Transcriptional Regulation of the *Arabidopsis* Circadian Clock Component *LHY*

Mark Spensley

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Biological Sciences

September 2007

Contents

List of Tables

List of Figures

Acknowledgments

I wish to thank Dr Isabelle Carré for her supervision and advice over the past four years.

In addition I am grateful to the members of the Molecular Cell Biology research group at the Univerity of Warwick for their technical advice and willingness to help wherever possible and Dr Laszlo Kozmar-Bognar for allowing us access to his *TOC1*ox transgenic line.

Declarations

I hereby declare that all the work presented in this thesis is my own, unless stated otherwise in the text, figure legends, or in the acknowledgements. and has not been submitted for a degree at any other institution

Mark Spensley

Abstract

Circadian clocks provide an endogenous mechanism by which an organism can anticipate and adapt to predictable daily changes in environmental conditions. In the model plant *Arabidopsis thaliana*, the circadian clock is believed to be composed of a number of coupled transcriptional negative feedback loops. The *LATE ELONGATED HYPOCOTYL* (*LHY*) gene is thought to form part of at least two of these transcriptional feedback loops, as well as playing a role in the perception of light signals by the clock.

To better understand how multiple transcriptional feedback loops might be integrated in the transcriptional regulation of *LHY*, we have performed an analysis of the *cis*-regulation of this gene. Through deletion analysis of reporter gene constructs, we have identified a 957 basepair region of the *LHY* promoter which contains sufficient sequence to direct the characteristic expression profile of *LHY*. Furthermore, we provide evidence that at least two circadian signals converge on this region. Electrophoretic mobility shift assays identified four classes of candidate *cis*-elements within the *LHY* promoter including a poly-CTT tract, an AAAAA motif, a candidate MYB-binding site and a G-box motif. Through mutational analysis of these elements, we have been able to to determine aspects of their *in vivo* regulatory function. We report that a G-box motif and the previously uncharacterized AAAAA element are implicated the regulation of *LHY* transcription by light signals. In etiolated seedlings, the region of the *LHY* promoter containing the MYB-binding site motif and multiple copies of the poly-CTT motif mediates regulation of *LHY* by both light-responsive and circadian signals.

Abbreviations

ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
bHLH	Basic helix-loop-helix
bZip	Basic leucine zipper
cDNA	Copy-DNA
Col	Columbia
dGTP	2′-deoxyguanosine 5′-triphosphate
DNA	Deoxyribonucleic acid.
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
EST	Expressed sequence tag
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HTGS	High throughput genomic sequence
LB	Luria-Bertani medium
MS	Murashige and Skoog medium
MS3	Murashige and Skoog medium with 3% by weight sucrose
OD_{600}	Optical density determined at 600 nm wavelength
PCR	Polymerase chain reaction
rpm	Revolutions per minute
RNA	Ribonucleic acid

RNAi	RNA interference
SDS	Sodium dodecyl sulphate
UTR	Untranslated region
WCC	White collar complex
Ws	Wassilewskija

Chapter 1

Introduction

1.1 Introduction

Circadian clocks are mechanisms that serve to regulate a wide variety of biochemical and physiological processes in such a way that they coincide with the appropriate time of day. To qualify as a circadian clock, such a mechanism must display certain functional properties (Pittendrigh, 1979).

Firstly, circadian clocks generate predictable output rhythms with a period of around 24 hours. These rhythms must persist even in the absence of any external timing cues, differentiating circadian regulation from diurnal regulation.

Finally, circadian clocks must be capable of entrainment. Entrainment describes the ability of a circadian oscillator to adapt its rhythmic parameters to correspond to external daily cycles, such as the daily cycle of dawn and dusk or daily cycles in ambient temperature. The two key parameters of circadian rhythms that are affected by entrainment are period and phase. The setting of the period of circadian oscillations, while possible in the laboratory, is not observed in the wild, since the period of daily cycles remains constant at 24 hours. On the other hand, the phase of circadian rhythms, which describes the

relationship between given reference points in the circadian rhythm and the entraining cycle, may vary to match seasonal changes in the entraining cycles in the wild. For example, at non-equatorial latitudes, the relative timings of dawn and dusk change with the progression of the seasons. In order that circadian regulated processes still occur at the correct time of day, relative to the timings of dawn and dusk, the phases of these processes must be adjusted to reflect the changes in the entraining signals. This property, described as photoperiodic entrainment, has been described for a variety of different rhythms in a diverse range of organisms (Pittendrigh & Daan, 1997)

Circadian clocks have been identified in a number of different organisms representing all kingdoms of life other than the eubacteria (Bell-Pedersen, D. and Cassone, V.M. and Earnest, D.J. and Golden, S.S. and Hardin, P.E. and Thomas, T.L. and Zoran, M.L., 2005). In those organisms in which the mechanisms of the circadian clock have been investigated, a number of common mechanistic properties are now beginning to emerge.

1.2 An Overview of Circadian Oscillator Mechanisms

1.2.1 The *Drosophila melanogaster* Clock

Since the early 1970s, when the first circadian mutant was identified (Konopka & Benzer, 1971), research into the mechanism of the *Drosophila* circadian oscillator has predominantly progressed ahead of research in other organisms.

Many of the genetic components of the *Drosophila* clock were discovered in screens to identify mutants defective in two relatively easily assayed circadian traits, namely locomotor activity rhythms and the time of emergence of adult flies from the pupal stage. Such mutants include the first to be identi-

fied, *period* (Konopka & Benzer, 1971) as well as *timeless* (Sehgal *et al.*, 1994). *dclock* (Allada *et al.*, 1998) and *cycle* (Rutila *et al.*, 1998). This focus on behavioural rhythms has led much of the research into the *Drosophila* circadian system to focus on an oscillator located within the *Drosophila* brain. Interestingly, however many other tissues express clock genes rhythmically, even in isolation from the central nervous system (Plautz *et al.*, 1997).

The *Drosophila* Molecular Oscillator

Current understanding of the *Drosophila* oscillator mechanism is built around a system of coupled transcriptional feedback loops centred around the transcription factor genes *dClock* and *Cycle*.

The first feedback loop in the *Drosophila* oscillator to be described is composed of the proteins dCLOCK (CLK), CYCLE (CYC), PERIOD (PER) and TIMELESS (TIM) (Darlington *et al.*, 1998; Rutila *et al.*, 1998). In this feedback loop, the genes *per* and *tim* are expressed in the early evening with the peak in protein accumulation occurring late in the night, at which point a complex of PER and TIM proteins translocates to the nucleus (Curtin *et al.*, 1995). Nuclear PER protein serves to repress the activity of a complex of dCLK and CYC proteins, which otherwise functions to activate expression of *per* and *tim* through binding to an E-box motif (CACGTG) on their promoters (Darlington *et al.*, 1998). Thus, the activation of *per* and *tim* expression is periodically repressed by the negative action of the PER-TIM complex on the positive acting dCLK-CYC complex, forming a negative feedback loop.

The negative feedback loop described above is coupled to further feedback loops incorporating the basic leucine zipper (bZip) transcription factors VRILLE (Glossop *et al.*, 2003) and PDP1 (Cyran *et al.*, 2003). In this feedback loop, the dCLK-CYC complex once again functions as a transcriptional activator, activating transcription of *vrille* and *Pdp1* by binding to E-box mo-

tifs in their promoters. VRILLE, which is expressed earlier than PDP1 , acts a repressor of *dClk* transcription closing the negative feedback loop. PDP1 acts later than VRILLE to activate *dClk* transcription, forming a positive feedback loop.

While the transcriptional feedback loops of the oscillator mechanism are believed to be largely responsible for the generation of circadian rhythms, various other components are important in the Drosophila regulation of oscillator mechanism. For example, the kinases SHAGGY, DOUBLETIME and CK2 have a critical role in regulating the period of the oscillator (Reviewed by Gallego & Virshup (2007)).

Light input to the *Drosophila* oscillator is known to perceived by the visual pathway (Menaker, 2003), but also through the action of the blue light receptor CRYPTOCHROME (Emery *et al.*, 1998). On activation by light, CRY binds to PER and TIM where it is believed to stimulate the degradation of TIM protein (Rosato *et al.*, 2001; Ceriani *et al.*, 1999; Lin *et al.*, 2001)

The system of transcriptional reedback loops that comprise the *Drosophila melanogaster* oscillator are represented in Figure 1.1.

Anatomical Organisation of the *Drosophila* Clock

Analysis of the spatial expression patterns of various clock components has allowed the circadian pacemaker in the fly brain to be mapped to a set of around 100 neurons clustered in 3 groups on each side of the brain: the dorsal neurons, the large and small ventral lateral neurons and the dorsal lateral neurons (Kaneko & Hall, 2000). While these different sub-groups of cells seem to share the same molecular oscillator mechanism, they do display certain functional differences. Specifically, those cells expressing the neuropeptide pigment dispersing factor (PDF), the ventral lateral neurons or morning cells (M-cells), regulate the morning peak of the circadian rhythm of behavioural

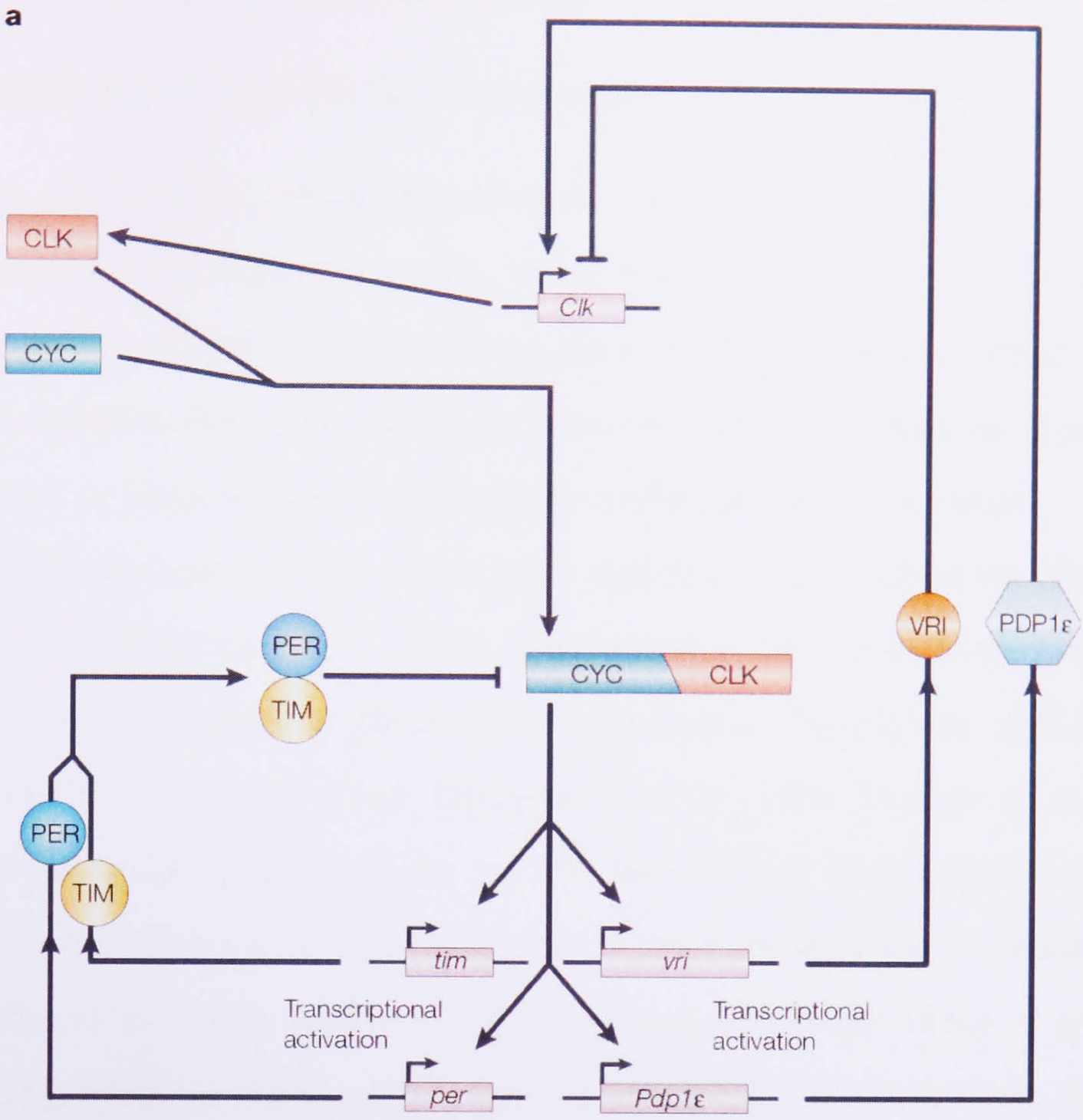


Figure 1.1: Structure of the *Drosophila melanogaster* oscillator mechanism. Figure taken from Bell-Pedersen, D. and Cassone, V.M. and Earnest, D.J. and Golden, S.S. and Hardin, P.E. and Thomas, T.L. and Zoran, M.L. (2005)

activity, whereas the evening activity peak requires the dorsal and dorsal lateral neurons, or evening cells (E-cells) (Grima *et al.*, 2004; Stoleru *et al.*, 2004). These two sets of neurons have been shown to interact as coupled oscillators in which the E-cells are reset daily by an as yet unidentified signal from the M-cells (Stoleru *et al.*, 2005). This coupling allows the central oscillator to shift the relative timings of morning and evening activity peaks to the appropriate time of day under entraining cycles of varying photoperiods.

