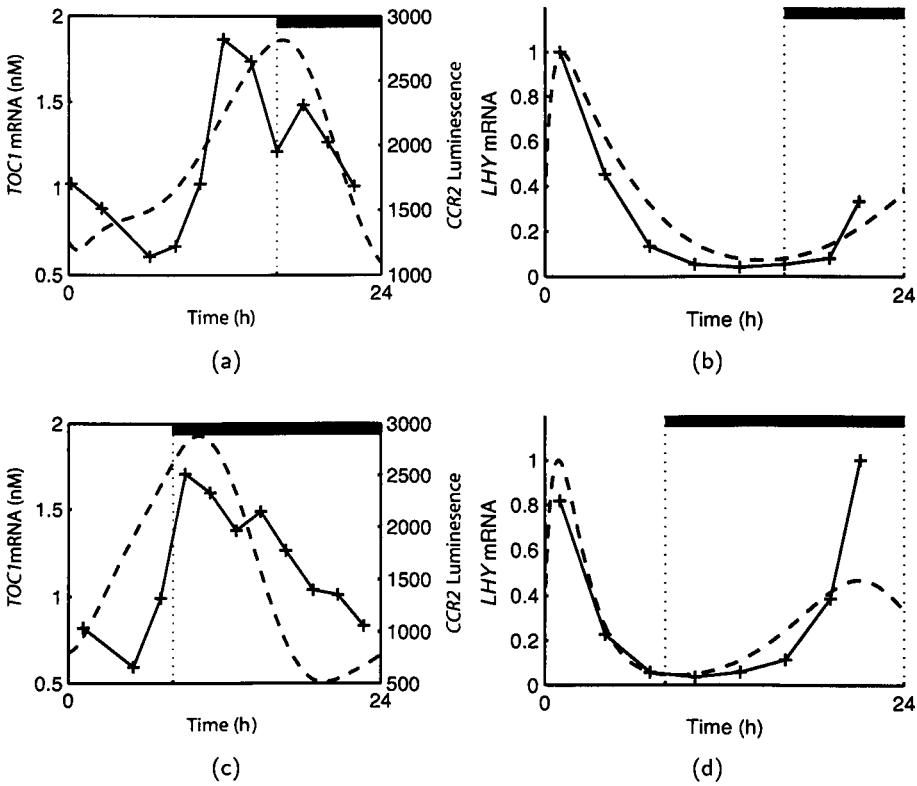


Figure 4.9: Effects of altered photoperiod on circadian rhythms. Simulations using the interlocked feedback network (dashed line), compared to data (solid line, arbitrary concentrations), under LD16:8 (upper panels) and LD8:16 (lower panels). CCR2 luciferase imaging data from [100] is used as a late evening marker to compare to simulations of *TOC1* mRNA (left-hand panels). *LHY* mRNA levels are from [120], highest value of data and simulation is set to 1 (right-hand panels).

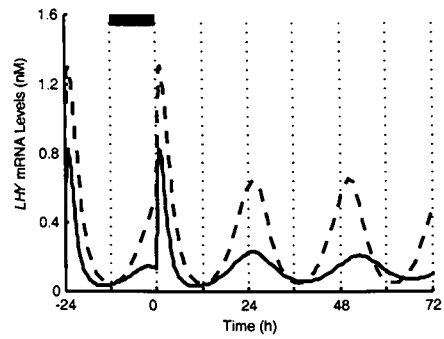


Figure 4.10: Effects of partial *y* loss-of-function in the interlocked feedback network. Simulation of *LHY* mRNA in WT (dashed line) and simulated *gi* mutant (solid line), simulated by halving the *Y* mRNA translation rate compared to WT.

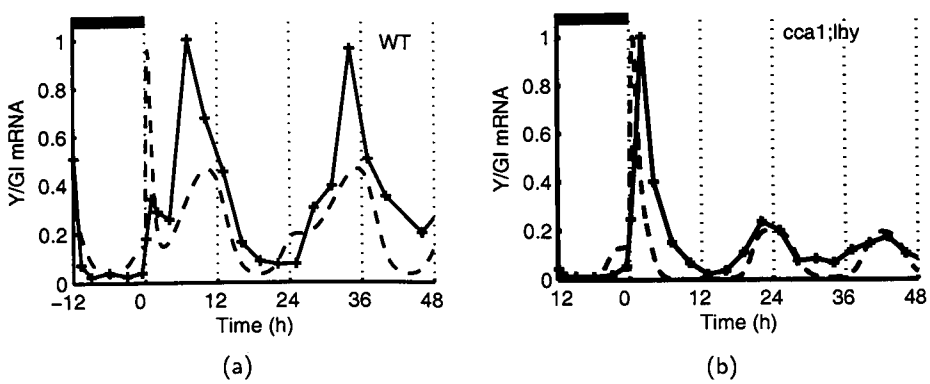


Figure 4.11: *GI* is a candidate gene for *Y*. Simulated *Y* mRNA levels under LD12:12 and LL (dashed line). Data for *GI* mRNA levels (crosses), assayed by quantitative PCR relative to the ACT2 control, from samples harvested at the times indicated. (a) WT. (b) *cca1;lhy*. Highest value of data and simulation is set to 1, for each panel. Experiment performed by Megan Southern, see Ref. [117].

the night (Figure 4.5). In the *cca1;lhy* double mutant, however, the light activation of *Y* at dawn is de-repressed, resulting in a much stronger activation than in WT, and causing *Y* mRNA levels to peak soon after dawn. No gene with this expression pattern had been observed experimentally.

In order to identify *Y*, Megan Southern analysed the transcript abundance of clock-affecting genes with peak RNA levels in the evening in WT and *cca1;lhy* double mutant seedlings. Tissue samples were harvested across the light-dark transitions in one LD cycle, followed by one cycle in LL. *TOC1*, *ELF3*, *ELF4* and *GI* mRNA levels were examined. *GI* was the only gene whose mRNA levels fitted well to our predicted mRNA profiles for *Y* (Figure 4.11). *GI* was shown to be significantly but transiently light activated in the WT and had a very strong light response in the double mutant. The subsequent circadian peak also fitted closely to the prediction for *Y* mRNA, including the 12h phase advance in the mutant relative to WT (Figure 4.11). The tentative identification of *Y* as *GI* allowed us to test whether Y in our model fitted additional, published results for *GI*; indeed, further data do support this proposed function of *GI*. *GI* mRNA is at a low, arrhythmic level in plants that over-express LHY [74] or TOC1 [59]: this agrees with model predictions and occurs because both LHY and TOC1 repress *Y* transcription (Figure 4.5). The sequence of the *GI* promoter includes several evening elements, the putative binding sites for LHY [51]. *gi* loss-of-function mutations result in low amplitude circadian rhythms, with low levels of *LHY* and *CCA1* RNA and either shorter or longer circadian periods [74, 73] or, in some conditions, in arrhythmia [77]. A simulated null mutation of *Y* indeed results in still lower *LHY* transcription and therefore in arrhythmia. If another gene in *Arabidopsis* can substitute for a fraction of *Y* function, then null mutants will avoid arrhythmia. Figure 4.10 shows the oscillation of *LHY* mRNA levels in a simulated partial loss-of-function *y* mutant, where *Y* translation rate has been halved compared to the WT rate. As observed in gi mutants, the oscillations have reduced *LHY* expression and a low amplitude both in LD cycles and in DD [64].

4.3 Discussion

We use a joint experimental and mathematical approach to understand the plant circadian clock, as an example of a regulatory sub-network that is not completely identified. We start from the first proposed feedback loop of the circadian clock mechanism in *Arabidopsis*, the LHY/CCA1-TOC1 network [72], Chapter 2. Comparing model predictions with experimental results, we have progressively incorporated additional components and interactions identified by molecular genetics or inferred from physiological analysis. Our final, interlocking loop model accounts for a greater range of data than the single-loop models, including the entrainable, short period oscillations in the *cca1;lhy* double mutant. In developing this model, we included two putative genes X and Y, and used experiments designed from the model predictions to identify *GI* as a candidate gene for Y.

The prediction of new components is a particularly beneficial outcome from formal modelling of a system that has not been completely identified by experiment. Mathematical models, in contrast to intuitive reasoning, can produce quantitative predictions of dynamic processes that allow detailed experimental design. This was important: the acute light activation of Y in WT was predicted to be very transient (peak 25 min after lights-on, Figure 4.11), allowing us to target our tissue sampling to the appropriate interval, whereas conventional sampling had obscured this induction of *GI* RNA [64]. The interlocked feedback model now highlights the importance of *GI* as a component of light input to the clock, a role that had not previously been emphasised and should now be tested in greater detail. The activation of *TOC1* by GI in an interlocked feedback loop is also a new proposal, which is consistent with the timing of peak *GI* expression before *TOC1*. Mutants that remove both of the loops, such as the lhy; cca1; gi triple mutant should now be tested to determine whether further oscillating sub-networks remain in their absence. A recent study has suggested the existence of a feedback loop between APRR9/APRR7 and *LHY*/CCA1 [84]. Including this loop would not affect our conclusions on the residual network in the *cca1;lhy* double mutant, which would lack

this additional loop. The component(s) that activate *CCA1* and *LHY* at the end of the night remain to be identified: the model predicts the likely accumulation pattern of such a component, *X*. The level of detail in such predictions is obviously limited by the data upon which the model is based, so including statistical measures of uncertainty with the predictions will be increasingly important [115].

Each model makes further, qualitative predictions that appear robust and readily testable. The constant activation of *TOC1* by light reproducibly caused arrhythmia of the LHY/CCA1–TOC1–X model under long photoperiods or constant light, for example, which is not observed in WT plants. This highlights the importance of rhythmic inhibition of the light input [121] which is a wide spread feature of clocks [122, 123]. It is reminiscent of the ELF3-dependent zeitnehmer function observed in *Arabidopsis* [88, 89]. In the interlocked loop model, repression of *Y* by LHY and by TOC1 are sufficient to gate the light activation of *Y*, so we had no justification for further additions to this model. Clearly such models should be interpreted with caution, because undiscovered components cannot be included explicitly. A model that accurately reproduces the regulation of known components is very likely to have captured the relevant effects of the 'hidden' components. The model can advance understanding and make useful predictions but might not capture the real number or mechanism of the hidden components. For example, we model the direct activator of *LHY* and *CCA1* as the product of a TOC1-activated gene, *X*, which could be a minor population of modified TOC1 protein or TOC1-dependent protein complex. We assume that *Y* mediates both the second light input and the additional feedback loop for parsimony, which is now supported by data on *GI*, though these functions could in principle be split among several components.

In order to create more complex models of the circadian clock network, a method to reduce the number of parameters to be optimised needs to be implemented. Our final interlocked loop model has 61 unknown parameters to be optimised, and increasing the complexity of the network by, for example including the output pathway genes, will

greatly increase this number. The best approach to reduce the number of unknown parameters is to measure them directly, although this is both time-consuming and difficult. Another approach is to reduce the number of free parameters in our equations - this is explored in the following chapter.

Chapter 5

Time delay models of the Arabidopsis clock

5.1 Introduction

In the previous chapters we developed a biochemical reactions model of the first multi-gene loop identified in the *Arabidopsis* circadian clock comprising a negative feedback loop, in which two partially redundant genes *LATE ELONGATED HYPOCOTYL* (*LHY*) and *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) repress the expression of their activator, *TIMING OF CAB EXPRESSION 1* (*TOC1*) (LHY/CCA1-TOC1 network, Chapter 2). We showed that this model could not explain all of the current experimental data for the *Arabidopsis* clock, and went on to extend this network in a series of further models, in order to gain a better understanding of the network responsible for the observed experimental data. Our final model consisted of two interlocking feedback loops (Chapter 4).

There is little or no direct biochemical data for the values of the numerous parameters that control the *Arabidopsis* clock. Our interlocked feedback loop model (Chapter 4) has 61 unknown parameters to be estimated by our optimisation scheme. Reduced models, where the equations representing protein translation and transport in

and out of the nucleus are replaced by a time delay [42, 43, 44], have two advantages over biochemical reaction models. Firstly they aid understanding, as they are simpler conceptually than biochemical reaction models. Secondly, the reduction in parameter numbers required to be optimised simplifies the parameter search. In this Chapter we generate time delay models for both the LHY/CCA1-TOC1 one-loop model and the interlocked feedback loop model. The parameters for the model are optimised using the same methods as in Chapter 3. Our aim is to both analyse models that are more simple to understand, and to test the accuracy of an approach that assumes a fixed delay between components of the network.

5.2 LHY/CCA1-TOC1 time delay model

The time delay model of the *Arabidopsis* clock network as outlined in Figure 5.1 required three coupled differential equations to model the central loop, with a total of 12 free parameters. This represents a reduction of 16 parameters and 4 equations compared to our biochemical reaction model (Chapter 2). As before, Michealis-Menten kinetics were used to describe enzyme mediated degradation of proteins, and Hill functions were used to describe the transcriptional activation term of the mRNA for LHY and $TOC1$. We took the following as our mathematical model for the central circadian network: a LHY-TOC1 feedback loop which involves the cellular concentrations $c_i^{(j)}(t)$ of the products of the i^{th} gene ($i = L$ labels LHY, $i = T$ labels TOC1) where $j = m, c, n$ denotes that it is the corresponding mRNA, or protein in the cytoplasm or nucleus respectively.

$$\frac{dc_L^{(m)}}{dt} = \Theta_{\text{light}}(t)\, q_1 c_P^{(n)} + \frac{n_1 c_T^{(n)}(t-\tau_1)^a}{g_1^a + c_T^{(n)}(t-\tau_1)^a} - \frac{m_1 c_L^{(m)}}{k_1 + c_L^{(m)}} \tag{5.1}$$

$$\frac{dc_T^{(m)}}{dt} = \frac{n_2 g_2^b}{g_2^b + c_L^{(n)}(t-\tau_2)^b} - \frac{m_2 c_L^{(m)}}{k_2 + c_T^{(m)}} \tag{5.2}$$

$$\frac{dc_P^{(n)}}{dt} = (1 - \Theta_{\text{light}}(t))\, p_1 - \frac{m_3 c_P^{(n)}}{k_3 + c_P^{(n)}} - q_2 \Theta_{\text{light}}(t)\, c_P^{(n)} \tag{5.3}$$

Here the rate constants parameterise transcription (n_k, g_k) and degradation

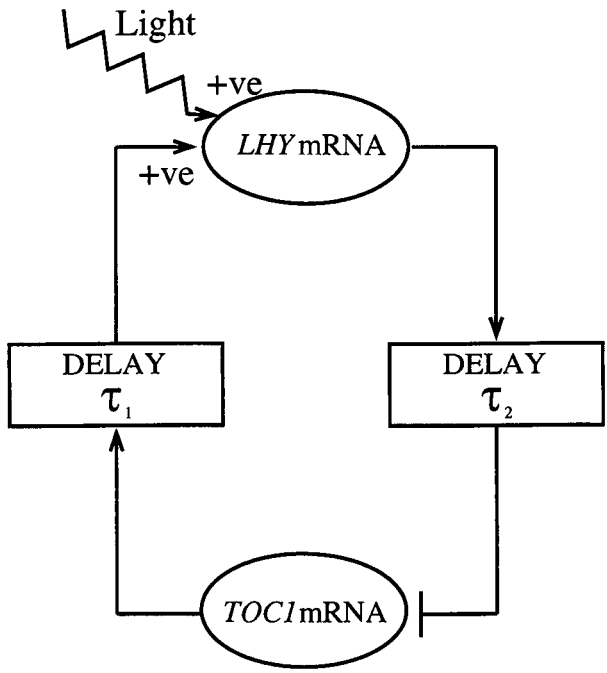


Figure 5.1: Time Delay model for the LHY/CCA1-TOC1 feedback loop. Transcription of the *LHY* gene is light activated, and represses *TOC1* transcription after a delay τ_1, which represents the time taken for the LHY protein to be translated and transported back into the nucleus. *TOC1* mRNA then goes on to activate LHY transcription, after a delay τ_2.

(m_k, k_k). Light was modelled as in Chapter 2, using a simple mechanism involving an interaction of a light sensitive protein P, with concentration $c_P^{(n)}$ with the LHY gene promoter. $\Theta_{\text{light}} = 1$ when light is present, 0 otherwise.

The 12 unknown parameters were estimated using the optimisation method as described in detail in Chapter 3, and briefly outlined here. The equations were numerically solved using the Matlab delay solver dde23 for 10^6 points in parameter space, and the qualitative cost function Δ was calculated (Eqn. 3.1). The delays τ_1 and τ_2 were bound to take values between 30 minutes and 15h, and the other rate constants were bound between 0 and 10 in units where nM=1 and hour=1, as before. The 50 parameter sets with the lowest Δ score were then passed on to a further simulated annealing routine, where the values of the parameters were not constrained. 10^5 annealing steps were calculated for each of the top 50 solutions from the sampling of parameter space.