1.2.2 The Mammalian Clock

The Molecular Oscillator in Mammals

While the mammalian circadian system is somewhat more complex than that of *Drosophila melanogaster*, strong similarities exist between the two. For example, many oscillator components appear to have been conserved through evolution between flies and mammals. Indeed several mammalian clock genes were cloned as homologues of *Drosophila melanogaster* clock genes.

While the mammalian *Clock* gene was discovered before its *Drosophila* counterpart (Vitaterna *et al.*, 1994), many other mammalian clock genes were identified as homologues of *Drosophila* clock genes. Specifically, *BMAL1*, the mammalian homologue of *Cycle* (Hogenesch *et al.*, 1998; Bunger *et al.*, 2000), three *Period* homologues, *mPer1*, *mPer2* and *mPer3* (Albrecht *et al.*, 1997; Tei *et al.*, 1997; Takumi *et al.*, 1998) and a *Timeless* homologue (Koike *et al.*, 1998; Zylka *et al.*, 1998) and two *Cry* homologues (van der Horst *et al.*, 1999) were all identified as clock components in mammals following their discovery in *Drosophila melanogaster*.

These mammalian clock components function in a feedback loop broadly comparable to that of *Drosophila melanogaster*. In this feedback loop, a complex of CLOCK and BMAL1 acts as a transcriptional activator of the *Cry* and *Per* genes, mediated by E-box motifs in their promoters. As CRY and PER proteins accumulate, they associate and enter the nucleus, where they down-regulate the activity of the CLOCK/BMAL complex, consequently negatively regulating their own transcription. (Shearman *et al.*, 2000). Just as the stability and nuclear translocation of the PER/TIM complex in *Drosophila* is regulated by DOUBLETIME protein kinase, the mammalian complex of CRY and PER proteins in similarly regulated by protein kinases CKIϵ and CKIδ (Akashi *et al.*, 2002: Eide *et al.*, 2002).

This feedback loop differs somewhat from the equivalent loop in *Drosophila melanogaster* in that the CRY proteins are core clock components in the mammalian oscillator, whereas CRY is not essential for the function of the central oscillator in *Drosophila* (Krishnan *et al.*, 2001). A second difference lies in the role of the *Timeless* gene. The *mTim* gene has been shown to be essential for circadian rhythmicity in mice, leading some to propose that it fulfils the same role in the mammalian clock as in the *Drosophila* clock (Barnes *et al.*, 2003). However, in contrast to *Drosophila Tim*, mouse *Tim* RNA is not rhythmically expressed in the brain, suggesting a difference in regulation (Koike *et al.*, 1998; Zylka *et al.*, 1998).

In the mammalian clock, the CLOCK/BMAL1 complex is also involved in a second feedback loop. In this loop, the CLOCK/BMAL1 complex activates the transcription of two other genes, *Rev-Erbα* and *Rora* (Preitner *et al.*, 2002; Sato *et al.*, 2004). In turn, the products of these genes feedback to regulate the transcription of *BmalI*, where REV-ERBα acts as a negative regulator (Preitner *et al.*, 2002) and RORa acts as a positive regulator (Sato *et al.*, 2004). While this feedback loop employs different components to the *Drosophila* oscillator, this loop is strongly reminiscent of the PDP1/VRILLE/CLOCK/CYCLE feedback loop in *Drosophila melanogaster* in that antagonistic positive and negative activities converge on *Bmal1* just as positive and negative activities converge on *dClk* in *Drosophila*.

A summary of the current model of te mammalian circadian oscillator is shown in Figure 1.2.

Organisation of the Mammalian Clock

Similarities also appear in the anatomical organisation of the mammalian and *Drosophila* circadian systems, with a master oscillator in the brain thought to entrain various peripheral oscillators in other tissues. The mammalian

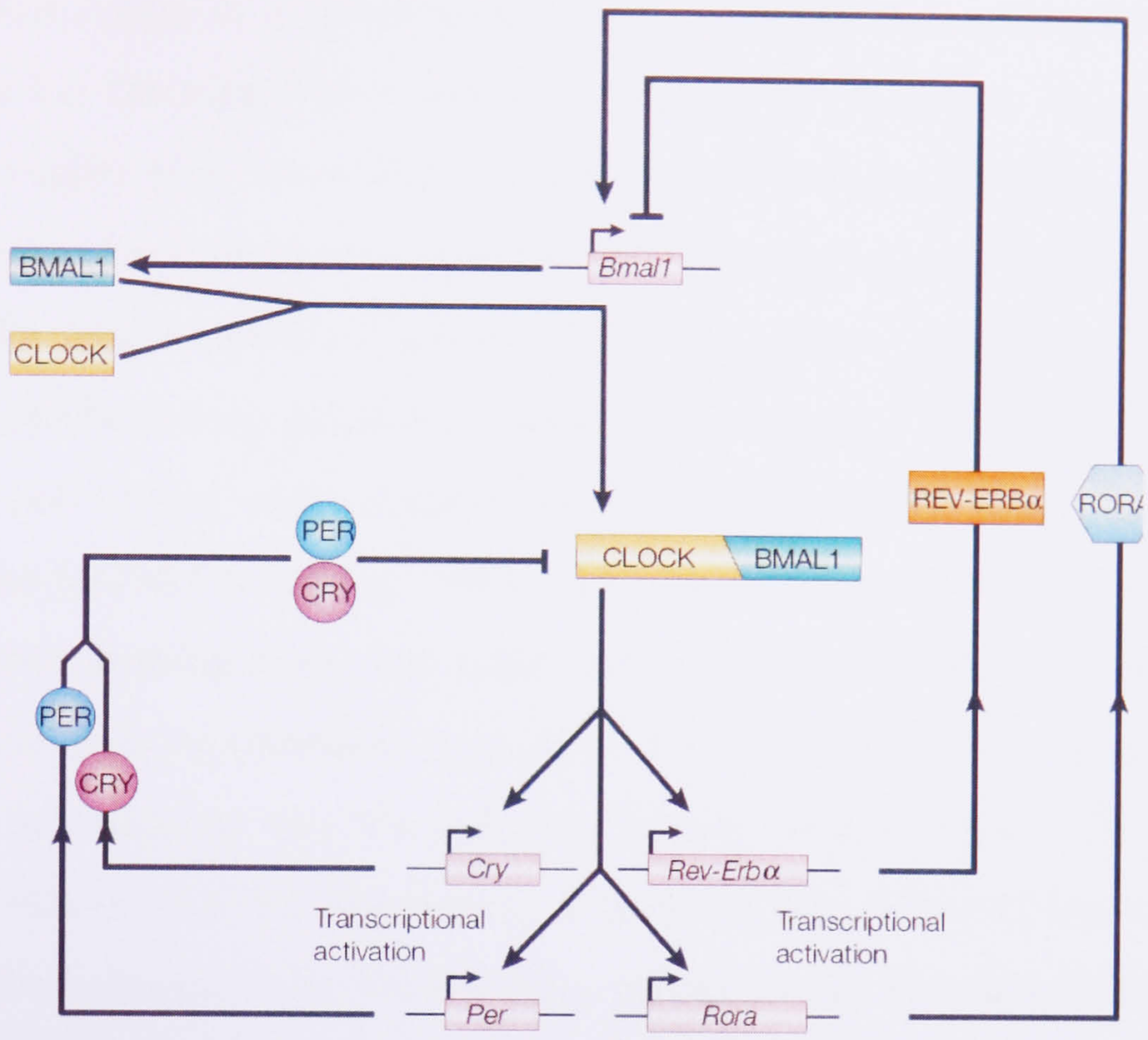


Figure 1.2: Structure of the mammalian oscillator mechanism. Figure taken from Bell-Pedersen, D. and Cassone, V.M. and Earnest, D.J. and Golden, S.S. and Hardin, P.E. and Thomas, T.L. and Zoran, M.L. (2005)

central oscillator is located in a region of the hypothalamus known as the suprachiasmatic nucleus which performs a function analogous to the clock cells in the *Drosophila* brain (Reppert & Weaver, 2002). As in the *Drosophila* central oscillator, different groups of cells in the mammalian SCN determine the phase of morning and evening activities allowing a degree of photoperiodic entrainment (Inagaki *et al.*, 2007).

1.2.3 The *Neurospora crassa* Clock

Neurospora crassa is a filamentous fungus that has also served as a model organism for circadian clock research. While the molecular components of the *Neurospora* clock show little conservation in vertebrate clocks, the overall architecture of coupled transcriptional feedback loops is very similar.

The best understood aspect of the *Neurospora crassa* oscillator is a negative feedback loop primarily consisting of the genes *frequency* (*frq*), *white collar-1* (*wc-1*) and *white collar-2* (*wc-2*). The first clock gene cloned in *N. crassa* was *frq* (McClung *et al.*, 1989). The *frq* gene is transcribed rhythmically with mRNA peaking in the late night and abundance of FRQ protein peaking in the early morning (Aronson *et al.*, 1994; Garceau *et al.*, 1997). Transcription of *frq* is activated by two transcription factors, White Collar-1 (WC-1) and White Collar-2 (WC-2), which form a complex (the White Collar Complex: WCC) which functions by binding to a *cis*-regulatory element, known as the Clock box (C-box), in the *frq* promoter (Froehlich *et al.*, 2003).

As FRQ protein accumulates during the day, it forms a complex with a second protein, FRQ-helicase (FRH) and the FRQ-FRH complex translocates to the nucleus (Cheng *et al.*, 2005). Once inside the nucleus, the FRQ-FRH complex functions to inhibit the activity of the WCC, thus negatively regulating activation of *frq* transcription, completing a negative feedback loop (Froehlich *et al.*, 2003). An additional positive feedback loop is formed by the activation of transcription of *wc-2* by FRQ, by some as yet unknown mechanism (Lee *et al.*, 2000; Cheng *et al.*, 2001).

As in the animal oscillator mechanisms described above, further regulation of this oscillator mechanism is imposed by the action of protein kinases on the oscillator components, notably casein kinases I and II, which serve to regulate the stability of FRQ (Liu & Bell-pedersen, 2006).

In addition to their role in the oscillator mechanism, WC-1 and WC-2

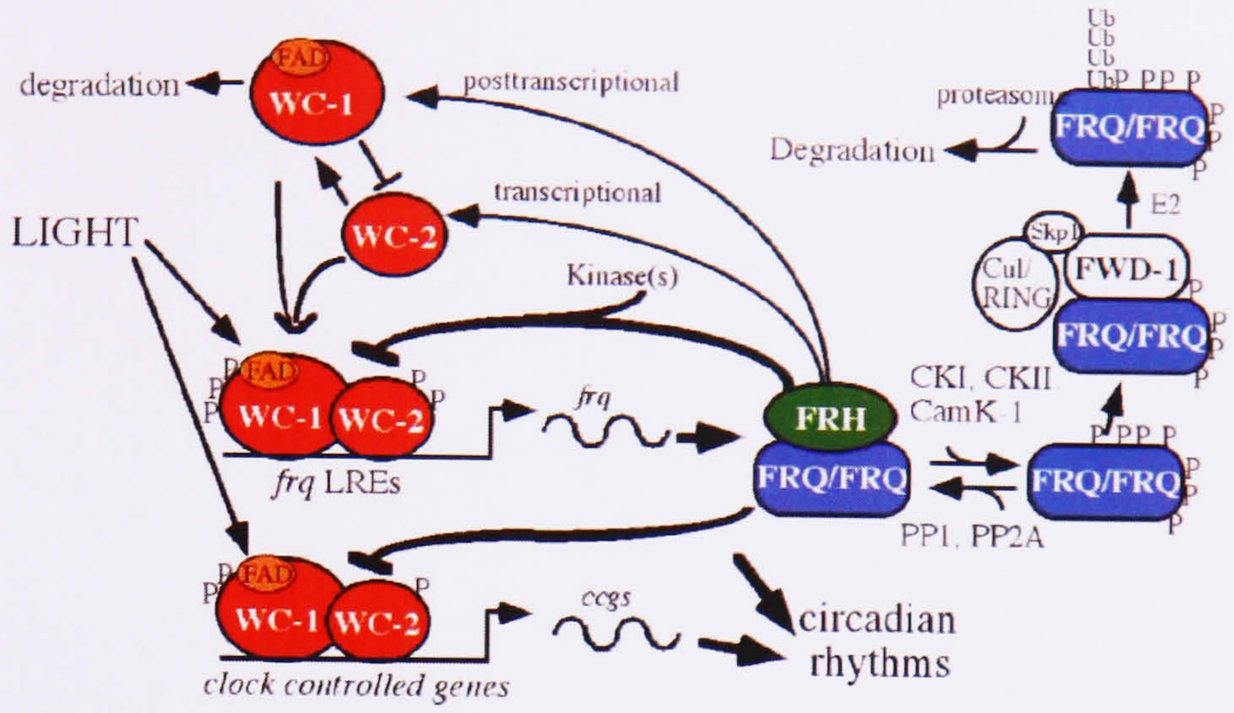


Figure 1.3: Structure of the *frq*-dependent oscillator mechanism of *N. crassa*. Figure taken from Liu & Bell-pedersen (2006).

also function in the light input pathway to the *N. crassa* oscillator. While WC-1 functions as a blue light receptor, WC-2 has no photoreceptor function but is required for the transcription factor activity of the complex (Cheng *et al.*, 2002; Froehlich *et al.*, 2002; He *et al.*, 2002). This light signalling pathway has been shown to act directly on the oscillator mechanism by activating transcription of *frq* (Crosthwaite *et al.*, 1995). The structure of this oscillator mechanism is summarized in Figure 1.3

In addition to the FRQ-WC-1/2 oscillator, one study has yielded evidence that suggested that a second oscillator mechanism might exist in *N. crassa* that is able to function in the absence of FRQ (Correa *et al.*, 2003). Correa *et al.* performed a large scale analysis of transcript abundance rhythms using oligonucleotide microarrays. This analysis identified three genes that were transcribed with a short period in a *frq* mutant strain that otherwise caused long period oscillations of most output genes. A later study demonstrated that these short period rhythms require WC-1 and WC-2 (de Paula *et al.*, 2006), raising the possibility that a second oscillator mechanism is cou-

pled to the FRQ-WC-1/2 oscillator.

1.2.4 The Cyanobacterial Clock

A fourth widely used model organism in circadian clock research is the cyanobacterium *Synechococcus elongatus*. The organization of the cyanobacterial oscillator appears to be somewhat different from eukaryotic oscillators in that it is able to function in the absence of transcription, to the extent that an oscillating system has been reconstituted *in vitro*, solely from purified clock proteins and ATP (Nakajima *et al.*, 2005).

Three essential components of the cyanobacterial clock are known: KaiA, KaiB and KaiC. These three components regulate the expression of each other through the mechanism of a transcriptional feedback loop in which KaiA acts as an activator of the *kaiBC* operon, and KaiC then feeds back to repress transcription of *kaiBC* (Ishiura *et al.*, 1998). However, this transcriptional feedback loop is not necessary for oscillator function since in darkness transcriptional activity ceases in *S. elongatus* but some circadian rhythms persist (Tomita *et al.*, 2005).

As the transcriptional feedback loop of the cyanobacterial clock components is not necessary for oscillator function, research has recently focused on phosphorylation of KaiC as a potential oscillator mechanism, since rhythmic phosphorylation of this protein persists in the absence of transcription (Tomita *et al.*, 2005). In this mechanism, the autokinase activity of KaiC serves to phosphorylate serine and threonine residues within the KaiC protein (Iwasaki *et al.*, 2002). This autokinase activity is enhanced by the presence of KaiA protein in complex with KaiC leading to a greater degree of KaiC phosphorylation. As KaiC phosphorylation increases, KaiB joins the KaiA:KaiC protein complex (Kageyama *et al.*, 2006; Nishiwaki *et al.*, 2007), where it inhibits the autokinase activity of KaiC (Kitayama *et al.*, 2003). Through an

as yet unknown mechanism, the cellular pool of KaiC protein then becomes de-phosphorylated and the complex dissociates, returning the system to the beginning of the cycle, closing the biochemical feedback loop.

The mechanism by which light signals are transduced to the oscillator in cyanobacteria is currently poorly understood. Just two genes have so far been implicated in light signalling to the cyanobacterial clock, *cikA* and *ldp1*. CikA protein is an atypical bacteriophytochrome with a histidine kinase domain and mutation of cikA causes defects in the phase resetting of the cyanobacterial oscillator by light signals (Schmitz *et al.*, 2000). A mutant allele of *ldpA*, however, causes a defect in the relationship of increasing light intensity to increasing free-running period of circadian rhythms (Katayama *et al.*, 2003). The function of these proteins in the light signalling pathway to the circadian oscillator remains unclear at the time of writing.

1.3 Common Themes in Clock Mechanisms

Despite the variation in clock components among different kingdoms, among the clock mechanisms that have been studied to date, several common themes have emerged. Firstly, transcriptional negative feedback loops seem to be a common structural motif in the circadian oscillator, forming the basis of the oscillator mechanism in insects, mammals and fungi though an interesting exception is noted in cyanobacteria, in which a post-translational feedback loop forms the core of the oscillator mechanism. Secondly, multiple feedback loops configured as coupled oscillators are beginning to emerge as a common paradigm in circadian oscillator structure.

It has been proposed that oscillators composed of multiple interlocked feedback loops might be more robust than simpler mechanisms, allowing reliable timekeeping in the face of environmental and stochastic perturbations

(Stelling *et al.*, 2004; Rand *et al.*, 2004). With sufficient light inputs to the oscillator, multiple feedback loops might also facilitate photoperiodic entrainment of the clock by allowing different loops to entrain to the timing of dawn or dusk (Rand *et al.*, 2004; Pittendrigh & Daan, 1997; Daan & Pittendrigh, 1978).

1.4 The Circadian Clock in *Arabidopsis thaliana*

Research into the molecular aspects of circadian clocks in plants was initiated rather later than in other eukaryotic models and the first plant circadian clock mutant, *toc1*, was identified in the model flowering plant *Arabidopsis thaliana* only in the mid 1990s (Millar *et al.*, 1995a).

As plants are sessile organisms that are reliant on light for photosynthesis, the perception and anticipation of daily and seasonal changes in the environment are implicated in a number of physiological processes from gene expression to growth and development. For example, in *Arabidopsis thaliana*, between 11% and 16% of genes are estimated to be regulated by the circadian clock at the level of transcription (Harmer *et al.*, 2000; Edwards *et al.*, 2006) and the circadian clock plays an important role in the photoperiodic control of flowering (Davis, 2002). Indeed, the circadian clock has been demonstrated to make a considerable contribution to fitness in *Arabidopsis* (Dodd *et al.*, 2005).

1.4.1 Transcriptional Feedback Loops in the *Arabidopsis thaliana* Circadian Oscillator

In common with other eukaryotic model organisms, the *Arabidopsis* circadian oscillator is believed to be composed of a system of interlocked transcriptional feedback loops with an additional level of regulation imposed at the post-translational level. The molecular components of the *Arabidopsis* oscillator,

however, seem to have little in common with the components of either the mammalian, *Drosophila melanogaster* or *Neurospora crassa* oscillators.

The LHY/CCA1-TOC1 Feedback Loop

The *toc1-1* (*timing of cab 1*) mutation was identified in a screen for mutants displaying aberrant expression rhythms of a *luciferase* reporter gene expressed from the clock-regulated *CAB* promoter (Millar *et al.*, 1995a). In *toc1-1* mutants, the circadian period of expression of this reporter was reduced from 24.7 hours to 20 hours.

Cloning of *TOC1* revealed that this gene encodes the founder member of the family of plant pseudo-response regulator proteins (Strayer *et al.*, 2000). This family of proteins share an amino-terminal domain similar to the response regulator receiver domain of two component signalling systems. however pseudo-response regulator proteins lack a conserved aspartate residue required for the function of true response regulator proteins (Mizuno, 2005). As one would expect of a clock-associated gene, *TOC1* is expressed rhythmically, with mRNA abundance peaking a few hours before dusk.

Functional analysis of *TOC1* has provided further evidence for its role as a component of the *Arabidopsis* circadian oscillator. Constitutive overexpression of *TOC1* led to an arrhythmic phenotype (Más *et al.*, 2003a), further demonstrating the role of *TOC1* in the generation of circadian rhtyhms. Interestingly, however, Más *et al.* also found that while reduction of *TOC1* mRNA levels by RNA interference (RNAi) caused a short period phenotype under constant white or blue light, consistent with the phenotype of the *toc1-1* mutation, rhythmicity was abolished under constant red light and in constant darkness, suggesting that the role of *TOC1* in the oscillator is closely associated with light conditions.

A second component of the *Arabidopsis* oscillator, *CCA1*, was identified

in a screen of a cDNA library for genes encoding factors able to bind the regulatory region of the *CAB2* gene (Wang *et al.*, 1997). This gene was found to encode a single MYB-domain DNA binding transcription factor that was shown to be rhythmically transcribed *in vivo* (Wang *et al.*, 1997; Wang & Tobin, 1998). A role for *CCA1* in the circadian clock was revealed when it was demonstrated that constitutive overexpression of *CCA1* caused arrhythmicity of clock outputs (Wang & Tobin, 1998). Interestingly, however, in the loss-of-function *cca1-1* mutant, clock outputs remained rhythmic, albeit with a markedly shortened period (Green & Tobin, 1999).

Almost concurrently with the identification of *CCA1* as a clock-associated gene, a second clock associated single MYB-domain transcription factor was identified in *Arabidopsis*. The *lhy-1* mutation was originally identified in a screen for effects on flowering time. This mutation was shown also to disrupt the circadian clock in *Arabidopsis*, causing arrhythmicity of various clock outputs (Schaffer *et al.*, 1998). Identification of the *LHY* locus revealed that this gene encodes a single-MYB domain transcription factor highly similar to CCA1. Further analysis of the *lhy-1* mutation demonstrated that this mutation causes overexpression of *LHY* (Schaffer *et al.*, 1998). However, like *CCA1*, *LHY* is not absolutely required for circadian clock function, as plants carrying *LHY* loss of function mutations or an *LHY* RNAi transgene display a rhythmic phenotype with a short period (Alabadí *et al.*, 2002; Mizoguchi *et al.*, 2002). However, loss of both *LHY* and *CCA1* abolishes clock function after several cycles in free-running conditions (Alabadí *et al.*, 2002; Mizoguchi *et al.*, 2002), demonstrating a degree of functional redundancy between LHY and CCA1.

These three components, *LHY*, *CCA1* and *TOC1*, have been shown to function together forming a transcriptional feedback loop (Alabadí *et al.*, 2001). Since levels of *LHY* and *CCA1* mRNA are depressed in *toc1-2* loss

of function mutants, *TOC1* was identified as an upstream activator of *LHY* and *CCA1* expression which is important for the up-regulation of these genes in the morning. Following the morning peak in transcription levels of *LHY* and *CCA1*, levels of LHY protein (and, presumably CCA1) peak during the mid-morning (Kim *et al.*, 2003) allowing LHY and CCA1 to fulfil their roles as transcriptional repressors of *TOC1* by binding to a *cis*-regulatory element in the *TOC1* promoter known as the evening element (Alabadí *et al.*, 2001). The resulting reduction in *TOC1* expression reduces the activation of *LHY* and *CCA1* expression, closing the transcriptional feedback loop.

It is not known how TOC1 serves as a transcriptional activator of *LHY* and *CCA1* since the only evidence for this function is genetic. However, TOC1 is unlikely to act directly on these genes since the TOC1 protein does not contain any known DNA binding motif (Strayer *et al.*, 2000). Furthermore, levels of TOC1 protein reach their maximum abundance during middle of the subjective night, several hours before the onset of the maximum rate of transcription of *LHY* and *CCA1* (Más *et al.*, 2003b). It therefore seems likely that some intermediate factor is required to transduce the transcriptional activation activity of TOC1 to the *LHY* and *CCA1* promoters.

This LHY/CCA1-TOC1 feedback loop is subject to further regulation at the post-translational level. For example, the cellular turnover of TOC1 protein by the proteasomal degradation pathway is modulated by the ZEITLUPE (ZTL) protein, which is required for the dark-dependent degradation of TOC1 (Más *et al.*, 2003b). This, in turn, is believed to play a part in determining the period of the circadian oscillator, since *ztl* mutants display a long period phenotype. While transcription of *ZTL* is not light-regulated, the dark dependence of this degradation suggests that the activity of ZTL might be regulated by light, raising the possibility that ZTL forms part of a light input pathway to the circadian oscillator mechanism (Más *et al.*, 2003b).

Levels of LHY protein are also subject to proteasomal degradation. regulated by the inhibitory factor DET1 (Song & Carré. 2005). This regulation is also believed to contribute to the period of the oscillator mechanism. since circadian rhythms run with a short period in *det1* mutants (Millar *et al.*, 1995b).

The *Arabidopsis* oscillator mechanism also appears to be subject to regulation by Casein Kinase 2 (CK2). Phosphorylation of CCA1 by CK2 is necessary for the DNA-binding activity or CCA1 and its function in the oscillator mechanism (Daniel *et al.*, 2004). However, it is not currently known what the role CCA1 phosphorylation might play in the regulation of oscillator function.

The LHY/CCA1-PRR7/9 Feedback Loop

The LHY/CCA1-TOC1 feedback loop described above is consistent with many experimental observations of the *Arabidopis* circadian clock (Locke *et al.*, 2005a), however, it does not account for the rhythmic phenotype of *toc1* mutants.