5.2.1 Results of Parameter search

The 50 parameter sets resulting from the optimisation scheme all had a cost function score $\Delta < 52$. Five of the optimised scores were discarded, as *LHY* and *TOC1* mRNA levels were seen to oscillate in phase in constant conditions in contradiction to experiment. This problem could probably be resolved by the addition of another term to the cost function - and did not occur in the optimisation of the equivalent biochemical reactions model (Chapter 2).

The remaining 45 optimised solutions had very similar values for the delays. τ_1 = 10.22h $\pm$2.187, τ_2 = 0.418h $\pm$0.262. The *LHY* mRNA and *TOC1* mRNA profiles for the parameter set with the lowest cost function score, $\Delta = 24$, is shown in figure 5.2. There are qualitative differences between the output of the optimal parameter set for the delay model and for the biochemical reactions model (See Figure 2.6). The delay model solution fits well to experimental data for *LHY* mRNA with the simulated levels of *LHY* mRNA being correctly low at night. However simulated *TOC1* mRNA levels incorrectly saturate and remain at a high level throughout the night. The oscillations in simulated

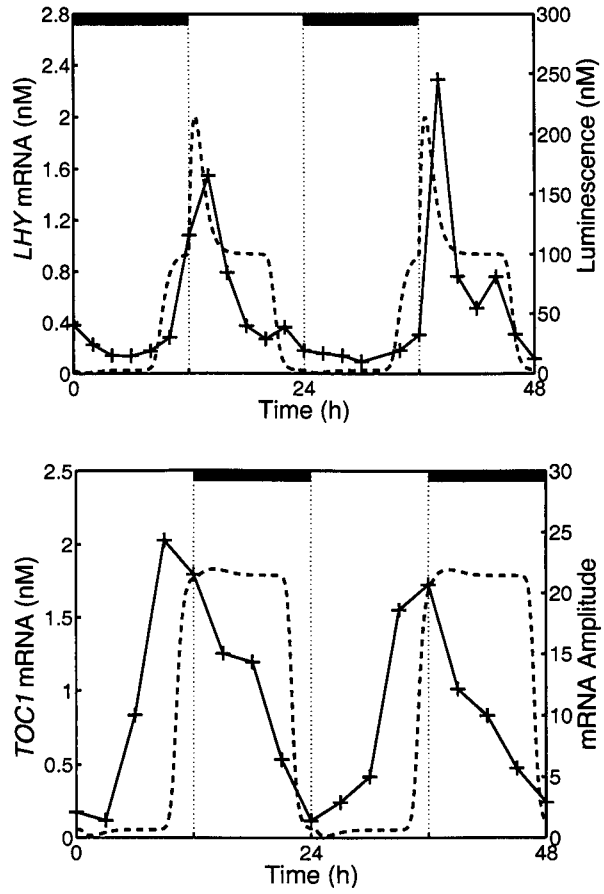


Figure 5.2: Comparison between experimental concentrations (solid curves) with those obtained from our optimal LHY/CCA1-TOC1 time delay model (dashed curves). Simulation models phases of LHY and $TOC1$ mRNA correctly. LHY anticipates dawn in the simulation to a greater degree than experiment. Experimental traces show $TOC1$ levels are falling during the night before LHY levels start to rise, suggesting that our model is missing some factor that would serve to bring $TOC1$ down at the end of the day. $\Delta = 24$. $q_1 = 5.3687\ h^{-1}$, q_2=1.0 h^{-1}, n_1=12.2968 nM/h, n_2=6.7059 nM/h, g_1=7.3107 nM, g_2=0.1826 nM, m_1=9.0131 nM/h, m_2=33.5039 nM/h, m_3=1.2 nM/h,p_1=0.5h^{-1}, k_1=2.5545 nM,k_2=7.3262 nM,k_3=1.2 nM, τ_1=11.2943h, τ_2 = 0.4670h, $a = 1$, $b = 2$.

LHY and *TOC1* mRNA levels under constant dark conditions are also less sinusoidal than seen experimentally. The optimal biochemical reactions model solution showed *TOC1* mRNA expression levels correctly being reduced over night in LD cycles, but only by increasing *LHY* mRNA levels earlier than shown experimentally. Both solutions point to the same result - there must be another mechanism to reduce *TOC1* mRNA levels at night, as *LHY* mRNA is low at this time.

The set of optimised delay models does give some information, as it clearly shows that a long (around 10h) delay is required between *TOC1* transcription and TOC1 activation of *LHY*, and that a short delay is required between *LHY* transcription and TOC1. A comparison of the annealing scheme for the optimised solution for the delay and biochemical reactions model shows that the optimising scheme works faster in the reduced parameter space (Figure 5.3).

We have carried out a stability analysis on our optimal LHY/CCA1-TOC1 delay model (Figure 5.4) as we did for the biochemical reactions network model (Figure 2.7). Here, the only parameters to affect the period of the oscillation were the parameters setting the delay, τ_1 and τ_2. All the remaining parameters only affected the amplitude of oscillation. The period of the oscillator being fixed is a major disadvantage of the time delay model. This can be seen clearly when we attempt a simulated mutant analysis (Figure 5.5), where we double and half the transcription rates of *LHY* and *TOC1* and examine the effect on the period. The period is only very slightly affected by this parameter variation, due to the period being completely set by the time delays. This means that gene dosage effects on period can not be analysed using our time delay model.

5.3 The interlocked loop time delay model

The interlocked loop model (See Chapter 4 for a full description) consists of a feedback loop between *LHY*, which represents the function of both *CCA1* and *LHY* and is acutely light activated, and *TOC1*, and an additional loop between *TOC1* and a proposed gene

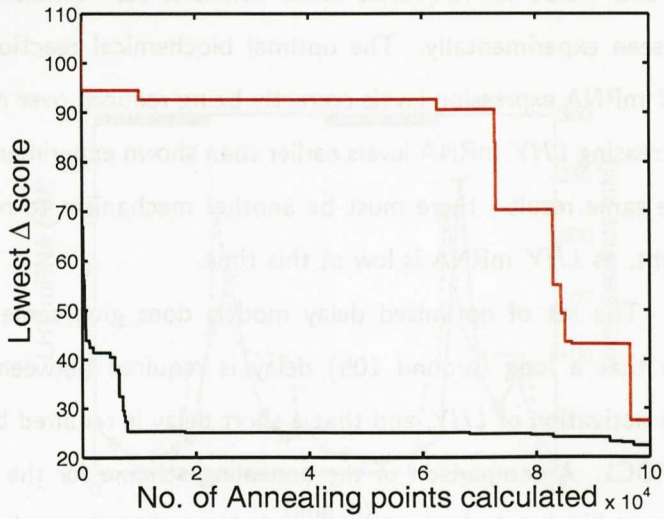


Figure 5.3: Comparison of annealing progress for optimal solution for LHY/CCA1-TOC1 biochemical reaction model (Red line) and for LHY/CCA1-TOC1 time delay model (black line).

Y, which is also light activated. An additional gene *X* is also proposed to be activated by TOC1 and then go on to activate *LHY* transcription, as TOC1 levels are low at dawn when *LHY* transcription is activated. For our delay model (Figure 5.6), 4 delays are used: τ_1 represents the delay between *X* mRNA transcription and *LHY* activation by X, τ_2 represents the delay between *Y* mRNA transcription and *TOC1* activation by Y, τ_3 represents the delay between *LHY* mRNA transcription and *TOC1* and *Y* repression by LHY and τ_4 represents the delay between *TOC1* mRNA transcription and *X* activation and *Y* repression by TOC1.

The time delay model of the interlocked loop network as outlined in Figure 5.6 required five coupled differential equations to model the central loop, with a total of 32 parameters to be fitted. This is a reduction of 31 parameters and 8 equations compared

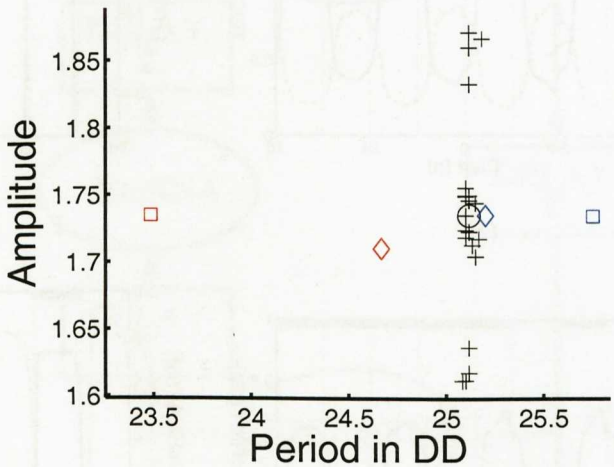


Figure 5.4: Stability analysis on optimal solution for the LHY/CCA1-TOC1 time delay model. The mean period (h) and amplitude (nM) of *TOC1* mRNA over 300h in DD are calculated for a small perturbation, a 5 percent increase and decrease, for each parameter value in turn. The circle represents the period and amplitude of the original parameter values. The squares represent changes in τ_1, the delay between *TOC1* activation and its activation of *LHY*. Diamonds represent changes in τ_2, the delay between *LHY* activation and the repression of *TOC1* (red symbols indicates 5% decrease, blue symbols indicate 5% increase).

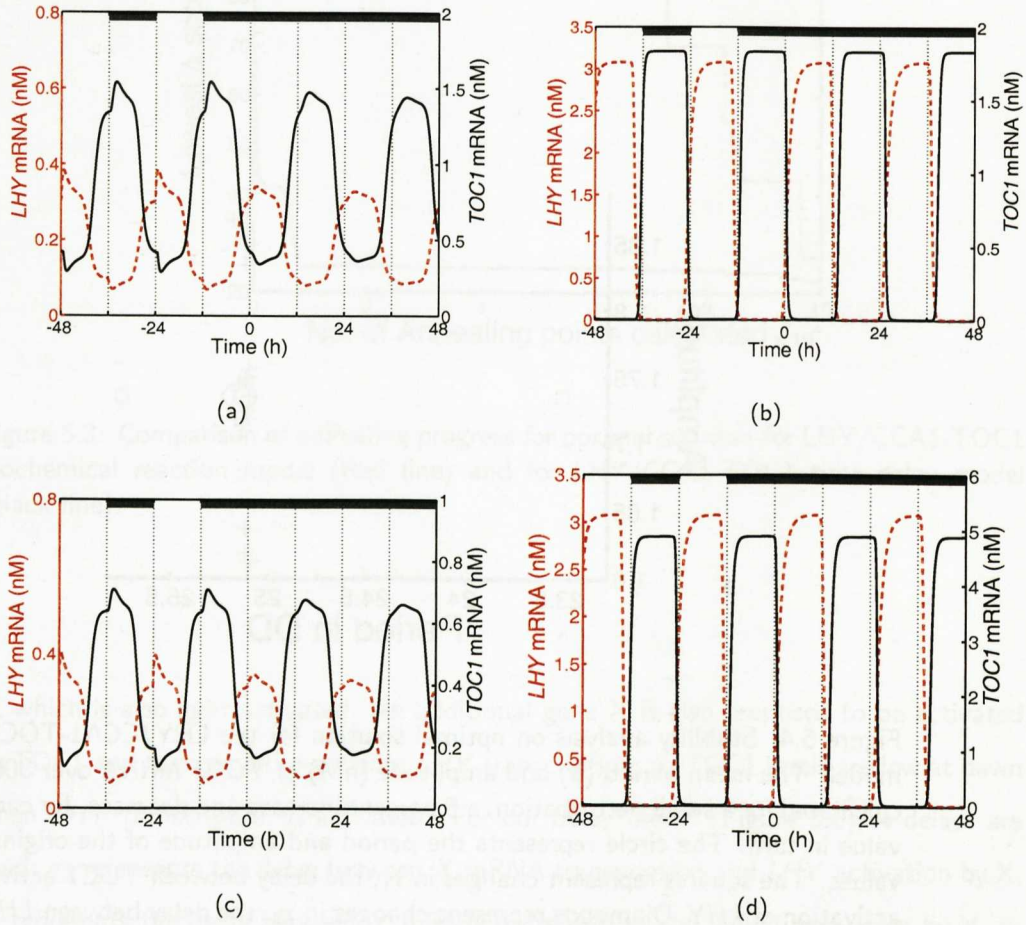


Figure 5.5: Mutant analysis on optimal solution for the LHY/CCA1-TOC1 time delay model. Period of *TOC1* mRNA for WT model is 25.12h. (a) Transcription rate of *LHY*, (p_1) halved, the average period over the 300h in DD, $\tau_d = 25.00$h. (b) p_1 doubled, $\tau_d = 25.04$h. (c) transcription rate of *TOC1*, (p_2), halved, $\tau_d = 24.81$h. (d) p_2 doubled, $\tau_d = 25.17$h.

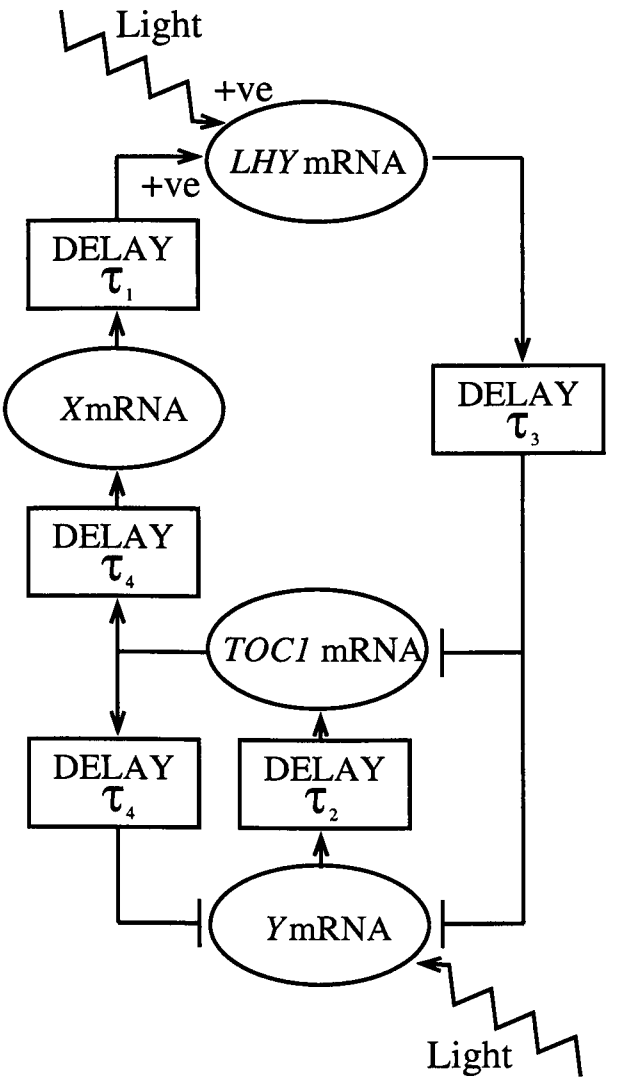


Figure 5.6: Interlocked loop time delay model network. The network differs from that defined in Figure 4.5, as the clock component protein levels have been replaced by explicit delays.

to the previously developed biochemical reaction model (Chapter 4)

$$\frac{dc_L^{(m)}}{dt} = \Theta_{\text{light}}(t)\, q_1 c_P^{(n)} + \frac{n_1 c_X^{(n)}(t-\tau_1)^a}{g_1^a + c_X^{(n)}(t-\tau_1)^a} - \frac{m_1 c_L^{(m)}}{k_1 + c_L^{(m)}} \tag{5.4}$$

$$\frac{dc_T^{(m)}}{dt} = \left(\frac{n_2 c_Y^{(n)}(t-\tau_2)^b}{g_2^b + c_Y^{(n)}(t-\tau_2)^b}\right)\left(\frac{g_3^c}{g_3^c + c_L^{(n)}(t-\tau_3)^c}\right) - \frac{m_2 c_T^{(m)}}{k_2 + c_T^{(m)}} \tag{5.5}$$

$$\frac{dc_X^{(m)}}{dt} = \frac{n_3 c_T^{(n)}(t-\tau_4)^d}{g_4^d + c_T^{(n)}(t-\tau_4)^d} - \frac{m_3 c_X^{(m)}}{k_3 + c_X^{(m)}} \tag{5.6}$$

$$\begin{aligned}\frac{dc_Y^{(m)}}{dt} &= \left(\Theta_{\text{light}}(t)\, q_2 c_P^{(n)} + \frac{(\Theta_{\text{light}}(t)\, n_4 + n_5) g_5^e}{g_5^e + c_T^{(n)}(t-\tau_4)^e}\right) \\ &\times \left(\frac{g_6^f}{g_6^f + c_L^{(n)f}(t-\tau_3)}\right) - \frac{m_4 c_Y^{(m)}}{k_4 + c_Y^{(m)}}\end{aligned} \tag{5.7}$$

$$\frac{dc_P^{(n)}}{dt} = (1 - \Theta_{\text{light}}(t))\, p_1 - \frac{m_5 c_P^{(n)}}{k_5 + c_P^{(n)}} - q_3 \Theta_{\text{light}}(t)\, c_P^{(n)} \tag{5.8}$$

Optimal parameters for the interlocked feedback loop delay model were identified as in Chapter 4 (see Appendix C for details). A preliminary search of 10^4 annealing points was carried out for the forty best parameter sets obtained from calculating the cost function for 10^6 Sobol points.