More recently, a second feedback loop has been proposed, converging on *LHY* and *CCA1* in which the TOC1-related pseudo-response regulator proteins PRR7 and PRR9 act redundantly as repressors of *LHY* and *CCA1* expression (Farré *et al.*, 2005).

Although loss of function of either PRR7 or PRR9 singly only produces mild long period phenotypes, plants in which both genes were mutated displayed an extremely long period phenotype of around 36 hours, suggesting that these genes play an important role in the *Arabidopsis* clock and share a high degree of functional redundancy (Farré *et al.*, 2005). Since *LHY* and *CCA1* were expressed with a greater amplitude and displayed an earlier onset expression in *prr7;prr9* double mutants, it was proposed that expression of

LHY and *CCA1* is normally repressed by PRR7 and PRR9. Interestingly, overexpression of *LHY* or *CCA1* led to elevated levels of expression of *PRR7* and *PRR9*, suggesting that LHY and CCA1, in contrast to their other role as repressors of *TOC1* expression, also act as activators of *PRR7* and *PRR9* expression. Together, these regulatory relationships were proposed to constitute a second transcriptional negative feedback loop in the *Arabidopsis* circadian oscillator mechanism.

1.4.2 Other Clock Associated Genes in *Arabidopsis*

Several other genes have been identified as forming part of the Arabidopsis clock although the placement of these genes in the circadian regulatory network remains to be determined. Many of these genes are themselves clock regulated raising the possibility that they might form additional regulatory feedback loops within the oscillator mechanism.

LUX

LUX ARRHYTHMO (*LUX*) encodes a single Myb-domain protein putative transcription factor which is required for free-running rhythmicity in constant light and robust rhythmicity in darkness (Hazen *et al.*, 2005). The *LUX* gene is expressed rhythmically with a similar phase to *TOC1* where expression peaks around dusk. Since, like *TOC1*, the *LUX* promoter contains an Evening Element motif, which is recognized by LHY and CCA1 *in vitro*, it has been suggested that *LUX* and *TOC1* might be co-regulated by LHY and CCA1.

It has also been suggested that *LHY* and *CCA1* are regulated by LUX, since expression levels of these genes are depressed in *lux* mutants, suggesting that LUX acts to activate their expression in wildtype plants (Hazen *et al.*, 2005). Constitutive overexpression of *LUX* causes rapid damping and eventual loss of circadian rhythmicity, including the rhythmic expression of

LHY, *CCA1*, *TOC1* and the endogenous *LUX* gene, demonstrating that *LUX* functions as part of an autoregulatory feedback loop (Onai & Ishiura, 2005). Whether LUX forms part of an already described feedback loop, by acting in concert with TOC1, for example, or an additional feedback loop is not yet known.

The *PSEUDO-RESPONSE REGULATOR Family*

As well as *TOC1*, *PRR7* and *PRR9*, the *Arabidopsis* pseudo-response regulator family contains two other members, *PRR3* and *PRR5* (Makino *et al.*, 2000).

Aside from the similarity of the proteins encoded by these genes, another interesting property of this family is that each member is transcribed with a circadian rhythm in a sequence beginning with *PRR9*, followed by *PRR7*, *PRR5*, *PRR3* and finally *TOC1*, with the peak of expression of each separated by 2 to 3 hours (Matsushika *et al.*, 2000). This sequential pattern of circadian regulation led to the suggestion that these genes might form a transcriptional cascade within the circadian clock, each activating transcription of the next, serving to add a time delay to the oscillator mechanism. However, this is unlikely to be wholly true since loss of function of *PRR9*, *PRR7* or *PRR5* has only mild period-altering effects (Eriksson *et al.*, 2003; Yamamoto *et al.*, 2003).

Null mutations of *PRR5* have a small, but significant effect on the circadian clock, causing the clock to run with a period 1 to 2 hours shorter than in wildtype plants (Eriksson *et al.*, 2003; Yamamoto *et al.*, 2003). Overexpression of *PRR5*, on the other hand, causes low amplitude expression of *LHY* and *CCA1*, which damps further in constant light (Sato *et al.*, 2002). Interestingly, overexpression of *PRR3* has a different effect, causing the clock to run with a long period in constant light with no effect on the amplitude of output rhythms, suggesting that *PRR3* and *PRR5* play different roles in the

Arabidopsis circadian system (Murakami *et al.*, 2004).

Altering the expression of either *PRR3* or *PRR5* by mutation or over-expression also causes notable changes in the de-etiolation of seedlings by red light, suggesting that both these genes play a part in light signal transduction (Murakami *et al.*, 2004; Yamamoto *et al.*, 2003). Again, these results suggest different functions for *PRR3* and *PRR5*, since overexpression of *PRR3* causes hypersensitivity of the de-etiolation response to red light (Murakami *et al.*, 2004), whereas overexpression of *PRR5* reduces sensitivity (Yamamoto *et al.*, 2003).

ELF3

The *ELF3* gene appears to play a light dependent role in the circadian clock since while putative loss of function mutations of *ELF3* have relatively little effect on rhythmicity under light-dark cycles or under constant darkness, certain alleles of *elf3* cause the clock to arrest under constant light (Hicks *et al.*, 1996). This conditional circadian phenotype suggested a role for *ELF3* outside, but closely associated with the central oscillator. It has been proposed that *ELF3* might function in "gating" signals to the oscillator mechanism, a process whereby a circadian rhythm is imposed on the perception of light signals by the clock . This hypothesis is supported by evidence that the circadian regulation of the light-inducibility of the *CAB* gene is abolished in the *elf3-1* null mutant (McWatters *et al.*, 2000).

Low levels of *LHY* mRNA expression in the absence of functional *ELF3* suggest that *LHY* may be a target for transcriptional activation by ELF3 (Schaffer *et al.*, 1998). Because ELF3 does not contain an obvious DNA-binding domain, it is expected that this protein might exert its action on *LHY* transcription through association with other factors (Liu *et al.*, 2001).

ELF4

The clock-regulated *ELF4* gene, which encodes a protein unrelated to ELF3, has also been implicated in the *Arabidopsis* circadian system. A mutation of *ELF4* has been described (*elf4-101*) which disrupts the regulation of circadian rhythms in both constant light and constant darkness (Doyle *et al.*, 2002). Interestingly, this mutation caused arrhythmicity of certain clock outputs whereas other clock outputs displayed variations in their circadian period between individual plants. Since the rhythmic expression of the oscillator component *CCA1* was disrupted in *elf4-101* plants, it seems likely that *ELF4* functions either within, or upstream of the circadian oscillator mechanism.

ELF4 has been implicated in light signalling pathways in *Arabidopsis* (Khanna *et al.*, 2003), and in the absence of a functional copy of *ELF4*, the induction by light of both *LHY* and *CCA1* expression in etiolated seedlings, and the following initiation of circadian rhythms are abolished (Kikis *et al.*, 2005). In a similar effect to that reported for elf3 mutations, loss of *ELF4* function also affects the gating of light signals by the circadian clock (McWatters *et al.*, 2007)

McWatters *et al.* (2007) have suggested that *ELF4* plays a role within the oscillator mechanism itself, since *LHY*, *CCA1* and *TOC1* are all expressed arrhythmically under constant conditions in the *elf4-1* mutant. Furthermore, this mutation was shown to cause various phase defects depending on the photoperiod of the entraining cycle, suggesting *ELF4* is required for correct photoperiodic entrainment of the clock. However, the molecular role of *ELF4* and its position in the circadian regulatory network remains unclear.

TIC

The *time for coffee* (*tic*) mutation was identified in a screen for mutations displaying a defective expression rhythm of a *CAB2::luc* reporter gene (Hall *et al.*,

2003). Though remaining rhythmic, the amplitude of rhythms of *CAB2::luc* expression was notably reduced in *tic* mutants and individual *tic* plants showed a range of lengthened circadian periods. Like mutations to *ELF3* and *ELF4*, *TIC* has also been implicated in the gating of light signalling by the clock, since this gating is less effective in the *tic-1* mutant (Hall *et al.*, 2003).

1.4.3 Light Signalling in *Arabidopsis thaliana*

The perception of light by the oscillator is, of course, vital for entrainment to environmental light:dark cycles. in *Arabidopsis*, both *LHY* and *CCA1* are known to be inducible by light (Viczián *et al.*, 2005; Wang & Tobin, 1998).

In *Arabidopsis*, light signals are perceived by a wide array of photoreceptors (Casal & Yanovsky, 2005). Five members of the phytochrome family are responsible for the perception of red and far red light (PHYA-PHYE) whereas blue light is detected by two cryptochrome (CRY1 and CRY2) proteins and two phototropins (PHOT1 and PHOT2). While the mechanism by which blue light signals are transduced to gene expression is poorly understood, the mechanisms of phytochrome-mediated light regulation of gene expression, especially induction of expression by red light, have been more extensively described.

On exposure to red light, PHYB under goes a reversible transition from its inactive (Pr) form to its active (Pfr) form and translocates to the nucleus (Yamaguchi *et al.*, 1999). Once in the nucleus, PHYA (Pfr) associates with the transcription factor PIF3, which consitutively binds to a G-box *cis*-element in target promoters, activating transcription (Ni *et al.*, 1998; Martínez-García *et al.*, 2000).

Interestingly, light input to the circadian oscillator is known to be gated by the imposition of a circadian rhtyhm of photoreceptor abundance (Toth *et al.*, 2001).

1.5 Aims

Both *LHY* and *CCA1* occupy key positions in the circadian oscillator of *Arabidopsis thaliana*. These genes are subject to regulation from a least two negative feedback loops as part of the oscillator mechanism as well as being downstream targets of a number of other clock related genes. As transcription of *LHY* and *CCA1* is also light regulated, it is likely that the regulation of expression of these genes plays a role in the entrainment of the *Arabidopsis* circadian clock by light.

Little is known of how the various regulators of *LHY* and *CCA1* transcription function at the molecular level, since few of these proteins contain domains associated with transcriptional regulation making it likely that they function as part of complexes with other, DNA-binding factors.

The work presented in this thesis aims to dissect the transcriptional regulation of *LHY* to determine:

1) How many circadian signals converge on the *LHY* promoter

2) How many light signals converge on the *LHY* promoter

3) Which cis-regulatory sequences mediate these signals

4) How these signals contribute to the regulation of *LHY* expression

It is hoped that this information will provide further insight into the role of *LHY* in the *Arabidopsis* circadian system and the wider structure of the circadian regulatory network.

Chapter 2

Methods

2.1 Methods for Molecular Cloning

Unless otherwise stated, routine molecular biology procedures such as phenol:chloroform extraction, ethanol precipitation of DNA, DNA ligation and agarose gel electrophoresis, were performed as described by Sambrook *et al.* (1989). Enzymatic manipulations of DNA were carried out as instructed by the enzyme suppliers.

Sequencing of DNA was carried out by the University of Warwick Molecular Biology Service. Sequencing reactions were performed using the Big Dye Terminator Version 3.1 system (Applied Biosystems, Warrington, UK) and were analysed with an Applied Biosystems 3130xl Genetic Analyser.

2.1.1 Extraction of DNA from Agarose Gels

To extract DNA from agarose gel slices, the conical section of a microfuge tube was cut from the remaining section of the tube and a hole pierced in the apex. A ball of polyester wool (available as fishtank filter medium), or approximately 100 mm^3, was placed in the apex of this section, which was then placed into the neck of an intact microfuge tube.

The gel slice was placed on top of the polyester wool substrate and the assembly was centrifuged at 13,000 x g to separate the DNA-containing liquid phase from the solid phase of the gel. Ethidium bromide was removed from the liquid phase by three rounds of phenol:chloroform extraction. The DNA was then ethanol precipitated and resuspended in the required volume of TE buffer [10 mM Tris, 1mM EDTA, pH 7.5]

2.1.2 Preparation and Transformation of Competent *Escherichia coli.*

Chemically competent *E. coli* cells, strain XL1 Blue, were prepared using the rubidium chloride method described by Hanahan (1985) and transformed using the 37 °C heat-shock method described by Sambrook *et al.* (1989).

Transformants were selected on Luria-Bertani (LB; Sambrook *et al.*, 1989) agar supplemented with either kanamycin sulphate (50 μg/ml) or ampicillin (100 μg/ml)

2.1.3 Purification of Plasmid DNA from *E. coli* cultures.

A single colony of *E. coli* cells was used to inoculate 5 ml of LB broth, containing the appropriate antibiotic for selection of the plasmid. The inoculated cultures were incubated at 37 °C, and shaken at 250 rpm, overnight. Cells were harvested from 3 ml of this culture by centrifugation and plasmid DNA was extracted using the QIAprep® Spin Miniprep Kit (QIAGEN Ltd.)

2.1.4 Site-directed Mutagenesis.

Site-directed mutagenesis of promoter constructs was carried out using the QuikChange^{TM} system (Stratagene, La Jolla, CA. USA) with an additional

pre-PCR primer elongation step to increase the efficiency of mutagenesis (Wang & Malcolm, 1999). Successfully mutated plasmids were identified by sequencing plasmid DNA extracted from transformed *E. coli* colonies.