5.3.1 Results of parameter search

All of the forty optimal solutions that obtained a good fit to both WT and *lhy/cca1* simulations showed an extremely long delay value for τ_3, representing the delay between *LHY* mRNA transcription and *TOC1* and *Y* repression by LHY. A typical solution is shown in Figure 5.7. *LHY* mRNA levels are correctly peaking at dawn and *TOC1* levels are correctly peaking at dusk in the WT simulations and *GI* are acutely activated at dawn and oscillate with a short period under constant conditions in the *cca1;lhy* simulation. However in this case τ_3 takes the value of 18.98h. This length of delay was not possible with the equivalent bio-chemical reaction network, and is not backed up by experimental data for LHY levels, which peak soon after *LHY* mRNA levels at the beginning of the

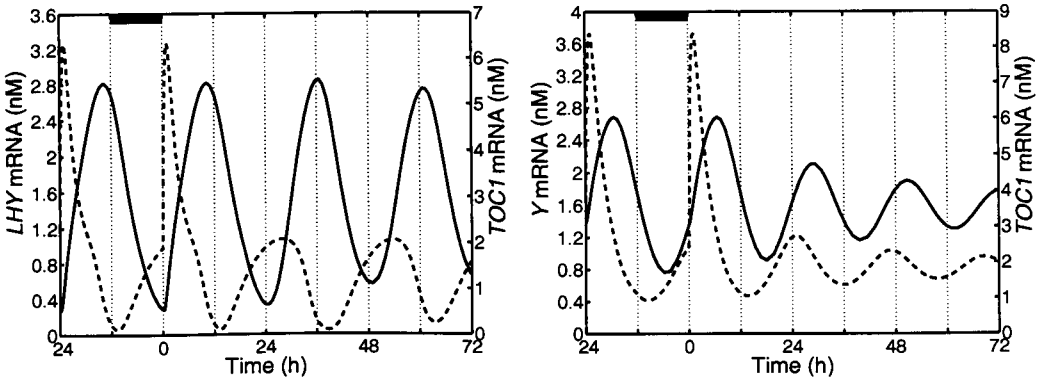


Figure 5.7: Simulated mRNA levels for WT and *cca1;lhy* backgrounds using the optimal interlocked loop delay model. (a) *LHY* mRNA levels (dotted line, left axis) and *TOC1* mRNA levels (black line, right axis) for WT simulated solution. (b) *Y* mRNA levels (dotted line, left axis) and *TOC1* mRNA levels (black line, right axis) for *lhy;cca1* simulated solution. Parameters are $\tau_1 = 6.7073$h, $\tau_2 = 0.2723$h, $\tau_3 = 18.9823$h and $\tau_4 = 5.3419$h. $q_1 = 7.6517\ h^{-1}$, q_2=4.8681 h^{-1}, n_1=0.9704 nM/h, n_2=1.6243 nM/h, n_3= 2.9133nM/h, n_4=0.2116 nM/h, n_5=1.1241 nM/h, g_1=1.7837 nM, g_2=0.8356 nM, g_3=1.9081 nM, g_4=1.1387 nM, g_5=2.5948 nM, g_6=1.3836 nM, m_1=7.5432 nM/h, m_2=5.2839 nM/h, m_3= 2.8703 nM/h, m_4=5.6216 nM/h, p_1=0.5h^{-1}, k_1=9.93 nM, k_2=19.9658 nM, k_3=0.4766 nM, k_4=8.1527 nM, a=2.0611, b=2.8744, c=2.8744, d=1.3959, e=1.4907, f=1.5126.

day [96]. This suggests that for future optimisations, τ_3 should be constrained to be less than $5h$ to better agree with experimental results.

5.4 Discussion

Our time delay model for the LHY/CCA1-TOC1 network is easy to understand conceptually, and clearly shows that another factor is required to reduce *TOC1* mRNA levels at night. The estimated delay of around $10h$ between *TOC1* transcription and *LHY* activation by TOC1 required for the clock components to oscillate with the correct phase is of interest as it is unlikely that such a long delay is created just by the transport of TOC1 protein into the nucleus. However, there is a limit to the scope of analysis possible

using the time delay model. Due to the fixed nature of the delays our stability analysis (Figure 5.4) and simulated mutant experiment (Figure 5.5) were not as informative as for the equivalent biochemical reactions model (Chapter 2).

Our preliminary parameter search for the interlocking loop delay model suggests that it may be possible to generate a delay model comparable in accuracy to the equivalent biochemical reactions model (Chapter 4). Our optimisation should be repeated with further constraints on τ_3.

The reduction in parameter values and equations made the computational optimisation time of our time delay models faster than the biochemical reaction models. Time delay models could be used in future to quickly test different network topologies before generating a complete biochemical reactions model of the more promising topologies found. However the greater speed in optimisation needs to be balanced against the time required to develop two separate models of the same network.

Chapter 6

The 3-loop Arabidopsis clock model

In this Chapter we extend our interlocked feedback loop model to include the recently proposed feedback loop between PSEUDO-RESPONSE REGULATOR 7 (PRR7) and 9 (PRR9), and LHY/CCA1 [84]. We use the resulting 3-loop network to further test our prediction that *GI* acts in a feedback loop with *TOC1*, and investigate the possibility of the feedback loops in the clock de-coupling. This work does not affect our predictions for the residual network in the *cca1;lhy* double mutant in Chapter 4, which lacks the additional loop.

6.1 Introduction

The interlocked loop model (Chapter 4) was able to fit to several key pieces of experimental data. However, there were experimental results that the model could not reproduce. A reduction in Y translation rate caused the period of our model to increase slightly (Figure 4.10). Although one mutant allele of *GI*, *gi-2*, was seen to have a long period expression of CAB compared to WT [73], all other mutant alleles cause the clock to go short-period [76, 64, 74]. This causes a problem with our prediction that GI is a

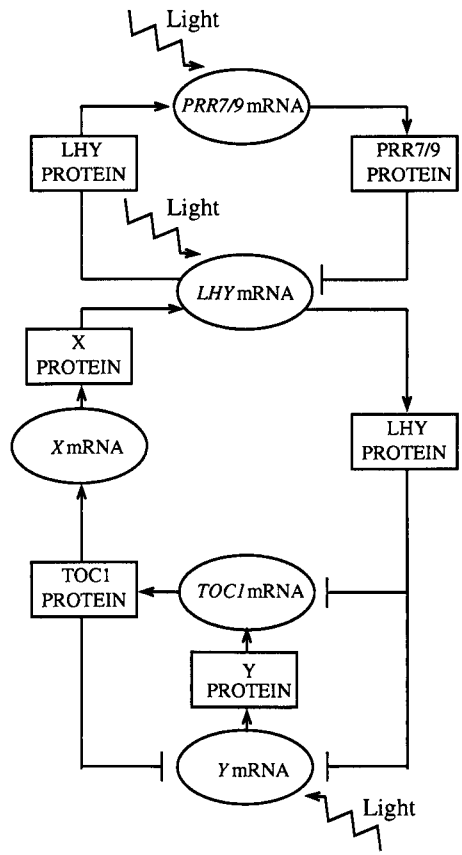


Figure 6.1: 3-loop model network diagram. *LHY* and *CCA1* are modelled as a single gene, *LHY* (genes are shown encircled). Nuclear and cytoplasmic protein levels are grouped for clarity (shown boxed) and degradation is not shown.

component of Y in our network. Also, *toc1* mutant plants show a residual short-period oscillation [72, 62] similar to that seen in a *cca1;lhy* plant (Chapter 4). This cannot be reproduced using the interlocked loop model, as *toc1* mutants are arrhythmic in our simulations.

A feedback loop has recently been proposed to act between *LHY/CCA1* and the genes *PRR7* and *PRR9* [84, 83]. LHY and CCA1 are proposed to activate *PRR7* and *PRR9* transcription. The evidence for this is that in LHY or CCA1 overexpressing plants *PRR7* and *PRR9* mRNA levels are elevated compared to WT, and *PRR7* and *PRR9* levels are low in *lhy*, *cca1* backgrounds, and extremely low in *lhy/cca1* plants [84]. CCA1 has also been shown to specifically bind to the *PRR7* and *PRR9* promoters. PRR7 and PRR9 are proposed to repress *LHY* and *CCA1* transcription forming an additional feedback loop. The evidence for this repression is less strong, although *LHY* and *CCA1* levels are expressed at a higher level in a *prr7/prr9* plant. In this chapter we add the LHY/CCA1-PRR7/9 feedback loop to our network equations and analyse the effect this has on our simulations

6.2 Computational Methods

We have built upon our network equations for the proposed interlocked feedback loop model for the *Arabidopsis* circadian clock [117] (see Chapter 4) to include the LHY/CCA1-PRR7/9 loop.

The interlocked loop model builds on the network described in Chapter 4, namely a feedback loop between *LHY*, which represents the function of both *CCA1* and *LHY* and is acutely light activated, and *TOC1*, and an additional loop between *TOC1* and a proposed gene *Y*, which is also light activated. An additional gene *X* is also proposed to be activated by TOC1 and then go on to activate *LHY* transcription, as TOC1 levels are low at dawn when *LHY* transcription is activated. We have added to this network an additional loop; *PRR7* and *PRR9* transcription is proposed to be activated by LHY/CCA1, and then PRR7 and PRR9 go on to repress *LHY* and *CCA1* transcription

[84]. We denote this as the 3-loop model for the clock (Figure 6.1).

We incorporated the *PRR7/9 - LHY/CCA1* feedback loop into our clock model as follows. *PRR9* [124] and *PRR7* [125] are two genes that peak at the beginning and middle of the day respectively. As there is little experimental evidence to distinguish their roles in the clock we modelled them with a single gene, called *PRR7/9* (Eqns.6.14 to 6.16). *PRR7/9* transcription is given both an acute light activation term, as *PRR9* has been seen to be acutely light activated experimentally [124], and a constant light activation term of the form previously used for *Y*, and is activated by nuclear LHY protein . Future work should explore whether it is possible for the acute light induction of *PRR9* transcription seen at dawn to be indirect, through the light induction of LHY/CCA1.

We modified our terms for *LHY* mRNA levels to include the role of PRR7/9 (Eqn.6.1). PRR7/9 represses both *LHY*'s light activation and the activation by *TOC1*. In addition to the acute light response, we gave *LHY* mRNA levels a constant light activation term $\Theta_{\text{light}}(t)\, n_0$ as *LHY* transcription appears to be light activated throughout the day in an *prr7/prr9* plant lacking functioning genes for *PRR7/9* [84]. $\Theta_{\text{light}} = 1$ when light is present, 0 otherwise.

We took the following as our mathematical model for the central circadian network, which involves the cellular concentrations $c_i^{(j)}(t)$ of the products of the i^{th} gene ($i = L$ labels LHY, $i = T$ labels TOC1, $i = X$ labels X, $i = Y$ label Y, $i = A$ labels PRR7/9) where $j = m, c, n$ denotes that it is the corresponding mRNA, or protein in the cytoplasm or nucleus respectively Eqns. 6.4 to 6.13 are unchanged from the interlocking feedback loop model in Chapter 4 but are listed here for completeness.

$$\frac{dc_L^{(m)}}{dt} = \left(\frac{g_0^\alpha}{g_0^\alpha + c_A^{(n)\alpha}}\right)\left(\Theta_{\text{light}}(t)\left(q_1 c_P^{(n)} + n_0\right) + \frac{n_1 c_X^{(n)a}}{g_1^a + c_X^{(n)a}}\right) - \frac{m_1 c_L^{(m)}}{k_1 + c_L^{(m)}} \tag{6.1}$$

$$\frac{dc_L^{(c)}}{dt} = p_1 c_L^{(m)} - r_1 c_L^{(c)} + r_2 c_L^{(n)} - \frac{m_2 c_L^{(c)}}{k_2 + c_L^{(c)}} \tag{6.2}$$

$$\frac{dc_L^{(n)}}{dt} = r_1 c_L^{(c)} - r_2 c_L^{(n)} - \frac{m_3 c_L^{(n)}}{k_3 + c_L^{(n)}} \tag{6.3}$$

$$\frac{dc_T^{(m)}}{dt} = \left(\frac{n_2 c_Y^{(n)b}}{g_2^b + c_Y^{(n)b}}\right)\left(\frac{g_3^c}{g_3^c + c_L^{(n)c}}\right) - \frac{m_4 c_T^{(m)}}{k_4 + c_T^{(m)}} \tag{6.4}$$

$$\frac{dc_T^{(c)}}{dt} = p_2 c_T^{(m)} - r_3 c_T^{(c)} + r_4 c_T^{(n)} - ((1 - \Theta_{\text{light}}(t))m_5 + m_6)\frac{c_T^{(c)}}{k_5 + c_T^{(c)}} \tag{6.5}$$

$$\frac{dc_T^{(n)}}{dt} = r_3 c_T^{(c)} - r_4 c_T^{(n)} - ((1 - \Theta_{\text{light}}(t))m_7 + m_8)\frac{c_T^{(n)}}{k_6 + c_T^{(n)}} \tag{6.6}$$

$$\frac{dc_X^{(m)}}{dt} = \frac{n_3 c_T^{(n)d}}{g_4^d + c_T^{(n)d}} - \frac{m_9 c_X^{(m)}}{k_7 + c_X^{(m)}} \tag{6.7}$$

$$\frac{dc_X^{(c)}}{dt} = p_3 c_X^{(m)} - r_5 c_X^{(c)} + r_6 c_X^{(n)} - \frac{m_{10} c_X^{(c)}}{k_8 + c_X^{(c)}} \tag{6.8}$$

$$\frac{dc_X^{(n)}}{dt} = r_5 c_X^{(c)} - r_6 c_X^{(n)} - \frac{m_{11} c_X^{(n)}}{k_9 + c_X^{(n)}} \tag{6.9}$$

$$\frac{dc_Y^{(m)}}{dt} = \left(\Theta_{\text{light}}(t)\, q_2 c_P^{(n)} + \frac{(\Theta_{\text{light}}(t)\, n_4 + n_5) g_5^e}{g_5^e + c_T^{(n)e}}\right) \times \left(\frac{g_6^f}{g_6^f + c_L^{(n)f}}\right) - \frac{m_{12} c_Y^{(m)}}{k_{10} + c_Y^{(m)}} \tag{6.10}$$

$$\frac{dc_Y^{(c)}}{dt} = p_4 c_Y^{(m)} - r_7 c_Y^{(c)} + r_8 c_Y^{(n)} - \frac{m_{13} c_Y^{(c)}}{k_{11} + c_Y^{(c)}} \tag{6.11}$$

$$\frac{dc_Y^{(n)}}{dt} = r_7 c_Y^{(c)} - r_8 c_Y^{(n)} - \frac{m_{14} c_Y^{(n)}}{k_{12} + c_Y^{(n)}} \tag{6.12}$$

$$\frac{dc_P^{(n)}}{dt} = \left(1 - \Theta_{\text{light}}(t)\right) p_5 - \frac{m_{15} c_P^{(n)}}{k_{13} + c_P^{(n)}} - q_3 \Theta_{\text{light}}(t)\, c_P^{(n)} \tag{6.13}$$

$$\frac{dc_A^{(m)}}{dt} = \Theta_{\text{light}}(t)\left(q_4 c_P^{(n)}\right) + \frac{n_6 c_L^{(n)g} + \Theta_{\text{light}}(t)\, n_7}{g_7^g + c_L^{(n)g}} - \frac{m_{16} c_A^{(m)}}{k_{14} + c_A^{(m)}} \tag{6.14}$$

$$\frac{dc_A^{(c)}}{dt} = p_6 c_A^{(m)} - r_9 c_A^{(c)} + r_{10} c_A^{(n)} - \frac{m_{17} c_A^{(c)}}{k_{15} + c_A^{(c)}} \tag{6.15}$$

$$\frac{dc_A^{(n)}}{dt} = r_9 c_A^{(c)} - r_{10} c_A^{(n)} - \frac{m_{18} c_A^{(n)}}{k_{16} + c_A^{(n)}} \tag{6.16}$$

Here the various rate constants n_j, g_j etc parameterise transcription (n_j, g_j), degradation (m_j, k_j), translation (p_j), and the nuclear $\leftrightarrow$ cytoplasmic protein transport (r_j). The Hill coefficients are represented by α, a, b, c, d, e, f, g. The acute light response was modelled as in Chapters 2 and 4. A preliminary parameter search was carried out, see Appendix E for details. We took as our starting parameters the optimal solution for the interlocked loop model. We then treated this network as representing a *prr7/prr9* double mutant plant and re-optimised the parameters for *LHY* transcription in order to give a 30h period in LL. We then modified the interlocked loop equations to include *PRR7/9*, and optimised the *PRR7/9* parameters to give an oscillation with a period of 24 hours in constant light for a WT plant, and a short period ($< 20h$) oscillation in a *toc1* plant (Appendix E).

6.3 Results of Parameter Search

The optimum model from our parameter search (Appendix F) gives a good fit to WT data similar to that seen for the interlocking loop model, see Figure 6.2. *LHY* levels peak at dawn and are low during the night, as shown by experiment, and *TOC1* levels correctly peak at dusk.

The PRR7/PRR9 - LHY/CCA1 loop is capable of generating the short-period

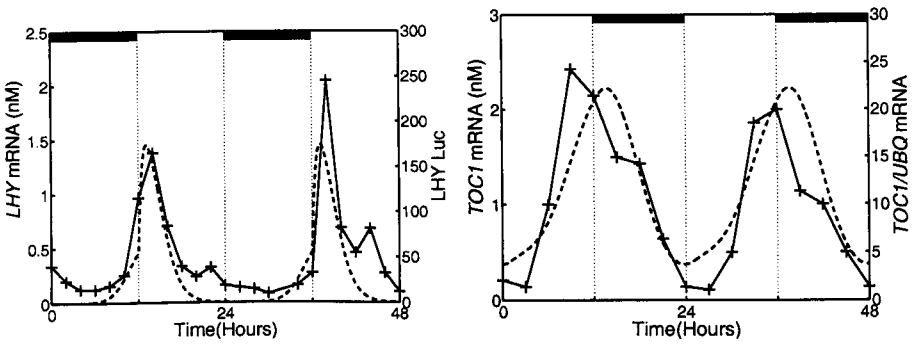


Figure 6.2: Comparison of simulations using the optimal 3-loop model (dashed line) under LD to data (solid line). (a) *LHY* mRNA levels in WT plants entrained to LD12:12, left axis; data from [96], right axis. (b) *TOC1* mRNA levels in WT plants entrained to LD12:12, left axis; *TOC1* mRNA levels relative to UBIQUITIN (UBQ) from [59], right axis.

rhythm of *toc1* mutant plants [62] (see Figure 6.3), and its absence results in the 30h period of *prr7/prr9* double mutants [84], (see Figure 6.4).

We used our 3-loop model to further investigate our prediction that *GI* is a component of *Y*. A simulation of a *gi* mutation using the interlocked feedback loop model caused the period of the clock to increase slightly (Chapter 4). A simulated *gi* mutation using the 3-loop model (modelled by reducing Y translation by 70%) gives a 1h reduction in period (Figure 6.5a) that appears to agree with the observed period of a *gi* null mutant allele (*gi-11* [74], Figure 6.5b) Experimental work carried out by Peter Gould and Anthony Hall. According to our model, the Y-TOC1 feedback loop generates the 18h rhythm seen in a *cca1;lhy* mutant (Figure 6.5d). A reduction in *Y* function in the *cca1;lhy* mutant background should therefore reduce the robustness of this residual rhythm. In fact, simulation of the *cca1;lhy;gi* triple mutation results in complete arrythmia (Figure 6.5c). The single oscillation of *LHY* mRNA levels predicted by the simulation is due to the residual component of Y remaining in a simulated *gi* mutant. The very strong phenotype encouraged us to test the rhythms of *cca1;lhy;gi* triple

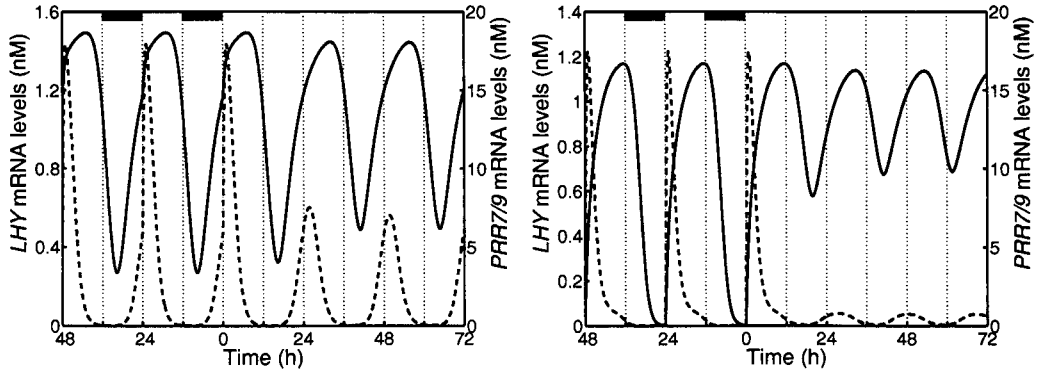


Figure 6.3: Comparison of simulated *LHY* (dotted line) and *PRR7/9* (solid line) mRNA levels using the optimal 3-loop model in; (a) a WT background; (b) a *toc1* mutant background (translation rate of *TOC1* reduced to $1/1000^{\text{th}}$ of WT value).

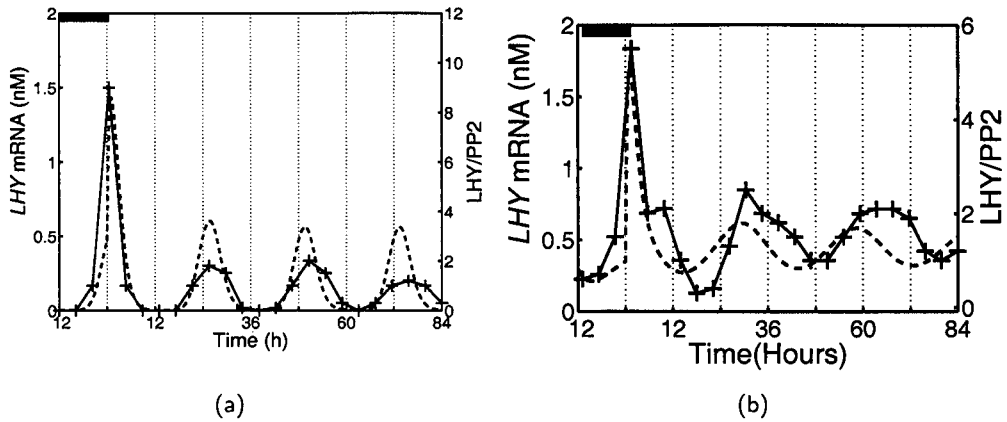


Figure 6.4: Comparison of simulated *LHY* mRNA levels in the WT and *prr7/9* mutant background (dotted line) using the optimal 3-loop model, to experimental data extracted from Ref. [84] (solid line).

mutant plants (Figure 6.5d). Encouragingly, with no further adjustment of parameters we find excellent agreement between the model and data; the *cca1;lhy;gi* mutant is also arrhythmic. *GI* operates in a feedback loop separate from *LHY* and *CCA1*, which is required for the maintanence of residual rhythms in the *cca1;lhy* background.

Our new model allows futher investigation into the structure of the *Arabidopsis* clock circuit. The model predicts that two short-period oscillators, the PRR7/9-LHY/CCA1 loop and the TOC1-Y/GI loop, are coupled together by the LHY/CCA1-TOC1-X loop (Figure 6.2). If the coupling were broken, the two oscillators might run with different periods within one cell. This is predicted by simulation of an *x* mutant (Figure 6.6), where *LHY* mRNA levels oscillated with a 20.4h period under constant light conditions, and TOC1 levels oscillated with a 17.3h period. This causes the oscillations of *LHY* and *TOC1* mRNA to go in and out of phase with each other.

6.4 Discussion

By extending our interlocked feedback loop model to include the PRR7/9-CCA1/LHY feedback loop, we are able to fit to several new pieces of experimental data. We are able to recreate the long period in *LHY* mRNA oscillations seen in a *prr7/9* mutant, and the PRR7/9-CCA1/LHY loop is capable of generating the short period oscillation observed in a *toc1* mutant background. A simulation of a *gi* mutant by reducing Y translation rates to 70% of the WT value results in a 1h short period oscillation that matches well to experiment. The arrythmicity seen experimentally in a *cca1;lhy;gi* plant in constant light conditions matches well with our simulations and provides evidence that our prediction that *gi* operates in a feedback loop with TOC1 is correct.

Our results concerning the possible decoupling of the feedback loops of the *Arabidopsis* clock leads to some very interesting possibilities. Future genetic screens could target mutations that cause desynchronisation of *LHY* and *TOC1* mRNA rhythms (Figure 6.6). Period differences among rhythms in the same plant have been observed repeatedly and in some cases can be interpreted as evidence for desynchronisation of

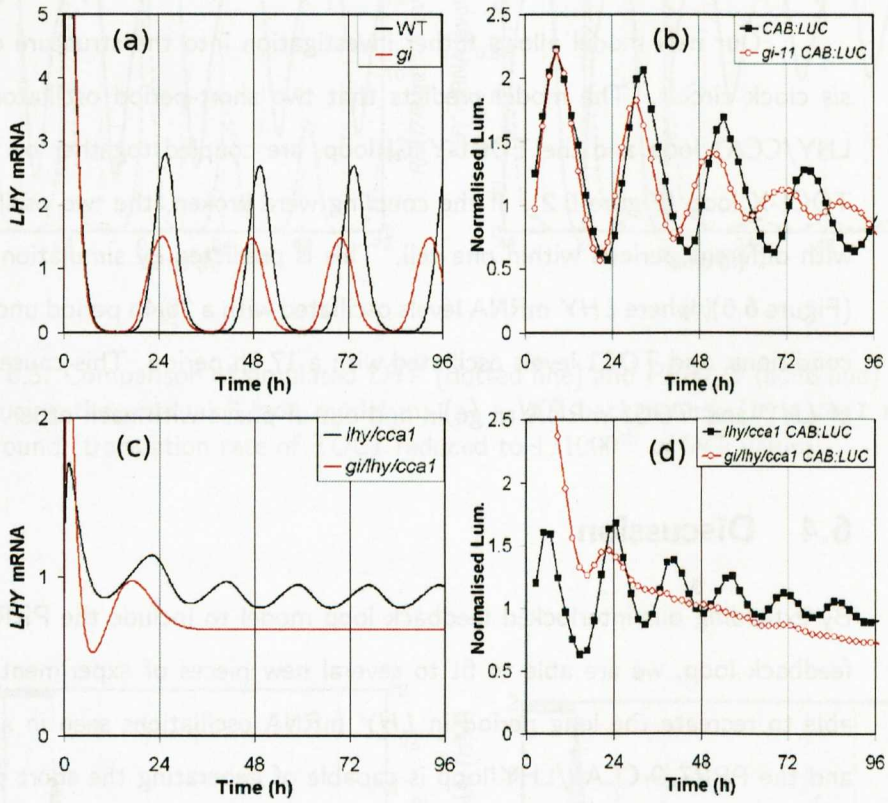


Figure 6.5: *GI* acts as *Y* in a feedback Loop with *TOC1*; (a) Simulation of WT (black line) and mutant (Y translation rate reduced by 70%, red line) under constant light (LL); (b) Corresponding experimental data assaying circadian control of CAB:LUCIFERASE expression by video imaging; (c) Simulation of *LHY* mRNA under LL in *cca1;lhy* (black line) and *cca1;lhy;gi* mutants (red line). Translation rate of *LHY* mRNA in simulated *cca1;lhy* and *cca1;lhy;gi* mutants was sset to 1/1000 of WT value. (d) Corresponding experimental data assaying CAB:LUCIFERASE expression. The *cca1;lhy;gi* mutant is arrhythmic. All data was normalised to the average level of expression. Experimental data from Peter Gould and Anthony Hall.

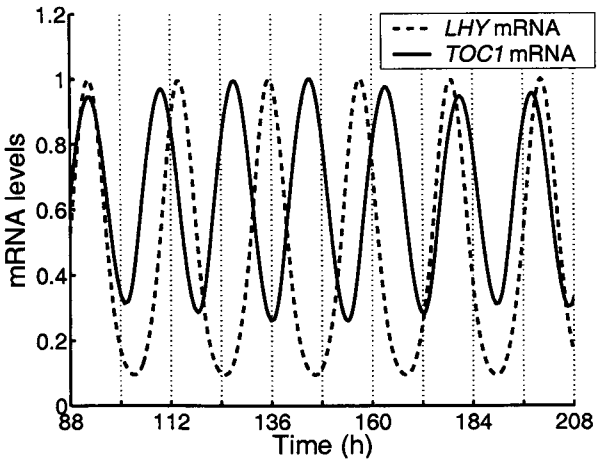


Figure 6.6: An *x* mutation can de-couple the two clocks Simulation of *LHY* mRNA levels and *TOC1* mRNA levels in an *x* mutant background, shown for an interval of free running rhythm in LL. Translation rate of *X* mRNA in simulated mutant is $1/1000^{\text{th}}$ WT value. Peak levels of *TOC1* and *LHY* mRNA can be seen to go in and out of phase with each other. Data was normalised to the maximum level of expression.

two intracellular oscillators, although cell-type-specific effects cannot be excluded (for example,[126, 93]). The 3-loop model provides a mechanism for such intracellular desynchronisation, if the various rhythmic processes are controlled by different loops and coupling between loops is weakened in some conditions. This flexibility of circadian regulation is expected to offer a selective advantage, particularly where the changing photoperiod varies the relative timing of dawn and dusk [127]. There is strong evidence in *Drosophila* [128] and mammals [129] for separate control of morning and evening processes by oscillators in different cells, which are coupled together by cell-cell signalling. Plant clocks are coupled only weakly between cells, if at all [91], but the 3-loop circuit suggests that an analogous architecture can be constructed within a single cell, by coupling the loop of morning-expressed genes LHY/CCA1 and PRR7/9 to the evening-expressed TOC1-GI loop.

Chapter 7

Conclusions

I have achieved, through close collaboration with Andrew Millar's lab, the construction of a detailed and predictive mathematical model of the *Arabidopsis* circadian clock. Using an iterative process of experimentation and mathematical modelling, we extended our circadian clock model from a one loop to a three loop network. Our final 3 loop model for the *Arabidopsis* clock incorporates several of the known clock genes. We have been able to design experiments using our models in order to test our prediction of *GI*'s role in a feedback loop with *TOC1*. Throughout this work we have used a parameter optimisation scheme in order to rigorously determine the consistency between our network model and experimental data. The implications of each addition to our *Arabidopsis* clock models are discussed at the end of each chapter, but I here summarise the key points of the work and place them in the wider context of modelling genetic networks.

Our first model of the *Arabidopsis* clock was able to pinpoint areas where the LHY/CCA1-TOC1 network was unable to explain experimental data (Chapter 2). The model failed to reproduce experimental data that *TOC1* mRNA levels are reduced before *LHY* mRNA levels rise in the night. Also, simulations using the model were incorrectly unaffected by changes in photoperiod. However, the model was able to reproduce several important experimental features, such as the phases of the peaks of *LHY* and *TOC1* mRNA LD cycles, and the long period seen in oscillations in DD. As such, it was a useful

starting point for our examination of the *Arabidopsis* clock, and suggested a specific phase of the circadian cycle where further experimentation continues to be required to identify additional components of the circuit.

Using a qualitative cost function as opposed to directly fitting to experimental data in my optimisation process allowed us to avoid the many technical issues with fitting to noisy and varied experimental data (Chapter 3). Our optimisation scheme also represents a step forward from the manual choice of parameters seen in many previous works, in which a very limited exploration of parameter space has carried the drawback that discrepancies cannot be assigned with confidence to choice of network over poor choice of parameters. The approach produced optimised solutions that fitted very well to experimental data.