The primers used for site-directed mutagenesis are described in Table 2.1

2.2 Growth of *Arabidopsis thaliana.*

2.2.1 Sterile Growth

Seeds were surface-sterilized by washing in a solution of 50% (v/v) bleach (Fisher Scientific, Loughborough UK), 0.05% (v/v) Tween® 20 (Sigma-Aldrich, Gillingham UK) for 5 to 10 minutes followed by three rinses in sterile, distilled water.

Sterile seed was sown either on Murashige and Skoog (MS) agar [4.3 g/l Murashige and Skoog basal salt mix (Sigma-Aldrich), pH 5.7, 1.5% (w/v) agar], or, where stated, MS agar with the addition of 3% (w/v) sucrose (MS3).

Once sown, seeds were incubated at 4 °C for four days to synchronize germination. Growth conditions are described below, as appropriate.

2.2.2 Greenhouse Growth

Plants grown in soil were potted in a mixture of equal parts compost (B&Q) and vermiculite (B&Q) at 22 °C in a glasshouse under ambient light conditions supplemented with 16 hours per day sodium lamp light.

2.2.3 Genetic Crosses

To perform genetic crosses between plants, the sepals, petals and stamens were removed from an unopened floral bud of the female parent. The exposed stigma

Primer	Sequence
5′ pal F	GCCACTACAATATCACCACGTGGTGATCTGCGATGACTTCTGG
5′ pal R	CCAGAAGTCATCGCAGATCACCACGTGGTGATATTGTAGTGGC
3′ pal F	GCCACTACAATATCGACACGTGTCGATCTGCGATGACTTCTGG
3′ pal R	CCAGAAGTCATCGCAGATCGACACGTGTCGATATTGTAGTGGC
St Cl II F	GACAGCCACTACAATATCGTCACGTGACGATCTGCGATGACTTC
St Cl II R	GAAGTCATCGCAGATCGTCACGTGACGATATTGTAGTGGCTGTC
Wk Cl II F	GACAGCCACTACAATATCCTCACGTGAGGATCTGCGATGACTTC
Wk Cl II R	GAAGTCATCGCAGATCCTCACGTGAGGATATTGTAGTGGCTGTC
Δctr1 F	CTTCTTCTTCTTCTTCTTCAGTTATGTTTTTTAAATTTATTTTTAGAG
Δctr1 R	CTCTAAAAATAAATTTAAAAAACATAACTGAAGAAGAAGAAGAAGAAG
Δctr2 F	CAGTCTTCTTCAGCCTAAAACAGTGTTTTTTAAATTTATTTTTAGAG
Δctr2 R	CTCTAAAAATAAATTTAAAAAACACTGTTTTAGGCTGAAGAAGACTG
Δctr3 F	GGCTGAGATTGCTTCTGGCTTCTGTTTTTTAAATTTATTTTTAGAG
Δctr3 R	CTCTAAAAATAAATTTAAAAAACAGAAGCCAGAAGCAATCTCAGCC
Site 1M F	GCAACTTGAGATATATGTCAAAGTGCAGTAGACAGCCACTAC
Site 1M R	GTAGTGGCTGTCTACTGCACTTTGACATATATCTCAAGTTGC
Site 2M F	GATCTGCGATGACTTCTGTTTTGACAATTTATACCCTTGGTGTTCC
Site 2M R	GGAACACCAAGGGTATAAATTGTCAAAACAGAAGTCATCGCAGATC
Site 345M1 F	CCAAAAATTAGGTGTCAAATTGTTGTGGCTGAGATTGCTTCTGGC
Site 345M1 R	GCCAGAAGCAATCTCAGCCACAACAATTTGACACCTAATTTTTGG
Site 345M2 F	GCCTCAAATAAACTTTTCAATTAAAATTTACTGTCAAATTAGGTG
Site 345M2 R	CACCTAATTTGACAGTAAATTTTAATTGAAAAGTTTATTTGAGGC

Table 2.1: Primers used in site directed mutagenesis

was then dusted with pollen by brushing with anthers taken from opened flowers of the male parent. Seeds were harvested once the silique had ripened and turned yellow.

2.3 Generation of Transgenic *Arabidopsis thaliana* Plants

Wild-type (Wt) and mutant plants used in this work were either Columbia (Col) or Wassilewskija (Ws) accessions, as stated.

All genetic constructs for eventual transformation into *Arabidopsis thaliana* were constructed in the pGreen binary vector (Hellens *et al.*, 2000).

2.3.1 Preparation and Transformation of Competent *Agrobacterium tumefaciens*

Competent *Agrobacterium tumefaciens* cells were prepared from strain C58, harbouring the pSoup plasmid (Hellens *et al.*, 2000). A 50 ml culture of *A. tumefaciens* was grown at 28 °C in LB broth, supplemented with 10 μg/ml tetracycline, to an OD_{600} of 0.8. These cells were pelleted by centrifugation at 2000 x *g* for 15 minutes at 4 °C and resuspended in 1 ml of sterile 20 mM calcium chloride solution. Cells were then snap frozen in liquid nitrogen and stored at -70 °C.

Competent cells were transformed using the freeze-thaw method of Höfgen & Willmitzer (1988), with the exception than LB broth was used in place of YEB broth. Following transformation, cells were plated on LB agar supplemented with 50 μg/ml kanamycin sulphate, to select for the pGreen plasmid, and 10 μg/ml rifampicin to select for *A. tumefaciens* C58. Plates were incubated at 28 °C for upto three days until colonies appeared.

2.3.2 Transformation of *Arabidopsis thaliana*

Seeds were sown directly on to soil, eight to ten per pot, and incubated under cloches until seedlings emerged. Primary bolts were removed to encourage further bolting. Plants were transformed by the method of floral dip (Clough & Bent, 1998) when secondary bolts were between 5 cm and 10 cm tall.

2.3.3 Selection of Transformed Plants

To select transformed seedlings of the first transgenic generation (T1), groups of ten to twenty seeds were arrayed on MS agar. After five days sterile growth, seedlings were sprayed with 5 mM luciferin spray solution and assayed for luciferase activity using a photon counting camera (see below). Seedlings from any groups displaying luciferase activity were then transferred to fresh MS agar plates and spaced approximately 1 cm apart. After a further seven days growth, these seedlings were once again treated with luciferin spray solution and re-assayed for luciferase activity to identify individual transformed seedlings. Once identified, transformed seedlings were transferred to soil and allowed to self-fertilize to generate T2 seed.

2.4 *In vivo* Analysis of *luciferase* Transcriptional Reporters

Surface sterilized seeds were sown on MS3 agar either 1 cm to 1.5 cm apart or in groups of 10 to 60 seeds, bounded by "collars" made from the cut-off necks of microfuge tubes.

Plants were grown under the stated photoperiod and light quality at 22 °C for seven days prior to imaging.

For light-induction experiments, plants were exposed to ambient light

conditions for 2 hours to stimulate seed germination before 7 days incubation at 22 °C in darkness. Dark-grown plants were additionally wrapped in aluminium foil during the incubation period as a safeguard against accidental light exposure.

2.4.1 Image Acquisition

One day before the start of the imaging experiment, plants were pre-sprayed with luciferin spray solution [5mM luciferin, 0.01% (v/v) Triton® X-100] to destabilize accumulated luciferase protein. On the first day of the experiment, plants were sprayed once more with luciferin spray solution before being transferred to the imaging chamber.

During imaging, ambient temperature was maintained at 22 °C. Light conditions were as described in figure legends. Illumination was provided by custom made red or blue LED arrays.

Luminescence was monitored by digital imaging of plants using either the ORCAII c4742-98 CCD camera system (Hamamatsu (UK) Ltd. Welwyn Garden City, UK) or a liquid nitrogen cooled TEK 512x512DB CCD with an ST138 controller (Princeton Instruments Inc. Trenton, New Jersey). Automated imaging protocols, including control of illumination, were created and run using the Metamorph^{TM} software package (Molecular Devices Ltd. Wokingham, UK). Images were acquired either every two hours for experiments in illuminated conditions, or every thirty minutes for experiments using dark conditions.

2.4.2 Image Analysis

Numerical values for luminescence intensity were extracted from 16-bit image files generated by the cameras using the Metamorph^{TM} package. Background intensities were calculated for each image from regions containing no plants

and subtracted from each data point to correct for noise arising from the limitations of the cameras.

2.5 Electrophoretic Mobility Shift Assays

2.5.1 Preparation of *Arabidopsis* Whole-Cell Protein Extracts

Arabidopsis seedlings were grown on MS-agar for ten days under cycles of 12 hours of cool white light (90 μmol m^{-2} s^{-1}) and 12 hours of darkness at 22 °C. After ten days, plants were transferred to constant white light and harvested into liquid nitrogen at the stated intervals.

Frozen plant material was ground to a fine powder under liquid nitrogen. One gram of ground tissue was allowed to thaw in a mixture of 3ml of ice cold extraction buffer [15 mM HEPES, pH 7.5, 40 mM KCl, 10 mM NaF, 12.5 mM EDTA, 1 mM DTT, 1% (v/v) plant protease inhibitor cocktail (Sigma Aldrich)], and 0.3 ml 4 M ammonium sulphate. The thawed mixture was filtered through miracloth (Calbiochem, Nottingham UK) to remove larger solids and the filtrate was incubated on ice for 30 minutes with mild agitation before centrifugation at 100,000 x g for 1 hour. After centrifugation, the supernatant was decanted into a small beaker and stirred on ice while finely ground ammonium sulphate was slowly added to precipitate the protein. The suspension was stirred for a further 30 minutes and was then centrifuged at 14,000 x g for 10 minutes at 4 °C to recover the precipitaed proteins. Protein recovered by centrifugation was resuspended in 100 μl resuspension buffer [15 mM HEPES, pH 7.5, 40 mM KCl, 10 mM NaF, 12.5 mM EDTA, 10% glycerol, 1 mM DTT, 1% (v/v) plant protease inhibitor cocktail].

Protein extracts were desalted using 0.5 ml Zeba^{TM} desalt columns (Pierce, Cramlington UK), equilibrated with resuspension buffer, according to

Primer	Sequence
H286 D1New	TTAATGCCGGCTTCTGGCTCGTAGAGAAGCAAC
B157 R1New	GGGGCCGGCGGTATAATGGAAAAAACAGAAGTCATCG
H163 D1New	TAAGCCGGCCTTGGGGTGCTGTTCCAGCCTCA
B73 R1New	AAAAGCCGGCGGGGGATCCGCCAGAAGCAATCTCAG

Table 2.2: PCR primers used to generate EMSA probes. NgoMIV sites are underlined

the manufacturers protocol.

Protein concentrations were determined for desalted extracts by BCA protein assays (Pierce) calibrated against a set of bovine serum albumin standards of known concentration.

2.5.2 Preparation of Probes.

Probes for electrophoretic mobility shift assays were prepared either from cloned DNA or by annealing of complementary custom synthesized oligonucleotides (VHBio Ltd, Gateshead UK).

Cloned probes were generated by PCR amplification of the plasmid template p*LHY:luc* (Song, 2002) with Taq polymerase. The primers used in amplification, described in Table 2.2, were designed to incorporate an NgoMIV site at each end of the PCR product.

PCR products were ligated into the pCR®2.1 vector (Invitrogen, Paisley UK) and propagated in *E. coli*, strain XL1 Blue.

Plasmid DNA was purified from 15 ml of overnight culture and digested with restiction enzyme NgoMIV. Digested DNA fragments were separated by agarose gel electrophoresis and the appropriate band was excised and DNA purified from the gel slice.

Purified probe DNA was labelled by end-filling of NgoMIV sites with

α^{32}P-dGTP (Amersham, Amersham UK), catalysed by Klenow enzyme (Fermentas, York UK). Unincorporated nucleotides were subsequently removed by passing the reaction mixtures through 0.5ml Zeba^{TM} desalt columns.

Oligonucleotide-derived probes were generated by 5′ phosphorylation of single stranded oligonucleotides with γ^{32}P-ATP (Amersham) using T4 polynucleotide kinase (Fermentas), according to the manufacturers protocol. Again, unincorporated radiolabel was removed on Zeba^{TM} desalt columns. Labelled oligonucleotides were combined in annealing buffer [10 mM Tris, pH 7.5, 100 mM NaCl] and annealed by heating to 95 °C for 10 minutes followed by cooling to 20 °C at a rate of 1 °C per minute. Unlabelled competitor DNA was prepared by similarly annealing custom oligonucleotides.

2.5.3 Preparation of Binding Reactions

Binding reactions contained either 10 or 20 fmol of probe, as stated, and, where present, 5 μg of protein extract. Components of binding reactions were combined in 12 μl total volume of binding buffer [25 mM HEPES-NaOH, pH 7.5, 50 mM KCl, 0.1 mM EDTA, 20% glycerol, 1 mM DTT, 1% (v/v) plant protease inhibitor cocktail] and incubated at room temperature for 30 minutes prior to electrophoresis.