There are limitations to the qualitative cost function technique. To some extent, the choice of weight of each error term must be arbitrary, and these qualitative terms should be replaced by a direct fit to data once sufficient, reliable time series data are available. The cost function terms developed here are specific for the *Arabidopsis* clock, and therefore would need to be reformulated if applied to a different network. That said, we envision that it would require only minor alterations to apply this optimisation technique to models of the *Drosophila*, *Neurospora* or mammalian clock.

In order to overcome the limitations of the LHY/CCA1-TOC1 model, we added network components to the model based on inferences from experimental data. A more comprehensive technique would be to optimise parameters for random network topologies, in order to distinguish the most promising networks. This is not presently feasible, due to the level of computational power available, so additions to the network had to be informed from data. By extending our network to include two hypothetical components, genes X and Y, we were able to produce a far better fit to experimental data. The interlocked loop model (Chapter 4) simulates a residual short-period oscillation in the *cca1;lhy* mutant, as characterised by experiments carried out Victoria Hibberd. No single-loop model is able to do this. Our model also matches experimental data un-

der constant light (LL) conditions and correctly senses photoperiod. The model also predicts a distinctive pattern of *Y* mRNA accumulation in the WT and in the *cca1;lhy* double mutant, with *Y* mRNA levels increasing transiently at dawn. We designed an experiment carried out by Megan Southern to identify *Y* based on this prediction. *GIGANTEA (GI)* mRNA levels fit very well to our predicted profile for *Y* identifying *GI* as a strong candidate for *Y*.

The ability to predict new components of a genetic network using mathematical modelling is potentially very advantageous. As most systems' components have not been completely identified by experiment, mathematical modelling affords us a tool which allows us more precision in the design of our experiments. The mathematical models can produce quantitative predictions of the mRNA expression profile of missing components, for example. However, modelling may not accurately identify the mechanism by which hidden components work, or how many components are interacting.

The inclusion of the PRR7/9-LHY/CCA1 loop into our interlocked loop model further increased the accuracy of our modelling (Chapter 6). Using this 3-loop model we were able to reproduce the effect of a *prr7/prr9* mutation, and the low amplitude oscillation seen in a *toc1-2* plant. The almost exact match betwen our simulation of a *cca1;lhy;gi* triple mutant plant and that observed in experiments represents a step towards confirming our prediction that *GI* acts as a feedback loop with *TOC1*.

Our final 3-loop model of the *Arabidopsis* circadian clock is reaching the complexity of the mathematical models of the *Neurospora* and *Drosophila* circadian clocks. However, there is less experimental evidence for the components of our *Arabidopsis* network. Further experimentation is required to further characterise GI, PRR7 and PRR9s' role in the circuit. Neither can our model reproduce all known molecular data for the Arabidopsis clock. There are several further genes yet to be placed in the 3-loop network, such as *ELF3*, *PRR3* and *ELF4*. In order to be able to include further genes it will be necessary to perform several additional experiments. The paucity of data for *ELF3* and *PRR5*, for example, makes it impossible to accurately understand their role.

Experiments need to be carried out in order to produce a standard set of results for each clock gene. Also, experiments need to be carried out in order to measure parameters which are currently unknown in our model - such as the transcription and translation rates of *LHY*. Our final 3-loop network model has > 70 unknown parameters and is reaching the limits of the capability of available optimisation schemes. As computer processing power continues to increase this will become less of a problem, but it will always be preferable to measure rather than infer, however accurately, parameter values. Mathematical models will be invaluable in suggesting which parameters to accurately measure first, in order to make best use of available experimental resources.

In order to reduce the number of parameters to be optimised and produce easier to conceptualise models, we developed *Arabidopsis* clock models using timedelays rather than modelling the protein levels of our clock components. (Chapter 5). Our reduced model of the LHY/CCA1-TOC1 network clearly shows the long delay required between *TOC1* expression and *LHY* transcription, although the scope of analysis capable with time delay models was limited due to the fixed nature of the delays. Reduced models can be used for a preliminary examination of the properties of a network, but can not replace biochemical reaction models for detailed analysis.

We can envision several extensions to this work. The importance of the light input pathways in our models was to be expected, because the plant circadian system is known to interact with multiple photoreceptor pathways in a complex fashion (reviewed in [130, 131]). The tracking of multiple phases during entrainment is thought to require at least two light inputs to two feedback loops [31], for example, which are present in our interlocking loop model. Our final model, including the PRR7/9-CCA1/LHY loop, has three light inputs to three feedback loops. A study should be made of how the network structures affect entrainment and whether the models are able to track multiple phases. The known input photoreceptors could in future be explicitly included, providing quantitative estimates of their function for comparison to data from plant photobiology.

The models will help to reveal how the circadian output pathways allow the

few genes of the clock to control over a thousand rhythmically-regulated genes in the *Arabidopsis* genome [51]. However, the complexity of such biological networks is likely to limit the quantitative accuracy of early models, so the potential value of simplified experimental model systems that facilitate the link to mathematical analysis is clear. These will include synthetic gene networks in microbial hosts but also reduced systems: For example by characterising circadian rhythms in seedlings without light exposure, the complexity of the circadian system and the number of clock-controlled target genes would be greatly reduced.

Our work modelling the circadian clock of *Arabidopsis* fits into the broader aim of plant biologists to develop a 'Computable plant'. As our biological knowledge and computational resources increase, it is becoming possible to use computational modelling to study in detail plant development (Reviewed in [132]). Our clock model can in future be linked into part of a larger network of models designed to create in silico a virtual plant. Using this software, it would, for example be possible to visualise the effect on leaf movement, and development in general, of a clock mutation.

Inferring network structure from system-wide transcriptomic data has already been shown to be a useful technique in systems biology [133]. In this thesis we have shown that a detailed mathematical model of even a partially-characterised genetic network can be accurately predictive, and that mathematical models can be a crucial tool in our attempts to understand complex regulatory networks.

Appendix A

Derivation of Michealis-Menten kinetics

Biochemical reactions generally involve enzymes as specific catalysts, which react on substrates to form products. For example haemoglobin in red blood cells is an enzyme and oxygen, with which it combines, is a substrate. Michaelis and Menten [134, 27] proposed a basic enzyme reaction as follows:

$$S + E \underset{k_1}{\overset{k_2}{\leftrightarrows}} SE \tag{A.1}$$

$$SE \xrightarrow{k_3} P + E \tag{A.2}$$

Where the k's are constant parameters relating to the reaction rates. The first of these reactions is a reversible creation of the complex SE from the substrate S using the enzyme E. The non-reversible break-down of this complex yields the product P and the original enzyme. Hence the enzyme is not used up in the overall reaction. The Law of Mass action states that the rate of a reaction is proportional to the product of the concentrations of the reactants. Denoting the concentration of the enzyme, product, complex and substrate by e,p,c,s respectively, the law can be applied to equations A.1

and A.2 to give the following system of non-linear equations.

$$\frac{ds}{dt} = -k_1 es + k_2 c \tag{A.3}$$

$$\frac{de}{dt} = -k_1 es + (k_2 + k_3)c \tag{A.4}$$

$$\frac{dc}{dt} = k_1 es - (k_2 + k_3)c \tag{A.5}$$

$$\frac{dp}{dt} = k_3 c \tag{A.6}$$

The initial conditions are assumed to be $s(0) = s_0$, $e(0) = e_0$, $c(0) = 0$, $p(0) = 0$ so that only the enzyme and the substrate are initially present. In the mechanism described in A.1 and A.2 the enzyme E is merely a catalyst so its total concentration is a constant.

$$e(t) + c(t) = e_0 \tag{A.7}$$

Hence the four equations in A.3-A.6 can be reduced to two, namely the differential equations for the concentration of the substrate and the complex. However, analytic solutions are not attainable, so an approximation on the relative amounts of reactants needs to be made. Very little of the enzyme is needed compared to the amount of substrate, so the approximation $e_0 << s_0$ can be made. This suggests that the time-scale of complex formation is much shorter than the timescale for product formation, and we can assume steady state conditions for the concentration of the complex. For a detailed analysis of Michaelis-Menten kinetics please see [27]. Here, we use the assumption of steady state kinetics to rewrite the complex formation equation as follows:

$$k_1 es - (k_2 + k_3)c = 0 \tag{A.8}$$

This can be rearranged to

$$c = \frac{es}{K_m} \tag{A.9}$$

where K_m, the Michaelis constant, is given by

$$K_m = \frac{k_2 + k_3}{k_1} \tag{A.10}$$

The Michaelis constant represents the dissociation constant (affinity for substrate) of the enzyme-substrate (ES) complex. Low values indicate that the ES complex is held together very tightly and rarely dissociates without the substrate first reacting to form product. Substituting Eqn.A.7 into Eqn.A.9 gives

$$c = \frac{(e_0 - c)s}{K_m} \tag{A.11}$$

Rearranging Eqn.A.11 gives

$$c = e_0 \frac{1}{1 + \frac{K_m}{s}} \tag{A.12}$$

Substituting Eqn,A.12 into Eqn.A.6 gives

$$\frac{dp}{dt} = k_3 e_0 \frac{s}{K_m + s} = V_{max} \frac{s}{K_m + s} \tag{A.13}$$

V_{max} is the maximum rate of the reaction.

Many enzymes have more than one binding site, and reactions involving these enzymes can be cooperative. The above procedure can be applied in a more general way to give the following equation for cooperative reactions, called a Hill function [27]:

$$\frac{dp}{dt} = V_{max} \frac{s^n}{{K_m}^n + s^n}. \tag{A.14}$$

where n is called the Hill coefficient and gives the degree of cooperativity in the system. We use Michaelis Menten kinetics terms of the form seen in Eqn. A.12 to describe the degradation of the clock components mRNA and protein through out this thesis, and Hill Functions of the form seen in Eqn. A.13 are used to describe transcription rates of the mRNA. This is because we have made similar assumptions, such as there being implicit enzymes that control transcription and degradation, and the time-scale of the complex formation of the enzyme and our substrate being much shorter than the time-scale for product formation. This is conventional in the literature.

Appendix B

Optimal parameter values for LHY/CCA1-TOC1-X network

Table B.1: Optimal parameter values for LHY/CCA1-TOC1-X network.

Parameter values for LHY/CCA1-TOC1-X Network			
Parameter Name	**Parameter Value**	**Parameter Description**	**Dimensions**
q1	12.2286	Coupling constant of light activation of LHY transcription	1/h
n1	9.4424	Maximum light-independent LHY transcription rate	nM/h
a	2	Hill coefficient of activation by X	
g1	2.0947	Constant of activation by protein X	nM
m1	8.0496	Maximum rate of LHY mRNA degradation	nM/h
k1	3.9155	Michaelis constant of LHY mRNA degradation	nM
p1	4.0188	Rate constant of LHY mRNA translation	1/h
r1	10.6578	Rate constant of LHY transport into nucleus	1/h
r2	1.0993	Rate constant of LHY transport out of nucleus	1/h
m2	2.1267	Maximum rate of cytoplasmic LHY degradation	nM/h
k2	0.2511	Michaelis constant of cytoplasmic LHY degradation	nM
m3	3.7925	Maximum rate of nuclear LHY degradation	nM/h
k3	8.4915	Michaelis constant of nuclear LHY degradation	nM
n2	3.4691	Maximum light-independent TOC1 transcription rate	nM/h
n3	1.2238	Maximum of light dependent activation of TOC1 transcription	1/h
b	2	Hill coefficient of repression by protein LHY	
g2	1.3859	Constant of repression by protein LHY	nM
m4	7.1075	Maximum rate of TOC1 mRNA degradation	nM/h
k4	2.2424	Michaelis constant of TOC1 mRNA degradation	nM
p2	2.1535	Rate constant of TOC1 mRNA translation	1/h
r3	0.6876	Rate constant of TOC1 movement into nucleus	1/h
r4	4.1674	Rate constant of TOC1 movement out of nucleus	1/h
m5	1.5743	Maximum rate of light dependent cytoplasmic TOC1 degradation	nM/h
m6	2.5529	Maximum rate of light independent cytoplasmic TOC1 degradation	nM/h
k5	1.8972	Michaelis constant of cytoplasmic TOC1 degradation	nM
m7	0.5879	Maximum rate of light dependent nuclear TOC1 degradation	nM/h
m8	0.9016	Maximum rate of light independent nuclear TOC1 degradation	nM/h
k6	2.6877	Michaelis constant of nuclear TOC1 degradation	nM
n4	2.6891	Maximum transcription rate of X mRNA	nM/h
c	2	Hill coefficient of activation by TOC1	
g3	1.9160	Constant of activation by TOC1	nM
m9	5.4578	Maximum rate of degradation of protein X mRNA	nM/h
k7	1.9433	Michaelis constant of protein X mRNA degradation	nM

Table B.1: (continued)

Parameter Name	Parameter Value	Parameter Description	Dimensions
p3	2.4201	Rate constant of X mRNA translation	1/h
r5	2.0076	Rate constant of protein X movement into nucleus	1/h
r6	20.0848	Rate constant of protein X movement out of nucleus	1/h
m10	2.1119	Maximum rate of degradation of cytoplasmic protein X	nM/h
k8	5.2738	Michaelis constant of cytoplasmic protein X degradation	nM
m11	2.1795	Maximum rate of degradation of nuclear protein X	nM/h
k9	18.1832	Michaelis constant of nuclear protein X degradation	nM
p4	0.5	Light dependent production of protein P	nM/h
q2	1.0000	Coupling constant of light activation of protein P degradation	1/h
m12	1.2000	Miaximum rate of protein P degradation	nM/h
k10	1.2000	Michaelis constant of protein P degradation	nM

Appendix C

Optimisation of interlocked loop model

We follow the optimisation technique as described in chapter 3. We constructed our cost function Δ as a sum of terms that each quantify the agreement between our model and a qualitative experimental feature. Small values of the cost function correspond to a model (or set of parameter values) that give a good qualitative agreement with the corresponding experimental features. The weighting of each term in the cost function was chosen so that an acceptable error within the range of expermental variability would add on the order of 1 unit to the cost function. In order to evaluate the terms in the cost function we solved the equations numerically over 600 hours, 300 hours in 12 hour light 12 hour dark cycles (LD), followed by 300 hours in darkness (DD) (the first 200 hours of the LD cycles of each solution are discarded as transitory). In order to find a set of optimal solutions for each network studied, the cost function was calculated for a cross section of parameter space chosen using a Sobol quasi-random number generator [108]. The best fifty solutions were then put through a further optimisation step using a simulated annealing routine [113].

The interlocked feedback loop network proposed here was scored both as a model of WT and of the *cca1;lhy* mutant. The double mutant was simulated by reducing the

translation rate of *LHY* to 1/1000th of its WT value. This simulated mutation led to arrhythmia in all the single loop models. Additional terms were introduced to the cost function to score models of the double mutant, specifying entrainment under LD12:12 with peak *TOC1* expression 5h after dawn and oscillations with a period of 18h or less in DD. To enable *TOC1* activation sufficiently early in the day in the double mutant, we required that *Y* transcription peaked sharply at dawn in the double mutant.

The 20 parameter sets with the lowest costs (which allowed the model to best fit the specified criteria) all simulated similar gene expression profiles in WT and *cca1;lhy* backgrounds. An optimal parameter set was chosen from these 20 by comparing the simulated rhythms to experimental data that were not included in the cost function (see chapter 4).

The equations were solved using MATLAB (Mathworks, Cambridge, UK). Parameter optimisation was carried out [116] by compiling MATLAB code into C and running the code on a task farm computer consisting of 62 x 2.6GHz Xeon CPUs. A user-friendly interface has been developed, "Circadian Modelling", to allow simulations using this and other circadian models, without MATLAB. This software and files for the interlocked loop model are available online [135].

Here we discuss the new cost function terms in detail. We added new terms to the cost function in order to optimise the interlocked feedback loop model to both WT and *cca1;lhy* mutant data. The equations were re-solved with the translation rate of *LHY* reduced to a thousandth of its WT value in order to simulate the double mutant. The cost function now becomes

$$\Delta = \delta_{\tau_{ld}} + \delta_{\tau_d} + \delta_{\phi} + \delta_{c_L} + \delta_{\text{size}} + \delta_{\phi d} + \delta^{dm}_{\tau_{ld}} + \delta^{dm}_{\tau_d} + \delta^{dm}_{\phi} + \delta^{dm}_{c_Y} + \delta^{dm}_{\text{size}} \tag{C.1}$$

where the label (dm) denotes the cost function for the *cca1;lhy* double mutant. One new WT cost function term $\delta_{\phi d}$ added represents a minor change to constrain an appropriate phase difference between the peak times of LHY (ϕ_L) and $TOC1$ mRNA (ϕ_T), $\Delta\Phi_d = \phi_T - \phi_L$ (modulo half the period), with a characteristic prefactor of 10h. This term

makes no discernable difference to the cost function when applied to the optimised one loop models. See term below:

$$\delta_{\phi d} = (10/\Delta\Phi_d)^2 \tag{C.2}$$

δ_{size} was also altered slightly in order to ensure both LHY mRNA and $TOC1$ mRNA oscillations do not decay too quickly when entering DD. This is necessary as in the interlocked feedback loop model $TOC1$ mRNA levels can oscillate through $TOC1$'s feedback loop with Y whilst LHY mRNA levels are arrhythmic. δ_{size} becomes

$$\delta_{\text{size}} = \sum_{i=L,T} \left[\left(\frac{1}{\langle \Delta c_i^{(m)} \rangle_{ld}} \right)^2 + \left(\frac{\tau_o}{\tau_e} \right)^2 \right]. \tag{C.3}$$

The first term remains the same, and the second term is now summed over LHY (L) and $TOC1$ (T). All the other WT cost function terms remain the same as for the one loop model optimisation.