2.5.4 Electrophoresis and Visualization of EMSAs.

Complexes and probe were resolved by polyacrylamide gel electrophoresis in either 0.25 x TBE (Sambrook *et al.*, 1989) or HEPES-EDTA buffer [40 mM HEPES, 0.2 mM EDTA, pH 7.5] as stated. Electrophoresis was carried out at 10 V cm^{-1}, at room temperature on 2 mm thick gels.

2.6 RNA Extraction and Northern Analysis

2.6.1 Extraction of RNA from plant tissue

Plants were grown as for whole-cell protein extracts and tissue was harvested at the required intervals after transfer to constant light conditions. Harvested tissue was ground to a fine powder under liquid nitrogen and stored at -70 °C until required.

Total RNA was extracted from around 0.75 ml of ground tissue using the RNeasy® Plant Mini Kit (QIAGEN), according to the manufacturers instructions and quantified by spectrophotometric absorbance at 280 nm.

2.6.2 Separation and Northern transfer of RNA species

Equal amounts (10 μg) of RNA from each sample were electrophoresed on a 1.2% formaldehyde-agarose gel, stained with ethidium bromide (1 ng/l) (Sambrook *et al.*, 1989). alongside 3 μg of 0.24-9.5 Kb RNA ladder (Invitrogen).

RNA was blotted to a Hybond NX nylon membrane (Amersham) as previously described (Sambrook *et al.*, 1989). After blotting, the membrane was allowed to dry completely before RNA was immobilized by treating both side of the membrane with cross-linking with 120 J of 254 nm ultraviolet light.

2.6.3 Generation and hybridization of probes

Radiolabelled probes to *LHY* and *luciferase* (*luc*$^+$) RNA were generated by random priming of gel purified *LHY* or *luc*$^+$ cDNA templates in the presence of α^{32}P-dCTP, using the High Prime DNA Labelling Kit (Roche Applied Science.) as directed by the manufacturer.

Pre-hybridization, hybridization and washing steps were carried out as directed by Sambrook et al. (1989). Between hybridizations, membranes were

stripped by washing in 0.1x SSC [150 mM NaCl, 15 mM sodium citrate], 0.1% SDS at 90 °C for 5 minutes.

2.6.4 Detection

Hybridization was visualized by exposing hybridized membranes, wrapped in clingfilm, to phosphorimager screens overnight. Screens were scanned using a phosphorimager. Hybridization was quantified using Metamorph^{TM} software to measure band intensities.

2.7 Bioinformatic Methods

2.7.1 Plant *cis*-Element Databases

Regions of the *LHY* promoter were scanned for previously described *cis*-regulatory elements using the PlantCARE database (Lescot *et al.*, 2001), accessible at:

`http://bioinformatics.psb.ugent.be/webtools/plantcare/html/`

and the PLACE database (Higo *et al.*, 1999), at:

`http://www.dna.affrc.go.jp/PLACE/`.

2.7.2 Identification of *cis*-Elements Through Homology

To identify orthologues of *LHY* in other plant species, BLAST searches were performed using the tBLASTn program to align the LHY and CCA protein sequences against the Genbank expressed sequence tag (EST) database, limited to entries from the Viridiplantae but excluding *Arabidopsis thaliana* entries. The default BLAST parameters were used other than that the low complexity filter was disabled.

For species for which high throughput genomic sequence (HTGS) data

was also available, EST BLAST hits were assembled manually into contiguous putative transcripts which were then used to search the Genbank HTGS database for ortholgues of the *Arabidopsis thaliana LHY* and *CCA1* genes using the BLASTn program. Again, the default BLAST parameters were employed, other than the omission of the low complexiy filter.

Chapter 3

Delimitation of the *cis*-Regulatory Regions of *LHY*

3.1 Introduction

As described in Chapter 1, *LHY* and *CCA1* are subject to regulation by several clock associated genes including among others, *TOC1*, *ELF3*, *ELF4*, *LUX*, *PRR7* and *PRR9*. Both *LHY* and *CCA1* are also targets of light signalling since both are rapidly induced by light signals in etiolated seedlings. The number and range of regulatory inputs that converge on *LHY* and *CCA1* place these genes at a key node in the *Arabidopsis* circadian clock. Further understanding of the regulation of expression of these genes is, therefore, an important first step in better understanding their role in the circadian clock and gaining further insight into the structure of the circadian regulatory network.

This chapter describes efforts to identify regions of the *LHY* upstream sequence and untranslated regions (UTRs) that contain the *cis*-regulatory elements that mediate the various circadian and light-responsive inputs that define *LHY* expression. Once these regions have been delimited, finer map-

ping of regulatory elements can be performed.

Here, expression of *LHY* was monitored *in vivo* using a *luciferase* reporter gene under the control of putative *LHY* regulatory sequences. Firefly *luciferase* is an especially valuable reporter gene in circadian research because changes in reporter expression can be monitored in intact, living organisms by monitoring luminescence with a photon counting CCD camera. Furthermore, because luciferase activity is unstable in the presence of its substrate, luciferin, levels of luminescence closely track levels of the *luciferase* mRNA. This property makes *luciferase* particularly useful for tracking dynamic changes in gene expression.

3.2 Circadian Regulation

3.2.1 Experimental Design

To investigate the circadian regulation of *LHY* expression, the expression of *luciferase* reporter constructs was examined under light:dark cycles (driven rhythmicity) and under constant conditions (free-running rhythmicity). Plants were grown under cycles of 12 hours of white light and 12 hours of darkness (12L 12D) for 7 days prior to data collection. Once transferred to the imaging apparatus, data were collected first under another 2 or 3 days of light-dark cycles, to examine driven rhythmicity, followed by 4 to 5 days of constant conditions, to examine free-running rhythmicity. After characterization of a pseudo-wildtype reporter construct, regulatory regions were identified by systematically deleting or mutating portions of the *LHY* sequence and testing for changes in temporal patterns of reporter expression in transgenic seedlings.

Several mutations affecting *LHY* expression cause light-quality dependent phenotypes. For example, plants in which *TOC1* expression is reduced by RNA interference, show arrythmic expression of *cab2:luc* and *ccr2:luc* re-

porters under constant red light and constant darkness but not under constant blue light Más *et al.* (2003a). Also, in *elf3-1* mutant plants, various clock outputs remain rhythmic upon transfer to constant darkness but become arrhythmic under constant light (Hicks *et al.*, 1996). With these observations in mind, reporter expression was examined under red light, blue light and, where appropriate, constant darkness.

3.2.2 The Regulatory Contribution of the *LHY* UTRs

It has previously been shown that expression of *LHY* is subject to translational control in response to light signals (Kim *et al.*, 2003). It is therefore possible that post-transcriptional regulation may contribute in some way to the temporal profile of *LHY* expression.

In many eukaryotic genes, the 5′ and 3′ untranslated regions (UTRs) play an important role in regulating expression of the gene. *Cis*-regulatory elements in these regions may modulate expression at the levels of transcription, transcript stability or translation (Mignone *et al.*, 2002). We therefore tested the effect on expression of altering the 5′ and 3′ UTRs of the reporter constructs.

The LUC3 construct generated by Dr Jae-Yean Kim was used as a reference (pseudo-wildtype) in this set of experiments. The LUC3 construct consists of 1618 base pairs of sequence upstream of the *LHY* translational start codon, extending to the border of the next gene upstream of *LHY*, At1g01070.1, joined as a transcriptional fusion to the enhanced firefly *luciferase* gene luc^+. This sequence was in turn fused to the *LHY* 3′ UTR (Figure 3.1 A).

Under light:dark cycles, the LUC3 construct, showed well defined peaks of expression in transgenic plants at around 2 hours after the onset of the light phase of the entraining cycle (Figure 3.2 A and B). Expression then fell to a

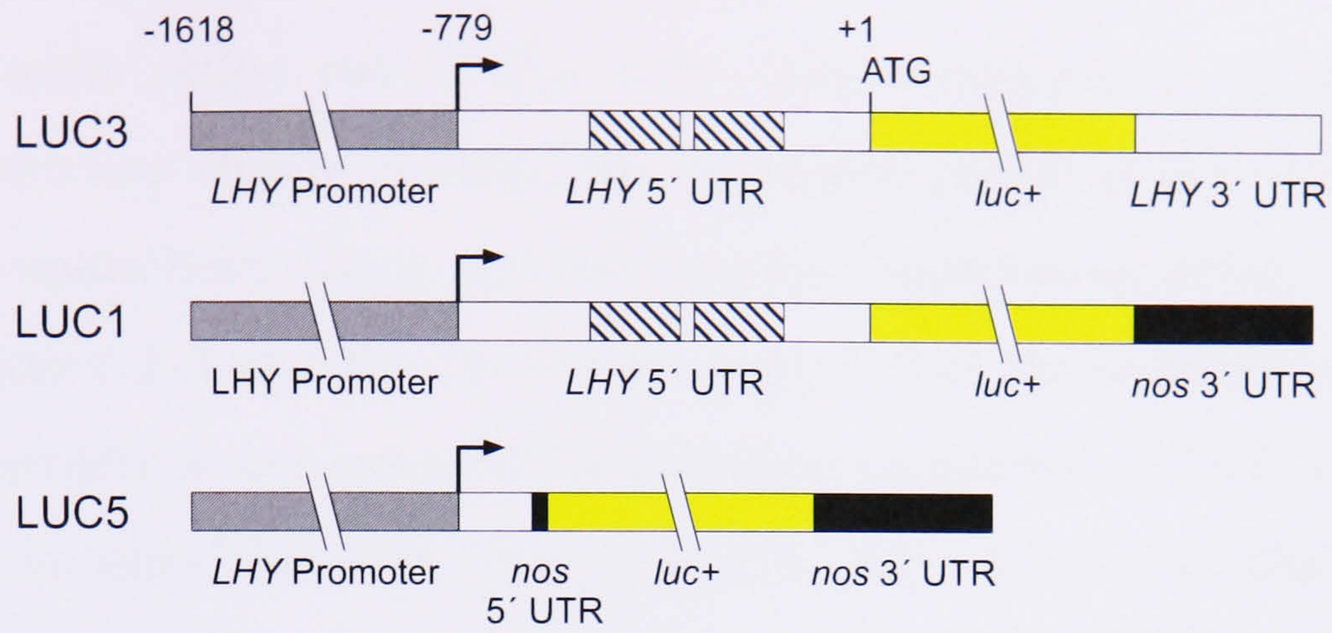


Figure 3.1: Schematic of UTR substitution constructs. Grey regions represent sequence upstream of the predicted transcriptional start site. White regions represent *LHY* untranslated sequence with predicted introns represented as hatched boxes. Black regions show substitution of *LHY* UTR sequence with the corresponding untranslated regions of the *nopaline synthase* gene of *Agrobacterium tumefaciens*. The enhanced firefly *luciferase* gene, luc^+, is shown in yellow. The *LHY* upstream sequence and luc^+ reporter are not shown to scale.

trough during the dark phase of the cycle before rising again in anticipation of the next light phase. This anticipation of the onset of dawn is characteristic of circadian regulation. During free-run, the profile of reporter expression changed somewhat, showing much less sharply defined peaks during the subjective day though remaining clearly rhythmic. These data are consistent with expression pattern of LHY mRNA described by Mizoguchi *et al.* (2002).

To test the regulatory contributions of the *LHY* UTRs, two further constructs were derived from LUC3 (Figure 3.1 A). In the LUC1 construct, the *LHY* 3′ UTR was substituted with the 3′ UTR sequence from the constitutively expressed *nopaline synthase* (*nos*) gene. In the LUC5 construct, the *LHY* 3′ UTR was again replaced with the corresponding *nos* 3′ UTR but, in addition, the majority of the *LHY* 5′ UTR was also replaced with the corresponding *nos* sequence.

Replacement of either *LHY* UTR with the corresponding *nopaline syn-*

thase UTR had no effect on the pattern of either driven or free-running rhythms under either red or blue light since expression of the LUC1 and LUC5 constructs closely matched the expression profile of the LUC3 control construct under both driven and free-running conditions in either red or blue light (Figure 3.2 A and B). These data suggest that the sequence replaced in these constructs is not essential for circadian regulation of *LHY* expression. It should be noted however, that in LUC5, only the 3′-most 638 basepairs of the *LHY* 5′ UTR were replaced leaving some possibility that the remaining 141 basepairs of *LHY* 5′ UTR sequence may mediate regulation of *LHY* expression.