Using the same methodology as for the WT terms, we define below the new double mutant terms of the cost function. The first new term,

$$\delta_{\tau_{ld}}^{dm} = \sum_{i=Y,T} \langle (24 - \tau_i^{(m)})^2 / 0.15 \rangle_{ld} \tag{C.4}$$

is the summed error in the period, τ, for Y (Y) and $TOC1$ (T) mRNA levels in LD cycles.

We penalise solutions with a period of $TOC1$ greater than 18 hours in the dark. $\delta_{\tau_{ld}}^{(m)} = 0$ if the period is less than 18 hours, otherwise:

$$\delta_{\tau_d}^{dm} = \langle (18 - \tau_T^{(m)})^2 / 0.1 \rangle_d \tag{C.5}$$

The next term δ_{ϕ}^{dm} is defined as:

$$\delta_{\phi}^{dm} = \sum_{i=Y,T} \left[\langle \Delta\Phi_i^2 \rangle_{ld} + (\sigma[\Delta\Phi_i])^2 \right] \tag{C.6}$$

Here the first term compares the mean difference in phase over the LD cycles, where $\Delta\Phi_i = \bar{\phi}_i - \phi_i$, ϕ_i is the phase (from dawn) of the RNA peak in the model and $\bar{\phi}_Y = 1h$,

$\bar{\phi}_T = 5h$ are the target phases of the peaks in $c_Y^{(m)}$ and $c_T^{(m)}$ respectively. The second term describes a cost of $O(1)$ for solutions whose variations in peak phases are 1h. Next,

$$\delta_{\text{size}}^{dm} = \sum_{i=Y,T} \left(\frac{1}{\langle \Delta c_i^{(m)} \rangle_{ld}} \right)^2 \tag{C.7}$$

This term costs for solutions in LD cycle with oscillation sizes, ($\Delta c_i^{(m)} = {c_i^{(m)}}_{max} - {c_i^{(m)}}_{min}$), less than 1nm. Finally,

$$\delta_{c_Y}^{dm} = \sum_{i=2,-2} \langle \left(\frac{2/3 c_Y^{(m)}(t_{\text{p}})}{c_Y^{(m)}(t_{\text{p}}) - c_Y^{(m)}(t_{\text{p}} + i)} \right)^2 \rangle_{ld} \tag{C.8}$$

This term checks that the Y mRNA expression profile has a sharp peak in LD cycles, with an O(1) contribution if Y's expression level has dropped by 2/3 of its oscillation size within 2 hours before and after its peak of expression. As for the single loop optimisations, throughout the implementation the cost function was "capped" at $\Delta_{\text{max}} = 10^4$, such that $\Delta \rightarrow Min(10^4, \Delta)$. The sum of the double mutant cost function terms was also capped at 10^3.

Appendix D

Optimal parameter values for interlocked feedback network

Table D.1: Optimal parameter values for interlocked feedback loop network.

Parameter values for interlocked feedback loop network			
Parameter Name	Parameter Value	Parameter Description	Dimensions
q1	2.4514	Coupling constant of light activation of LHY transcription	1/h
n1	5.1694	Maximum light-independent LHY transcription rate	nM/h
a	3.3064	Hill coefficient of activation by protein X	
g1	0.8767	Constant of activation by protein X	nM
m1	1.5283	Maximum rate of LHY mRNA degradation	nM/h
k1	1.8170	Michaelis constant of LHY mRNA degradation	nM
p1	0.8295	Rate constant of LHY mRNA translation	1/h
r1	16.8363	Rate constant of LHY transport into nucleus	1/h
r2	0.1687	Rate constant of LHY transport out of nucleus	1/h
m2	20.4400	Maximum rate of cytoplasmic LHY degradation	nM/h
k2	1.5644	Michaelis constant of cytoplasmic LHY degradation	nM
m3	3.6888	Maximum rate of nuclear LHY degradation	nM/h
k3	1.2765	Michaelis constant of nuclear LHY degradation	nM
n2	3.0087	MaximumTOC1 transcription rate	nM/h
b	1.0258	Hill coefficient of activation by protein Y	
g2	0.0368	Constant of activation by protein Y	nM
g3	0.2658	Constant of repression by LHY	nM
c	1.0258	Hill coefficient of repression by LHY	
m4	3.8231	Maximum rate of TOC mRNA degradation	nM/h
k4	2.5734	Michaelis constant of TOC mRNA degradation	nM
p2	4.3240	Rate constant of TOC1 mRNA translation	1/h
r3	0.3166	Rate constant of TOC1 movement into nucleus	1/h
r4	2.1509	Rate constant of TOC1 movement out of nucleus	1/h
m5	0.0013	Maximum rate of light dependent cytoplasmic TOC1 degradation	nM/h
m6	3.1741	Maximum rate of light independent cytoplasmic TOC1 degradation	nM/h
k5	2.7454	Michaelis constant of cytoplasmic TOC1 degradation	nM
m7	0.0492	Maximum rate of light dependent nuclear TOC1 degradation	nM/h
m8	4.0424	Maximum rate of light independent nuclear TOC1 degradation	nM/h
k6	0.4033	Michaelis constant of nuclear TOC1 degradation	nM
n3	0.2431	Maximum transcription rate of protein X	nM/h
d	1.4422	Hill coefficient of activation by TOC1	

Table D.1: (continued)

Parameter Name	Parameter Value	Parameter Description	Dimensions
g4	0.5388	Constant of activation by TOC1	nM
m9	10.1132	Maximum rate of degradation of protein X mRNA	nM/h
k7	6.5585	Michaelis constant of protein X mRNA degradation	nM
p3	2.1470	Rate constant of X mRNA translation	1/h
r5	1.0352	Rate constant of protein X movement into nucleus	1/h
r6	3.3017	Rate constant of protein X movement out of nucleus	1/h
m10	0.2179	Maximum rate of degradation of cytoplasmic protein X	nM/h
k8	0.6632	Michaelis constant of cytoplasmic protein X degradation	nM
m11	3.3442	Maximum rate of degradation of nuclear protein X	nM/h
k9	17.1111	Michaelis constant of nuclear protein X degradation	nM
q2	2.4017	Coupling constant of light activation of Y mRNA transcription	1/h
n4	0.0657	Light dependent component of Y transcription	nM/h
n5	0.1649	Light independent component of Y transcription	nM/h
g5	1.1780	Constant of repression by TOC1	nM
g6	0.0645	Constant of repression by LHY	nM
e	3.6084	Hill coefficient of repression by TOC1	
f	1.0237	Hill coefficient of repression by LHY	
m12	4.2970	Maximum rate of degradation of protein Y mRNA	nM/h
k10	1.7303	Michaelis constant of protein Y mRNA degradation	nM
p4	0.2485	Rate constant of Y mRNA translation	1/h
r7	2.2123	Rate constant of protein Y movement into nucleus	1/h
r8	0.2002	Rate constant of protein Y movement out of nucleus	1/h
m13	0.1347	Maximum rate of degradation of cytoplasmic protein Y	nM/h
k11	1.8258	Michaelis constant of cytoplasmic protein Y degradation	nM
m14	0.6114	Maximum rate of degradation of nuclear protein Y	nM/h
k12	1.8066	Michaelis constant of nuclear protein Y degradation	nM
p5	0.5000	Light dependent production of protein P	nM/h
k13	1.2000	Michaelis constant of protein P degradation	nM
m15	1.2000	Miaximum rate of protein P degradation	nM/h
q3	1.0000	Coupling constant of light activation of protein P degradation	1/h

Appendix E

Optimisation of 3-loop Model

The parameter values (Appendix D) and equations for the optimal solution of the interlocked feedback loop model (Chapter 4) were taken as our starting parameter set and network.

The parameters for *LHY* transcription in Eqn. 6.1 were re-optimised to take into account that in a *prr7/prr9* plant the period of the clock is approx 30h in LL [126]. In order to model the *prr7/prr9* mutant the equations were solved for 10^6 simulated annealing points in order to minimise the qualitative cost function as defined in Appendix C which quantifies the goodness of fit of the solutions to several key pieces of experimental data. The cost function error term for the WT period in DD, δ_{τ_d} was replaced with an error term for the period in LL, $\delta_{\tau_{ll}}$ in order to find a solution in LL with a period of 30h, as opposed to a DD solution with a period of 25h.

$\delta_{\tau_{ll}}$ is given by:

$$\delta_{\tau_{ll}} = \sum_{i=L,T} \langle (30 - \tau_i^{(m)})^2 / f \rangle_{ll}. \tag{E.1}$$

This represents the summed error in the period, τ, for LHY (L) and $TOC1$ (T) mRNA levels in constant light conditions, where $\langle\rangle_{ll}$ gives the average over the cycles between $300 < t < 600$. The biological evidence strongly indicates that the free running period of the clock in an *prr7/9* mutant plant is not less than 30h [84], but we have less

confidence in assigning a precise value hence we adopt values of $f = 0.05$ if $\tau_i^{(m)} \leq 30$ and $f = 2$ if $\tau_i^{(m)} > 30$.

The equations of the interlocked loop model were then altered to include the effects of PRR7/9 (Chapter 6.1). The equation for *LHY* transcription was modified to include the effect of PRR7/9 repressing *LHY* transcription (Eqn. 6.1, and three equations were added to describe PRR7/9 mRNA, cytoplasmic and nuclear protein levels (Eqs. 6.14 to 6.16).

The parameters for *PRR7/9* were then optimised (Eqs. 6.14 to 6.16) in order to model a WT plant. In order to reduce parameter space, the acute light activation term for *PRR7/9* q_4 was set to the same value as the acute light response for the *LHY* promoter q_1, the Hill coefficient of *PRR7/9* activation by LHY, g, was set to the same value as c. The Hill coefficient of *TOC1* repression by LHY, and g_0, the constant of repression of *LHY* by PRR7/9 was set to 1.

As in Chapter 3 the equations were solved for 1 million quasirandom points in parameter space, and $\delta_{\tau_{ll}}$ was altered in order to search for a period in LL of 24h rather than 30h $\delta_{\tau_{ll}} = \sum_{i=L,T} \langle (24 - \tau_i^{(m)})^2 / f \rangle_{ll}$. The costfunction was also altered to find a short period oscillation in the *toc1* background [62], as opposed to a short period oscillation in a *lhy/cca1* background (Appendix C). This gives a cost function:

$$\Delta = \delta_{\tau_{ld}} + \delta_{\tau_{ll}} + \delta_{\phi} + \delta_{c_L} + \delta_{\text{size}} + \delta_{\phi d} + \delta_{\tau_{ld}}^{toc1} + \delta_{\tau_d}^{toc1} + \delta_{\phi}^{toc1} + \delta_{c_Y}^{toc1} + \delta_{\text{size}}^{toc1} \tag{E.2}$$

where the label (toc1) denotes the cost function for the *toc1* mutant plant. We define below the terms for the new *toc1* mutant terms of the cost function:

$$\delta_{\tau_{ld}}^{toc1} = \sum_{i=A,L} \langle (24 - \tau_i^{(m)})^2 / 0.15 \rangle_{ld} \tag{E.3}$$

is the summed error in the period, τ, for A (PRR7/9) and T (*LHY*) mRNA levels in LD cycles. We penalise solutions with a period of *PRR7/9* greater than 20 hours under

constant light conditions. $\delta_{\tau_{ld}}^{(m)} = 0$ if the period is less than 20 hours, otherwise:

$$\delta_{\tau_d}^{toc1} = \langle (20 - \tau_A^{(m)})^2 / 0.1 \rangle_d \tag{E.4}$$

The next term δ_{ϕ}^{dm} is defined as:

$$\delta_{\phi}^{toc1} = \left[\langle \Delta\Phi_L^2 \rangle_{ld} + (\sigma[\Delta\Phi_L])^2 \right] \tag{E.5}$$

Here the first term compares the mean difference in phase over the LD cycles, where $\Delta\Phi_i = \bar{\phi}_L - \phi_L$, ϕ_L is the phase (from dawn) of the *LHY* mRNA peak in the model and $\bar{\phi}_L = 1h$ is the target phase of the peak in $c_L^{(m)}$. The second term describes a cost of $O(1)$ for solutions whose variations in peak phase are 1h. Next,

$$\delta_{\text{size}}^{toc1} = \sum_{i=A,L} \left(\frac{1}{\langle \Delta c_i^{(m)} \rangle_{ld}} \right)^2 \tag{E.6}$$

This term costs for solutions in LD cycle with oscillation sizes, ($\Delta c_i^{(m)} = {c_i^{(m)}}_{max} - {c_i^{(m)}}_{min}$), less than 1nm. Finally,

$$\delta_{c_Y}^{toc1} = \sum_{i=2,-2} \langle \left(\frac{2/3 c_L^{(m)}(t_p)}{c_L^{(m)}(t_p) - c_L^{(m)}(t_p + i)} \right)^2 \rangle_{ld} \tag{E.7}$$

The first term checks that the *LHY* mRNA expression profile has a sharp peak in LD cycles, with an O(1) contribution if *LHY*'s expression level has dropped by 2/3 of its oscillation size within 2 hours before and after its peak of expression. As for the single loop optimisations, throughout the implementation the cost function was "capped" at $\Delta_{\max} = 10^4$, such that $\Delta \rightarrow Min(10^4, \Delta)$. The sum of the *toc1* cost function terms was also capped at 10^3.

As we are only optimising parameters in the equations for LHY and PRR7/9, the output of the model is the same as for the interlocked loop model when simulating a *lhy/cca1* plant, as PRR7/9 and LHY are no longer part of the functional clock in this case.

A further 10^5 simulated annealing points was carried out on the 10 best solutions found from the search of parameter space, to find the optimal parameter set.

Appendix F

Optimal parameter values for the 3-loop Model

Table F.1: Optimal parameter values for the 3-loop model.

Parameter values for 3 – loop model of Arabidopsis clock			
Parameter Name	Parameter Value	Parameter Description	Dimensions
q1	4.1954	Coupling constant of light activation of LHY transcription	1/h
n0	0.0500	Maximum light-dependent LHY transcription rate	nM/h
g0	1	Constant of repression by APPR7/9	nM/h
α	4.000	Hill coefficient of repression by APPR7/9	nM
n1	7.8142	Maximum light-independent LHY transcription rate	nM/h
a	1.2479	Hill coefficient of activation by protein X	
g1	3.1383	Constant of activation by protein X	nM
m1	1.9990	Maximum rate of LHY mRNA degradation	nM/h
k1	2.3920	Michaelis constant of LHY mRNA degradation	nM
p1	0.8295	Rate constant of LHY mRNA translation	1/h
r1	16.8363	Rate constant of LHY transport into nucleus	1/h
r2	0.1687	Rate constant of LHY transport out of nucleus	1/h
m2	20.4400	Maxjmum rate of cytoplasmic LHY degradation	nM/h
k2	1.5644	Michaelis constant of cytoplasmic LHY degradation	nM
m3	3.6888	Maximum rate of nuclear LHY degradation	nM/h
k3	1.2765	Michaelis constant of nuclear LHY degradation	nM
n2	3.0087	MaximumTOC1 transcription rate	nM/h
b	1.0258	Hill coefficient of activation by protein Y	
g2	0.0368	Constant of activation by protein Y	nM
g3	0.2658	Constant of repression by LHY	nM
c	1.0258	Hill coefficient of repression by LHY	
m4	3.8231	Maximum rate of TOC mRNA degradation	nM/h
k4	2.5734	Michaelis constant of TOC mRNA degradation	nM
p2	4.3240	Rate constant of TOC1 mRNA translation	1/h
r3	0.3166	Rate constant of TOC1 movement into nucleus	1/h
r4	2.1509	Rate constant of TOC1 movement out of nucleus	1/h
m5	0.0013	Maximum rate of light dependent cytoplasmic TOC1 degradation	nM/h
m6	3.1741	Maximum rate of light independent cytoplasmic TOC1 degradation	nM/h
k5	2.7454	Michaelis constant of cytoplasmic TOC1 degradation	nM
m7	0.0492	Maximum rate of light dependent nuclear TOC1 degradation	nM/h
m8	4.0424	Maximum rate of light independent nuclear TOC1 degradation	nM/h

Table F.1: (continued)

Parameter Name	Parameter Value	Parameter Description	Dimensions
k6	0.4033	Michaelis constant of nuclear TOC1 degradation	nM
n3	0.2431	Maximum transcription rate of protein X	nM/h
d	1.4422	Hill coefficient of activation by TOC1	
g4	0.5388	Constant of activation by TOC1	nM
m9	10.1132	Maximum rate of degradation of protein X mRNA	nM/h
k7	6.5585	Michaelis constant of protein X mRNA degradation	nM
p3	2.1470	Rate constant of X mRNA translation	1/h
r5	1.0352	Rate constant of protein X movement into nucleus	1/h
r6	3.3017	Rate constant of protein X movement out of nucleus	1/h
m10	0.2179	Maximum rate of degradation of cytoplasmic protein X	nM/h
k8	0.6632	Michaelis constant of cytoplasmic protein X degradation	nM
m11	3.3442	Maximum rate of degradation of nuclear protein X	nM/h
k9	17.1111	Michaelis constant of nuclear protein X degradation	nM
q2	2.4017	Coupling constant of light activation of Y mRNA transcription	1/h
n4	0.0857	Light dependent component of Y transcription	nM/h
n5	0.1649	Light independent component of Y transcription	nM/h
g5	1.1780	Constant of repression by TOC1	nM
g6	0.0645	Constant of repression by LHY	nM
e	3.6064	Hill coefficient of repression by TOC1	
f	1.0237	Hill coefficient of repression by LHY	
m12	4.2970	Maximum rate of degradation of protein Y mRNA	nM/h
k10	1.7303	Michaelis constant of protein Y mRNA degradation	nM
p4	0.2485	Rate constant of Y mRNA translation	1/h
r7	2.2123	Rate constant of protein Y movement into nucleus	1/h
r8	0.2002	Rate constant of protein Y movement out of nucleus	1/h
m13	0.1347	Maximum rate of degradation of cytoplasmic protein Y	nM/h
k11	1.8258	Michaelis constant of cytoplasmic protein Y degradation	nM
m14	0.6114	Maximum rate of degradation of nuclear protein Y	nM/h
k12	1.8066	Michaelis constant of nuclear protein Y degradation	nM
p5	0.5000	Light dependent production of protein P	nM/h
k13	1.2000	Michaelis constant of protein P degradation	nM
m15	1.2000	Miaximum rate of protein P degradation	nM/h
q3	1.0000	Coupling constant of light activation of protein P degradation	1/h

Table F.1: (continued)

Parameter Name	Parameter Value	Parameter Description	Dimensions
q4	2.4514	Coupling constant of light activation of LHY transcription	1/h
g	1.0258	Hill coefficient of activation by LHY	
n6	8.0706	Maximum light-independent APRR7/9 transcription rate	nM/h
n7	0.0002	Maximum light-dependent APPR7/9 transcription rate	$(nM/h)(nM)^g$
g7	0.0004	Constant of activation by LHY	nM
m16	12.2398	Maximum rate of degradation of APRR7/9 mRNA	nM/h
k14	10.3617	Michaelis constant of APRR7/9 mRNA degradation	nM
p6	0.2907	Rate constant of APRR7/9 mRNA translation	1/h
r9	0.2528	Rate constant of APRR7/9 protein movement out of nucleus	1/h
r10	0.2212	Rate constant of APRR7/9 protein movement into the nucleus	1/h
m17	4.4505	Maximum rate of degradation of cytoplasmic protein APRR7/9	nM/h
k15	0.0703	Michaelis constant of cytoplasmic protein APRR7/9 degradation	nM
m18	0.0156	Maximum rate of degradation of nuclear protein APRR7/9	nM/h
k16	0.6104	Michaelis constant of nuclear protein APRR7/9 degradation	nM

Bibliography

[1] J. C. Dunlap, J. L. Loros, and P. J. DeCoursey. *Chronobiology: Biological Time-Keeping*. Sinauer, Sunderland, 2003.

[2] A. N. Dodd, N. Salathia, A. Hall, E. Kevei, R. Toth, F. Nagy, J.M. Hibberd, A. J. Millar, and A. A. Webb. Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science*, 309:630–633, 2005.

[3] C. S. Pittendrigh. Circadian rhythms and the circadian organization of living systems. *Cold Spring Harb Symp Quant Biol.*, 25:159–184, 1960.

[4] J. C. Dunlap, J. J. Loros, Y. Liu, and S. K. Crosthwaite. Eukaryotic circadian systems: cycles in common. *Genes Cells*, 4:1–10, 1999.

[5] K. Lee, J. J. Loros, and J. C. Dunlap. Interconnected feedback loops in the *Neurospora* circadian system. *Science*, 289:107–110, 2000.

[6] N. R. Glossop, L. C. Lyons, and P. E. Hardin. Interconnected feedback loops within the Drosophila circadian oscilator. *Science*, 286:766–8, 1999.

[7] G. E. Duffield. DNA microarray analyses of circadian timing: the genomic basis of biological time. *J Neuroendocrinol.*, 15:991–1002, 2003.

[8] P. Smolen, D. A. Baxter, and J. H. Byrne. Modeling Transcriptional Control in Gene Networks–Methods, Recent Results, and Future Directions. *Bull Math Biol.*, 62:247–292, 2000.

[9] D. G. M. Beersma. Why and How Do We Model Circadian Rhythms? *Jour. Biol. Rhythms*, 20:304–313, 2005.

[10] C.S. Pittendrigh. Circadian systems. I. The driving oscillation and its assay in Drosophila pseudoobscura. *Proc. Natl. Acad. Sci.*, 58:1762–1767, 1967.

[11] E.L. Peterson and D.S. Saunders. The circadian eclosion rhythm in Sarcophaga argyrostoma: A limit cycle representation of the pacemaker. *J. Theor. Biol.*, 86:265–277, 1980.

[12] E.L. Peterson. A Limit Cycle Interpretation of a Mosquito Circadian Oscillator. *J. Theor. Biol.*, 84:281–310, 1980.

[13] C.A. Czeisler, J.F. Duffy, T.L. Shanahan, E.N. Brown, J.F. Mitchell, D.W. Rimmer, J.M. Ronda, E.J. Silva, J.S. Allan, J.S. Emens, D. Dijk, and R.E. Kronauer. Stability, Precision, and Near-24-Hour Period of the Human Circadian Pacemaker. *Science*, 284:2177–2181, 1999.

[14] D.B. Forger, M.E. Jewett, and R.E. Kronauer. A simpler model of the human circadian pacemaker. *Journal of Biol. Rhythms*, 14:538–543, 1999.

[15] C.H. Johnson. Forty years of PRCs–what have we learned? *Chronobiol. Int.*, 16:711–743, 1999.

[16] A. T. Winfree. Integrated view of resetting circadian clock. *J. Theor. Biol.*, 28:327–374, 1970.

[17] B. C. Goodwin. Oscillatory behavior in enzymatic control processes. *Advances in Enzyme Regulation*, 3:425–428, 1965.

[18] J. C. Leloup, D. Gonze, and A. Goldbeter. Limit cycle models for circadian rhythms based on transcriptional regulation in Neurospora and Drosophila. *J. Biol. Rhythms*, 14:433–448, 1999.

[19] P. Ruoff and L. Rensing. The temperature-compensated Goodwin model simulates many circadian clock properties. *J. Theor. Biol*, 179:275–285., 1996.

[20] P. Ruoff, C. Monnerjahn M. Vindjevik, and L. Rensing. The Goodwin model: Simulating the effect of light pulses on the circadian sporulation rhythm of Neurospora Crassa. *J. Theor. Biol.*, 209:29–42, 2001.

[21] P. Ruoff, C. Monnerjahn M. Vindjevik, and L. Rensing. The Goodwin oscillator: On the importance of degradation reactions in the circadian clock. *J. Biol. Rhythms*, 14:469–479, 1999.

[22] J. C. Dunlap and J. J. Loros. The Neurospora circadian system. *J Biol Rhythms.*, 19:414–424, 2004.

[23] A. C. Froehlich, Y. Liu, J. J. Loros, and J. C. Dunlap. White Collar-1, a circadian blue light photoreceptor, binding to the frequency promoter. *Science*, 297:815–819, 2002.

[24] C. Talora, L. Franchi, H. Linden, P. Ballario, and G. Macino. Role of a white collar-1-white collar-2 complex in blue-light signal transduction. *EMBO J.*, 18:4961–4968, 1999.

[25] B. D. Aronson, K. A. Johnson, J. J. Loros, and J.C. Dunlap. Negative feedback defining a circadian clock: autoregulation of the clock gene *frequency*. *Science*, 263:1578–1584, 1994.

[26] Liu Y., N. Y. Garceau, J. J. Loros, and J. C. Dunlap. Alternative initiation of translation and time-specific phosphorylation yield multiple forms of the essential clock protein FREQUENCY. *Cell*, 89:469–476, 1997.

[27] J. D. Murray. *Mathematical Biology*. Springer, New York, 1993.

[28] G. Kurosawa, A. Mochizuki, and Y. Iwasa. Comparative Study of Circadian Clock

Models, in Search of Processes Promoting Oscillation. *J. Theor. Biol.*, 216:193–208., 2002. doi:10.1006/jtbi.2002.2546.

[29] P. Smolen, D. Baxter, and J. H. Byrne. Modeling Circadian Oscillations with Interlocking positive and Negative Feedback Loops. *J. Neurosci*, 21:6644–6656, 2001.

[30] P. Francois. A model for the Neurospora circadian clock. *Biophys J.*, 88:2369–2383, 2005.

[31] D A Rand, B V Shulgin, D Salazar, and AJ Millar. Design principles underlying circadian clocks. *Interface*, 1:119–130, 2004.

[32] J. Stelling, E. D. Gilles, and F. J. Doyle III. Robustness properties of circadian clock arhitectures. *Proc. Natl. Acad. Sci.*, 101:13210–13215, 2004.

[33] C. Lee, K. Bae, and I. Edery. PER and TIM inhibit the DNA binding activity of a Drosophila CLOCK-CYC/dBMAL1 heterodimer without disrupting formation of the heterodimer: a basis for circadian transcription. *Mol Cell Biol.*, 19:5316–5325, 1999.

[34] Bae. K., P. E. Lee, C. Hardin, and I. Edery. dCLOCK is present in limiting amounts and likely mediates daily interactions between the dCLOCK-CYC transcription factor and the PER-TIM complex. *J Neurosci.*, 20:1746–1753, 2000.

[35] P. E. Hardin. The Circadian Timekeeping System of Drosophila. *Current Biology*, 15:R714–R722, 2005.

[36] M. P. Myers, K. Wager-Smith, A. Rothenfluh-Hilfiker, and M. W. Young. Light-induced degradation of TIMELESS and entrainment of the Drosophila circadian clock. *Science*, 271:1736–1740, 1996.

[37] I. Edery, L. J. Zwiebel, M. E. Dembinska, and M. Rosbash. Temporal phosphorylation of the Drosophila period protein. *Proc. Natl. Acad. Sci.*, 91:2260–2264, 1994.

[38] J. J. Tyson, C. I. Hong, C. D. Thron, and B. Novak. A simple model of circadian rhythms based on dimerization and proteolysis of PER and TIM. *J. Biophys.*, 77:2411–2417, 1999.

[39] H. R. Ueda, M. Hagiwara, and H. Kitano. Robust Oscillations within the Interlocked Feedback Model of Drosophila Circadian Rhythm. *J. Theor. Biol*, 210:401–406., 2001. doi:10.1006/jtbi.2000.2226.

[40] P. Ruoff, M. K. Christensen, and V. K. Sharma. PER/TIM-mediated amplification, gene dosage effects and temperature compensation in an interlocking-feedback loop model of the Drosophila circadian clock. *J Theor Biol.*, 237:41–57, 2005.

[41] P. Smolen, P.E. Hardin, B. S. Lo, D. A. Baxter, and J. H. Byrne. Simulation of Drosophila circadian oscillations, mutations, and light responses by a model with VRI, PDP-1, and CLK. *Biophys J.*, 86:2786–2802, 2004.

[42] M. A. Lema, D. A. Golombek, and J. Echave. Delay model of the circadian pacemaker. *J. Theor Biol.*, 204:565–573, 2000.

[43] K. Sriram and M. S. Gopinathan. A two variable delay model for the circadian rhythm of Neurospora crassa. *J. Theor Biol.*, 231:23–38, 2004.

[44] P. Smolen, D. A. Baxter, and J. H. Byrne. A reduced model clarifies the role of feedback loops and time delays in the Drosophila circadian oscillator. *Biophys J*, 83(5):2349–59, 2002.

[45] N. Barkai and S. Leibler. Circadian clocks limited by noise. *Nature*, 403:267–268, 2000.

[46] D. Gonze, J. Halloy, and A. Goldbeter. Robustness of circadian rhythms with respect to molecular noise. *Proc. Natl. Acad. Sci.*, 99:673–678, 2002.

[47] J. C. Leloup and A. Goldbeter. Toward a detailed computational model for the mammalian circadian clock. *Proc. Natl. Acad. Sci*, 100:7051–7056, 2003.

[48] D. B. Forger and C. S. Peskin. A detailed predictive model of the mammalian clock. *Proc. Natl. Acad. Sci.*, 100:14806–14811, 2003.

[49] J. de Marain. Observation botanique. *Histoire de l'Academie Royale des Sciences*, pages 35–36, 1729.

[50] B.M. Sweeney. *Rhythmic Phenomena in Plants*. Academic Press, San Diego, 1987.

[51] S. L. Harmer, J. B. Hogenesch, M. Straume, H. S. Chang, B. Han, T. Zhu, X. Wang, J. A. Kreps, and S. A. Kay. Orchestrated transcription of key pathways in Arabidopsis by the Circadian Clock. *Science*, 290:2110–2113, 2000.

[52] K. D. Edwards, P. E. Anderson, A. Hall, N. S. Salathia, J. C. W. Locke, J. R. Lynn, M. Straume, J. Q. Smith, and A. J. Millar. FLOWERING LOCUS C Mediates Natural Variation in the High-Temperature Response of the Arabidopsis Circadian Clock. *Plant Cell*, page In Press, 2006.

[53] Meyerowitz E. M. Prehistory and history of Arabidopsis research. *Plant Physiol.*, 125:15–19, 2001.

[54] P. F. Devlin and S.A. Kay. Cryptochromes–bringing the blues to circadian rhythms. *Trends Cell Biol.*, 9:295–298, 1999.

[55] M. Chen, J. Chory, and C. Fankhauser. Light signal transduction in higher plants. *Annu Rev Genet.*, 38:87–117, 2004.

[56] J.F. Martinez-Garcia, E. Huq, and P. H. Quail. Direct targeting of light signals to a promoter element-bound transcription factor. *Science*, 288:859–63, 2000.

[57] A. Viczian, S. Kircher, E. Fejes, A. J. Millar, E. Schafer, L. Kozma-Bognar, and Nagy F. Functional characterization of phytochrome interacting factor 3 for the Arabidopsis thaliana circadian clockwork. *Plant Cell Physiol.*, 46:1591–1602, 2005.

[58] A. Oda, S. Fujiwara, H. Kamada, G. Coupland, and T. Mizoguchi. Antisense suppression of the Arabidopsis PIF3 gene does not affect circadian rhythms but causes early flowering and increases FT expression. *FEBS Lett.*, 557:259–264, 2004.

[59] S. Makino, A. Matsushika, M. Kojima, T. Yamashino, and T. Mizuno. The APRR1/TOC1 quintet implicated in circadian rhythms of Arabidopsis thaliana: 1. characterization with APRR1-overexpressing plants. *Plant Cell Physiol*, 43:58–69, 2002.

[60] A. J. Millar, Carre I. A., C. A. Strayer, N. H. Chua, and S. A. Kay. Circadian clock mutants in arabidopsis identified by luciferase imaging. *Science*, 27:1161–1163, 1995.

[61] P. Mas. Circadian clock signaling in Arabidopsis thaliana: from gene expression to physiology and development. *Int J Dev Biol.*, 49:491–500, 2005.

[62] P Mas, D Alabadi, M J Yanovsky T Oyama, , and S A Kay. Dual role of TOC1 in the control of circadian and photomorphogenic responses in Arabidopsis. *The Plant Cell*, 15:223–236, 2003.