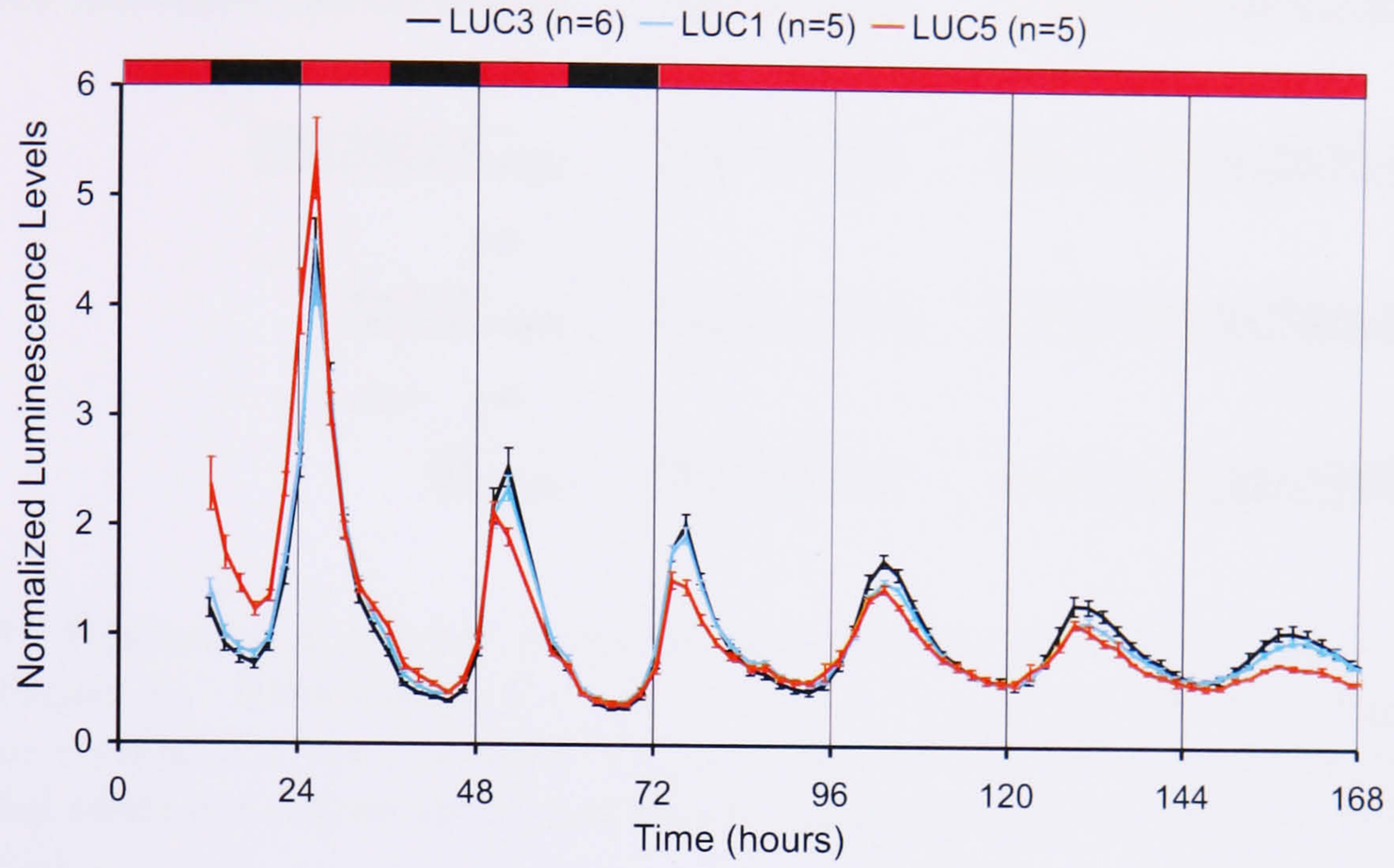
3.2.3 5′ Deletion Analysis of the *LHY* Promoter

To determine how much of the 1618 base-pairs of *LHY* upstream sequence is required for the correct regulation of *LHY* expression, 5′ deletion analysis was carried out in which the sequence upstream of the reporter gene was sequentially truncated from the 5′ end. Again these reporter constructs were constructed and transformed into *Arabidopsis thaliana* by Dr Jae-Yean Kim (Figure 3.3; full sequence in Appendix). These constructs were derived from the LUC1 construct described above, allowing the use of LUC1 as a pseudo-wildtype in these experiments.

Under both driven and free-running conditions, the rhythmic expression patterns of the constructs -1110p*LHY::luc*$^+$ and -957p*LHY::luc*$^+$, containing 1100 and 957 basepairs of *LHY* sequence upstream of the translational start site, respectively, were indistinguishable from the full length LUC1 construct. This observation indicated that no sequence necessary for correct circadian regulation had been removed in these constructs. The same result was obtained under both red and blue light (Figure 3.4 A and B).

A shorter construct, -847p*LHY::luc*$^+$, containing 847 basepairs of *LHY*

A



B

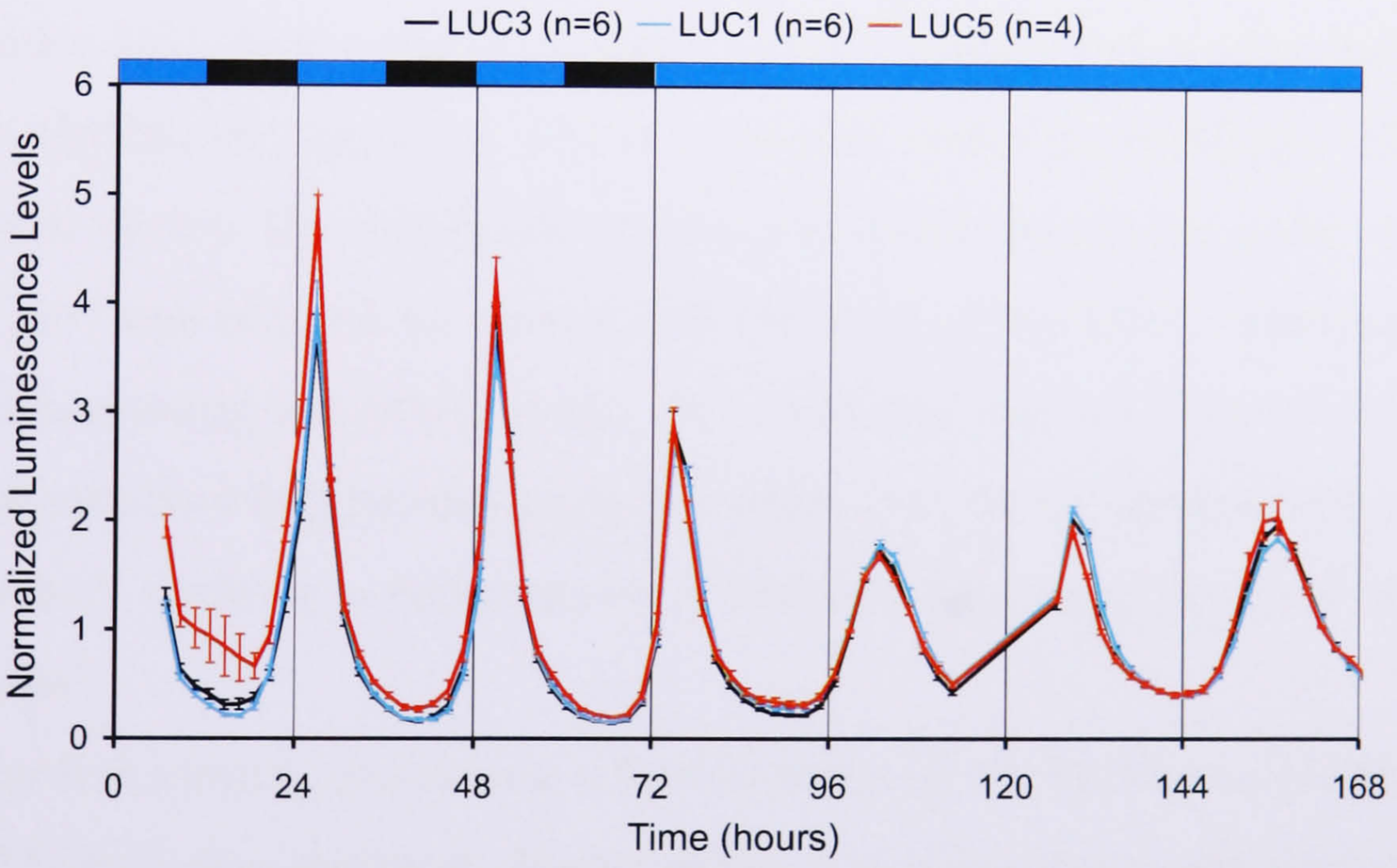


Figure 3.2: Circadian phenotypes of UTR substitution constructs. Between 40 and 60 seeds for each line were sown in collars on MS3 agar and stratified in the dark for 4 days at 4 °C. Plants were grown under 12L 12D cycles for 7 days before transfer to the imaging chamber. Plants were then subjected to a further 3 12L 12D cycles of either red light (A) or blue light (B) before release into constant light conditions. Data are averages of n transgenic lines, normalized to the average level of expression over the course of the experiment. Error bars represent standard errors of the means.

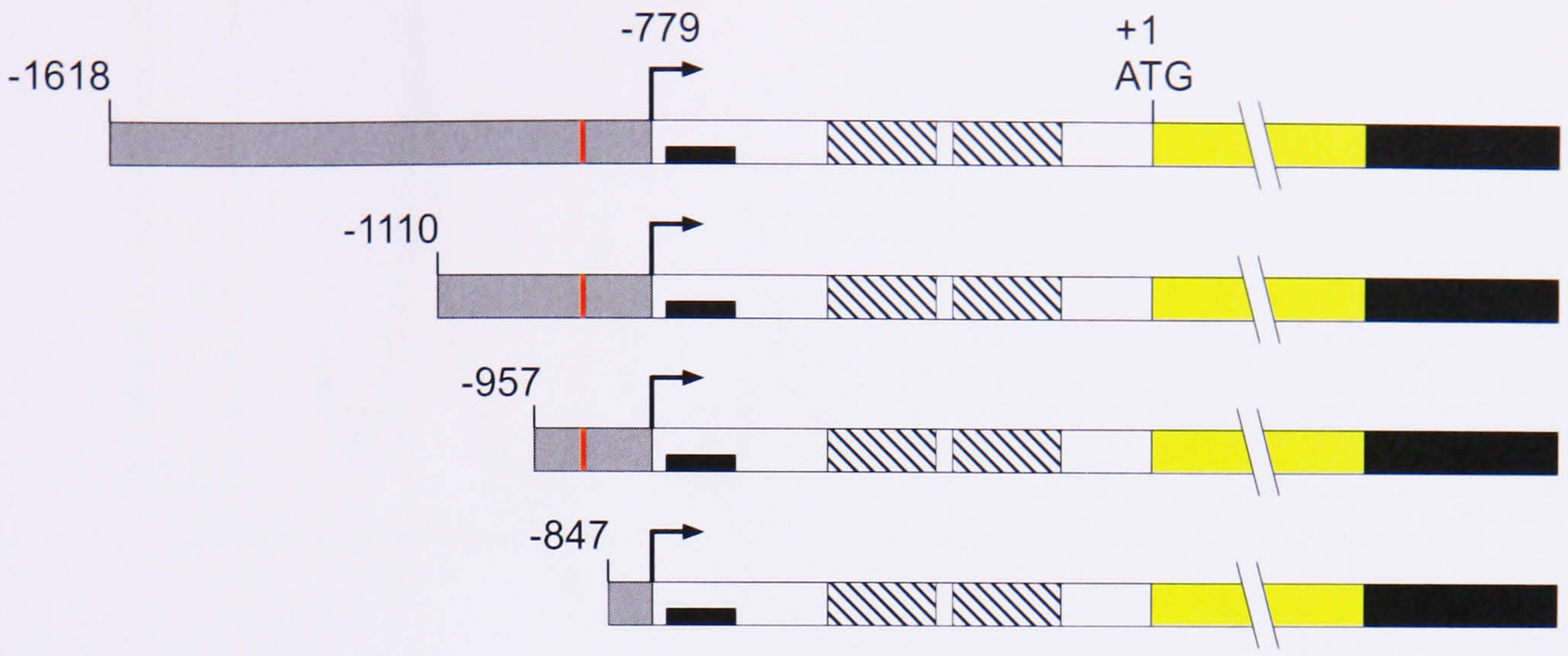
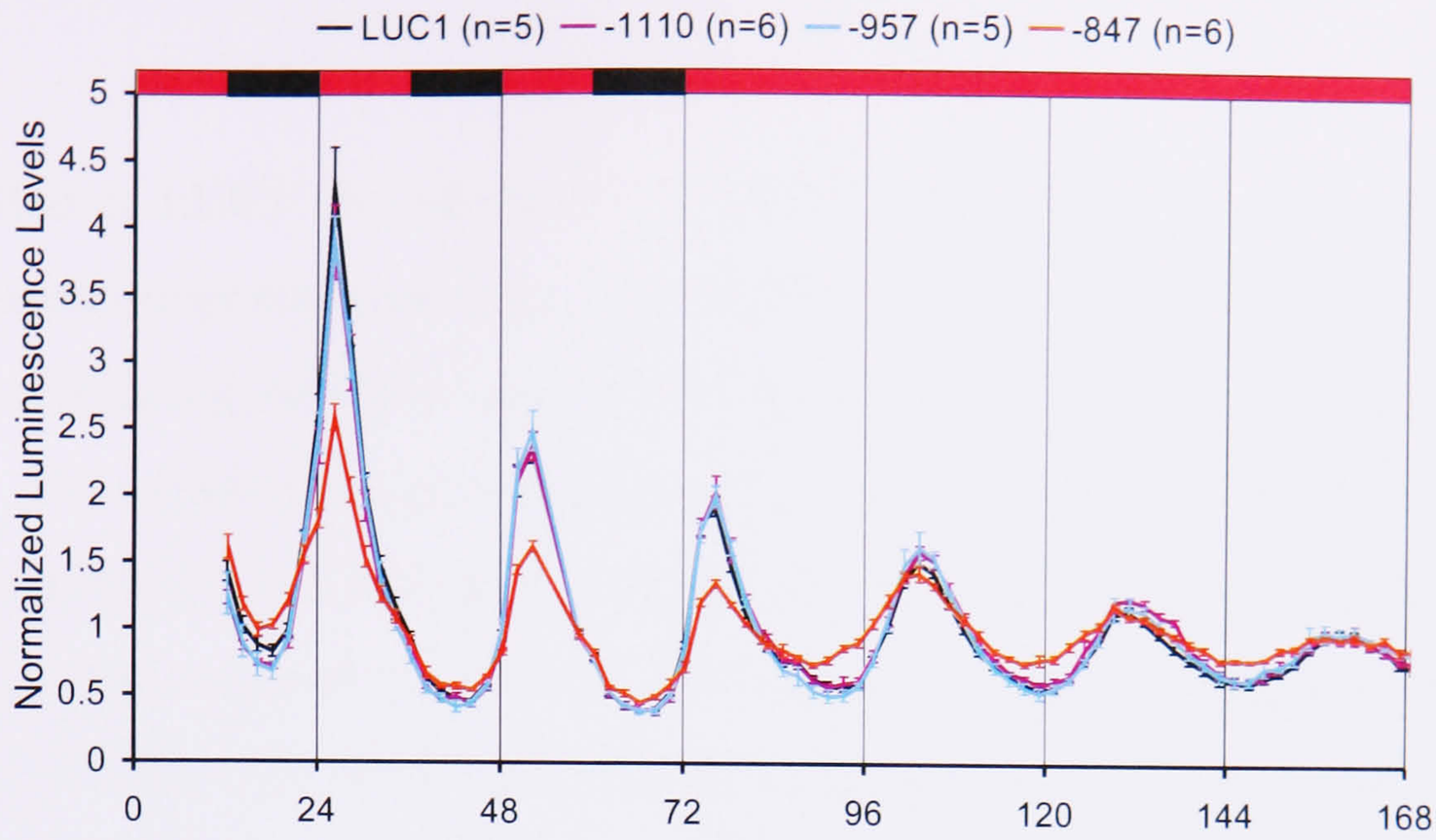