[63] C. Strayer, T. Oyama, T. F. Schultz, R. Raman, D. E. Somers, P. Mas, S. Panda, J. A. Kreps, and S. A. Kay. Cloning of the Arabidopsis clock gene TOC1, an autoregulatory response regulator homolog. *Science*, 289(5480):768–771, 2000.

[64] T. Mizoguchi, K. Wheatley, Y. Hanzawa, L. Wright, M. Mizoguchi, H. R. Song, I. A. Carre, and G. Coupland. LHY and CCA1 are partially redundant genes required to maintain circadian rhythms in Arabidopsis. *Dev Cell*, 2:629–641, 2002.

[65] P. Mas, W. Y. Kim, D. E. Somers, and S. A. Kay. Targeted degradation of TOC1 by ZTL modulates circadian function in Arabidopsis thaliana. *Nature*, 426:567–570, 2003.

[66] R. Schaffer, N. Ramsay, A. Samach, S. Corden, J. Putterill, I. A. Carre, and G. Coupland. The Late Elongated Hypocotyl mutation of Arabidopsis disrupts circadian rhythms and the photoperiodic control of flowering. *Cell*, 93:1219–1229, 1998.

[67] D. Alabadi, M. J. Yanovsky, P. Mas, S. L Harmer, and S. A Kay. Critical role for CCA1 and LHY in maintaining circadian rhythmicity in Arabidopsis. *Curr. Biol.*, 12:757–761, 2002.

[68] H. R. Song and I. A. Carre. DET1 regulates the proteasomal degradation of LHY, a component of the Arabidopsis circadian clock. *Plant Mol Biol.*, 57:761–771, 2005.

[69] Z. Y. Wang and E. M. Tobin. Constitutive expression of the circadian clock associated 1 (CCA1) gene disrupts circadian rhythms and suppresses its own expression. *Cell*, 93:1207–1217, 1998.

[70] S. Sugano, C. Andronis, M. S. Ong, R. M. Green, and E. M. Tobin. The protein kinase CK2 is involved in regulation of circadian rhythms in Arabidopsis. *Proc. Natl. Acad. Sci.*, 96:12362–12366, 1999.

[71] X. Daniel, S. Sugano, and E. M. Tobin. CK2 phosphorylation of CCA1 is necessary for its circadian oscillator function in Arabidopsis. *Proc. Natl. Acad. Sci*, 101:3292–3297, 2004.

[72] D. Alabadi, T. Oyama, M. J. Yanovsky, F. G. Harmon, P. Mas, and S. A. Kay. Reciprocal regulation between TOC1 and LHY/CCA1 within the Arabidopsis circadian clock. *Science*, 293:880–883, 2001.

[73] D. H. Park, D. E. Somers, Y. S. Kim, Y. H. Choy, H. K. Lim, M. S. Soh, H. J. Kim, S. A. Kay, and H. G. Nam. Control of circadian rhythms and photoperiodic flowering by the Arabidopsis GIGANTEA gene. *Science*, 285(5433):1579–1582, 1999.

[74] S. Fowler, K. Lee, H. Onouchi, A. Samach, K. Richardson, B. Morris, G. Coupland, and J. Putterill. GIGANTEA: a circadian clock-controlled gene that regulates photoperiodic flowering in Arabidopsis and encodes a protein with several possible membrane-spanning domains. *The EMBO Journal*, 18:4679–4688, 1999.

[75] E. Huq, J. M. Tepperman, and P. H. Quail. GIGANTEA is a nuclear protein involved in phytochrome signaling in Arabidopsis. *Proc. Natl. Acad. Sci.*, 97:9789–9794, 2000.

[76] T. Mizoguchi, L. Wright, S. Fujiwara, F. Cremer, K. Lee, H. Onouchi, A. Mouradov, S. Fowler, H. Kamada, J. Putterill, and G. Coupland. Distinct Roles of GIGANTEA in Promoting Flowering and Regulating Circadian Rhythms in Arabidopsis. *The Plant Cell*, 17:2255–2270, 2005.

[77] A Hall. Personal communication.

[78] M. R. Doyle, Seth J. Davis, R. M. Bastow, H. G. McWatters, L. Kozma-Bognar, F. Nagy, A. J. Millar, and R. Amasino. The ELF4 gene controls circadian rhythms and flowering time in Arabidopsis thaliana. *Nature*, 132:732–238, 2002.

[79] E. A. Kikis, R. Khanna, and P. H. Quail. ELF4 is a phytochrome-regulated component of a negative-feedback loop involving the central oscillator components CCA1 and LHY. *Plant J.*, 44:300–313, 2005.

[80] A. Matsushika, S. Makino, M. Kojima, and T. Mizuno. Circadian waves of expression of the APRR1/TOC1 family of pseudo-response regulators in Arabidopsis thaliana:an insight into the plant circadian clock. *Plant Cell Physiol.*, 41:1002–1012, 2000.

[81] T. Mizuno and N. Nakamichi. Pseudo-Response Regulators (PRRs) or True Oscillator Components (TOCs). *Plant Cell Physiol.*, 46:677–685, 2005.

[82] A. Matsushika, A. Imamura, T. Yamashino, and T. Mizuno. Aberrant expression of the light-inducible and circadian-regulated APRR9 gene belonging to the circadian-associated APRR1/TOC1 quintet results in the phenotype of early flowering in Arabidopsis thaliana. *Plant Cell Physiol.*, 43:833–843, 2002.

[83] P. A. Salome and C. R. McClung. PSEUDO-RESPONSE REGULATOR 7 and 9 are partially redundant genes essential for the temperature responsiveness of the Arabidopsis circadian clock. *Plant Cell.*, 17:791–803, 2005.

[84] E. M. Farre, S. L. Harmer, F. G. Harmon, M. J. Yanovsky, and S. A. Kay. Overlapping and distinct roles of PRR7 and PRR9 in the Arabidopsis circadian clock. *Curr Biol*, 15(1):47–54, 2005.

[85] N. Nakamichi, M. Kita, S. Ito, T. Yamashino, and T. Mizuno. PSEUDO-RESPONSE REGULATORS, PRR9, PRR7 and PRR5, together play essential roles close to the circadian clock of Arabidopsis thaliana. *Plant Cell Physiol.*, 46:686–696, 2005.

[86] K. A. Hicks, A. J. Millar, I. A. Carre, D. E. Somers, M. Straume, D. R. Meeks-Wagner, and S. A. Kay. Conditional circadian dysfunction of the Arabidopsis early-flowering 3 mutant. *Science*, 274:790–792, 1996.

[87] K. A. Hicks, T. M. Albertson, and D. R. Wagner. EARLY FLOWERING3 encodes a novel protein that regulates circadian clock function and flowering in Arabidopsis. *Plant Cell*, 13:1281–1292, 2001.

[88] M. F. Covington, S. Panda, X. L. Liu, C. A. Strayer, D. R. Wagner, and S. A. Kay. ELF3 modulates resetting of the circadian clock in Arabidopsis. *Plant Cell*, 13:1305–1315, 2001.

[89] H. G. McWatters, R. M. Bastow, A. Hall, and A. J. Millar. The ELF3 zeitnehmer regulates light signalling to the circadian clock. *Nature*, 6813:716–720, 2000.

[90] M. H. Hastings and E. D. Herzog. Clock genes, oscillators, and cellular networks in the suprachiasmatic nuclei. *J Biol Rhythms.*, 19:400–413, 2004.

[91] S. C. Thain, A. Hall, and A. J. Millar. Functional independence of circadian clocks that regulate plant gene expression. *Current Biology*, 10:951–956, 2000.

[92] A. Hall, L. Kozma-Bognar, R. M. Bastow, F. Nagy, and A. J. Millar. Distinct regulation of CAB and PHYB gene expression by similar circadian clocks. *Plant Journal*, 32:529–537, 2002.

[93] T. P. Michael, P. A. Salome, and C.R. McClung. Two Arabidopsis circadian oscillators can be distinguished by differential temperature sensitivity. *Proc. Natl. Acad. Sci.*, 100:6878–6883, 2003.

[94] H. Kitano. Systems Biology: A Brief Overview. *Biophys J.*, 295:1662–1664, 2002.

[95] G. Kurosawa and Y. Iwasa. Saturation of Enzyme Kinetics in Circadian Clock Models. *J. Biol. Rhythms*, 17:568–577, 2002.

[96] J. Y. Kim, H. R. Song, B. L. Taylor, and I. A. Carre. Light-regulated translation mediates gated induction of the Arabidopsis clock protein LHY. *EMBO J*, 22:935–944, 2003.

[97] K. A. Kaczorowski and P. H. Quail. Arabidopsis PSEUDO-RESPONSE REGULATOR7 is a signaling intermediate in phytochrome-regulated seedling deetiolation and phasing of the circadian clock. *Plant Cell*, 15:2654–2665, 2003.

[98] M. R. Doyle, Seth J. Davis, R. M. Bastow, H. G. McWatters, L. Kozma-Bognar, F. Nagy, A. J. Millar, and R. Amasino. The ELF4 gene controls circadian rhythms and flowering time in Arabidopsis thaliana. *Nature*, 132:732–238, 2002.

[99] D. Bauer, A. Viczian, S. Kircher, T. Nobis, R. Nitschke, T. Kunkel, K. C. Panigrahi, E. Adams, E. Fejes, E. Schafer, and F. Nagy. Constitutive photomorphogenesis 1 and multiple photoreceptors control degradation of phytochrome interacting factor 3, a transcription factor required for light signaling in Arabidopsis. *Plant Cell*, 16:1433–45, 2004.

[100] L. C. Roden, H. R. Song, S. Jackson, K. Morris, and I. A. Carre. Floral responses to photoperiod are correlated with the timing of rhythmic expression relative to dawn and dusk in Arabidopsis. *Proc. Natl. Acad. Sci.*, 99(20):13313–8, 2002.

[101] A. J. Millar and S. A. Kay. Integration of circadian and phototransduction pathways in the network controlling CAB gene transcription in Arabidopsis. *Proc. Natl. Acad. Sci.*, 93(26):15491–15496, 1996.

[102] I. A. Carre and J. Y. Kim. MYB transcription factors in the Arabidopsis circadian clock. *J. Exp. Bot.*, 53:1551–1557, 2002.

[103] M. E. Eriksson and A. J. Millar. The circadian clock. A plant's best friend in a spinning world. *Plant Physiol.*, 132:732–738, 2003.

[104] P. Mendes and D. B. Kell. Non-linear optimization of biochemical pathways:applications to metabolic engineering and parameter estimation. *Bioinformatics*, 14:869–883, 1998.

[105] J. W. Zwolak, J. J. Tyson, and L. T. Watson. Estimating Rate Constants in Cell Cycle Models. *Proc. High Performance Computing Symposium 2001, San Diego, CA*, ?:53–57, 2001.

[106] L. F. Shampine and M. W. Reichelt. The MATLAB ODE suite. *SIAM Journal on Scientific Computing*, 18:1–22, 1997.

[107] A. J. Millar, M. Straume, J. Chory, N. H. Chua, and S. A. Kay. The regulation of circadian period by phototransduction pathways in Arabidopsis. *Science*, 267:1163–1166, 1995.

[108] W. H. Press, S. A. Teukolsky, W. T. Vettering, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge, Cambridge, 1996.

[109] I. M. Sobol. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Comp. Math. and Math. Phys.*, 7:784–802, 1967.

[110] P. Bratley and B. L. Fox. ALGORITHM 659 Implementing Sobol's Quasirandom Sequence Generator. *ACM Transactions on Mathematical Software*, 14:88–100, 1988.

[111] S. Joe and F. Y. Kuo. Remark on ALGORITHM 659: implementing Sobol's quasirandom sequence generator. *ACM Trans. on Mathematical Software (TOMS) archive*, 29:49–57, 2003.

[112] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7:308–313, 1965.

[113] N. Metropolis, A. W. Rosenbluth, Rosenbluth M. N., Teller A. H., and Teller E. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[114] S. P. Brooks and B. I. T. Morgan. Optimization Using Simulated Annealing. *The Statistician*, 44:241–257, 1995.

[115] K. S. Brown and J. P. Sethna. Statistical mechanical approaches to models with many poorly known parameters. *Physical Review E*, 68, 2003.

[116] J. C. W. Locke, A. J. Millar, and M. S. Turner. Modelling genetic networks with noisy and varied data: The circadian clock in *arabidopsis thaliana. J. Theor. Biology*, 234:383–393, 2005.

[117] J C W Locke, M M Southern, L Kozma-Bognar, V Hibberd, P E Brown, M S Turner, and A J Millar. Extension of a genetic network model by iterative experimentation and mathematical analysis. *Mol Systems Biol*, 1:13, 2005.

[118] S. Makino, A. Matsushika, M. Kojima, Y. Oda, and T. Mizuno. Light response of the circadian waves of the APRR1/TOC1 quintet: when does the quintet start singing rhythmically in Arabidopsis? *Plant Cell Physiol*, 42(3):334–9, 2001.

[119] D. E. Somers, W. Y. Kim, and R. Geng. The F-box protein ZEITLUPE confers dosage-dependent control on the circadian clock, photomorphogenesis, and flowering time. *Plant Cell*, 16(3):769–82, 2004.

[120] M. J. Yanovsky and S. A. Kay. Molecular basis of seasonal time measurement in Arabidopsis. *Nature*, 419(6904):308–12, 2002.

[121] T. Roenneberg and M. Merrow. Life before the clock: modeling circadian evolution. *J Biol Rhythms*, 17(6):495–505, 2002.

[122] G. Fleissner and G. Fleissner. Feedback loops in the circadian system. *Disc. Neurosci.*, 8:79–84, 1992.

[123] M. E. Jewett, D. B. Forger, and R. E. Kronauer. Revised limit cycle oscillator model of human circadian pacemaker. *J Biol Rhythms*, 14(6):493–9., 1999.

[124] S. Ito, A. Matsushika, H. Yamada, S. Sato, T. Kato, S. Tabata, T. Yamashino, and T. Mizuno. Characterization of the APRR9 pseudo-response regulator belonging to the APRR1/TOC1 quintet in Arabidopsis thaliana. *Plant Cell Physiol.*, 44:1237–1245, 2003.

[125] Y. Yamamoto, E. Sato, T. Shimizu, N. Nakamich, S. Sato, T. Kato, S. Tabata, A. Nagatani, T. Yamashino, and T. Mizuno. Comparative genetic studies on the APRR5 and APRR7 genes belonging to the APRR1/TOC1 quintet implicated in circadian rhythm, control of flowering time, and early photomorphogenesis. *Plant Cell Physiol.*, 44:1119–1130, 2003.

[126] A. Hall, R. M. Bastow, S. J. Davis, S. Hanano, H. G. Mcwatters, V. Hibberd, M. R. Doyle, S. Sung, K. J. Halliday R. M. Amasino, and A. J. Millar. The TIME

FOR COFFEE gene maintains the amplitude and timing of Arabidopsis circadian clocks. *Plant Cell*, 15:2719–2729, 2003.

[127] C. S. Pittendrigh and S. Daan. A functional analysis of circadian pacemakers in nocturnal rodents. *Journal of Comparative Physiology A*, 106:333–355, 1976.

[128] D. Stoleru, Y Peng, J. Agosto, and M. Rosbash. Coupled oscillators control morning and evening locomotor behaviour of Drosophila. *Nature*, 431:862–868, 2004.

[129] A. Jagota, H. O. de la Iglesia, and W. J. Schwartz. Morning and evening circadian oscillations in the suprachiasmatic nucleus in vitro. *Nat. Neurosci.*, 3:372–376, 2000.

[130] C. Fankhauser and D. Staiger. Photoreceptors in Arabidopsis thaliana: light perception, signal transduction and entrainment of the endogenous clock. *Planta*, 216(1):1–16, 2002.

[131] A. J. Millar. A suite of photoreceptors entrains the plant circadian clock. *J Biol Rhythms*, 18(3):217–26, 2003.

[132] P Prusinkiewicz and A.G. Rolland-Lagan. Modeling plant morphogenesis. *Curr. Opin. Plant. Biol.*, 9:83–88, 2006.

[133] D. di Bernardo, M. J. Thompson, Gardner T. S., Chobot S. E., E. L. Eastwood, A. P. Wojtovich, S. J. Elliott, S. E. Schaus, and J. J. Collins. Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks. *Nat. Biotech.*, 23:377–383, 2005.

[134] L. Michaelis and M.I. Menten. Die Kinetik der Invertinwirkung. *Biochem. Z*, 49:333–369, 1913.

[135] P.E. Brown. Circadian Modelling. *http:www.amillar.org/Downloads.html*, 2004.