Figure 3.3: Schematic of deletion series constructs. The colour key is the same as used in Figure 3.1. Two notable features of the *LHY* sequence have been highlighted. The vertical red bar represents a G-box motif in the *LHY* promoter and the horizontal black bar represents a repetitive CT-rich region in the *LHY* 5′ UTR.

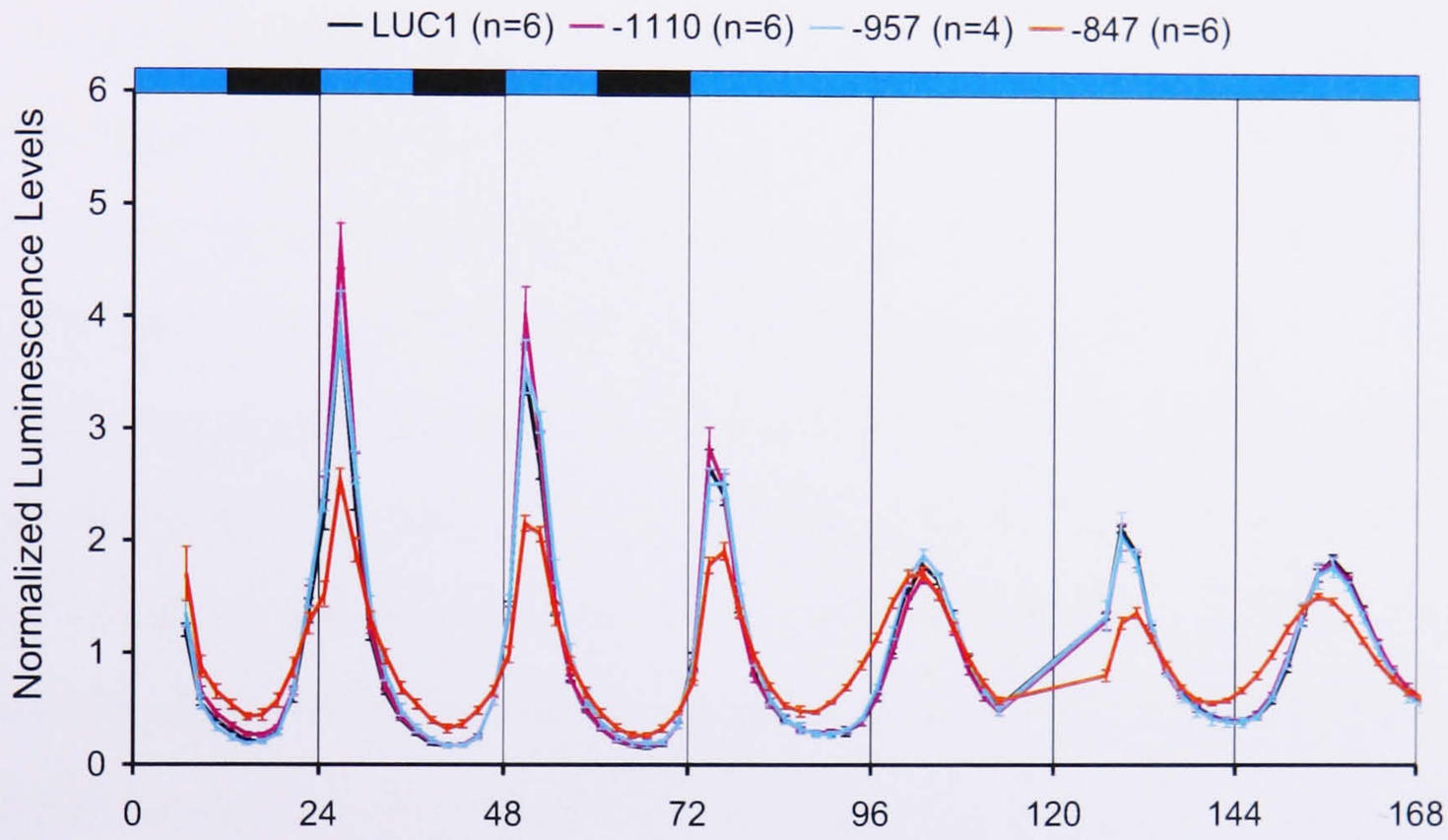
sequence upstream of the translational start site, also displayed rhythmic expression under light-dark cycles and under either constant red or blue light, though the rhythm of expression was of a notably different waveform. Under light:dark cycles, the amplitude of the post-dawn expression peak of -847p*LHY::luc*$^+$ was reduced to around half the level of the LUC1 construct. Since the free-running amplitude of the -847 construct was not reduced compared to the pseudo-wildtype construct, it is likely that the reduced amplitude under light:dark cycles is a consequence of reduced light responsiveness of -847p*LHY::luc*$^+$.

Under free running conditions, a further defect in the expression rhythm of -847p*LHY::luc*$^+$ was observed. Under either constant red or constant blue light, the onset of the peak of transcription of -847p*LHY::luc*$^+$ (where transcription begins to rise from trough levels) occurred around four hours earlier than was observed for the pseudo-wildtype construct. A smaller, but still detectable change in timing was also seen for the peak of expression of the -847p*LHY::luc*$^+$ construct, which occurred around one to two hours earlier

A



B



C

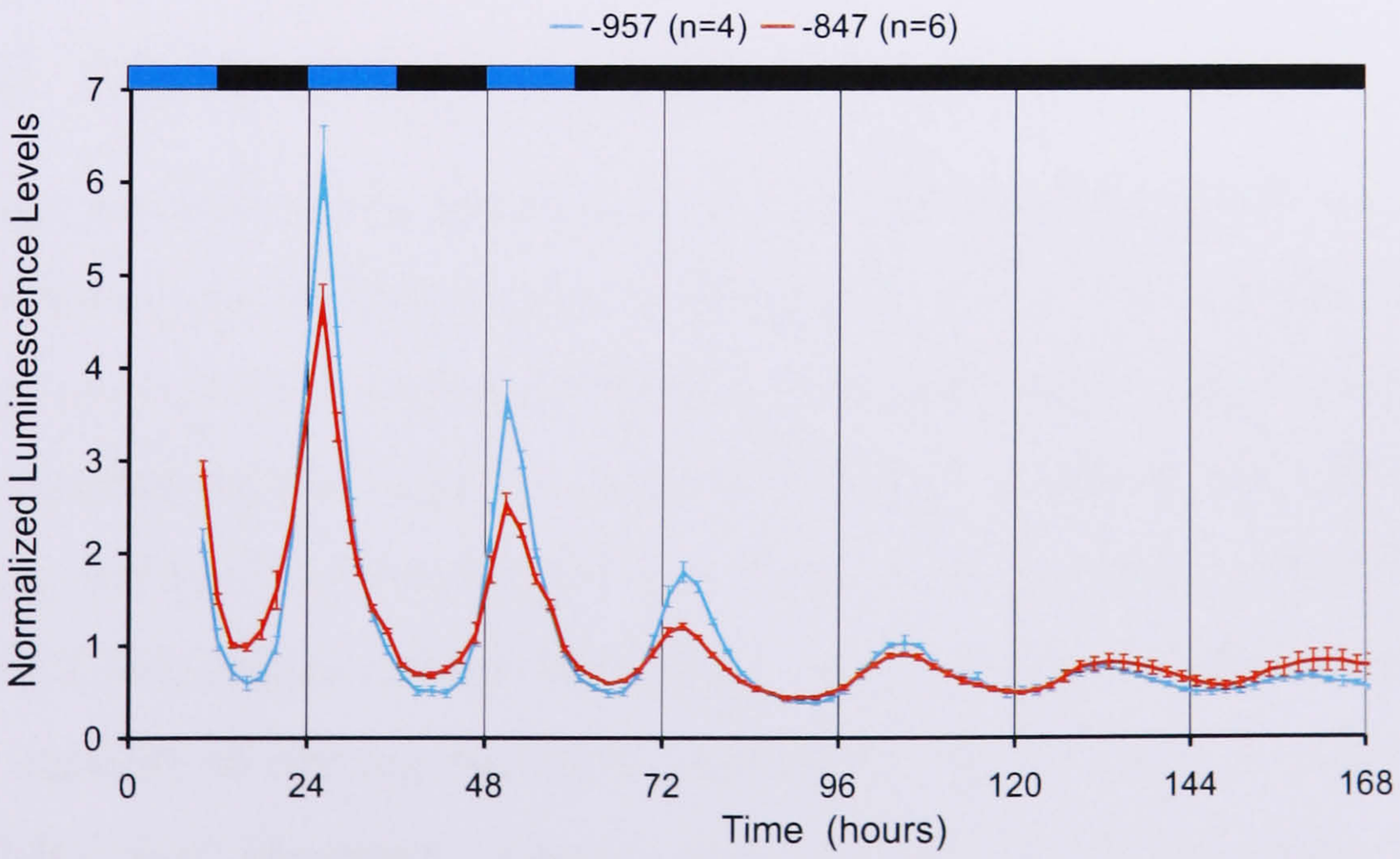


Figure 3.4: Circadian phenotypes of 5′ deletion series. Plants were grown under the same conditions as described for Figure 3.2. On transfer to the imaging chamber, plants were subjected to a further two 12L 12D cycles of either red light (A and C) or blue light (B) before transfer to either A) constant red light B) constant blue light or C) constant darkness. Light conditions are represented by bars above the graphs. Data are averages of *n* transgenic lines normalized to the average level of expression over the course of the experiment. Error bars represent standard errors of the means.

than that of LUC1. So, although -847p*LHY::luc*$^+$ is able to sustain rhythmic expression under constant light. -957p*LHY::luc*$^+$ is the shortest of the deletion series containing sufficient sequence to direct a wild-type expression profile.

The earlier onset of expression of the -847p*LHY::luc*$^+$ construct suggests that a repressive activity acting during the subjective night is unable to act through this truncated construct. Because this early de-repression was not seen with -957p*LHY::luc*$^+$, the putative repressive activity must act between -957 and -847 base-pairs upstream of the translational start codon.

A notably different phenotype of the -847p*LHY::luc*$^+$ construct was observed upon transfer from light-dark cycles to constant darkness (Figure 3.4 C). Under these conditions, strong damping of the rhythms of both the -957p*LHY::luc*$^+$ and -847p*pLHY::luc*$^+$ constructs were observed, though both constructs remained rhythmic for the 4 day duration of the experiment in constant darkness. Interestingly, the -847p*LHY::luc*$^+$ construct did not exhibit an early onset of expression relative to -957*LHY::luc*$^+$ in constant darkness. Instead, the peak of expression from the -847p*LHY::luc*$^+$ construct occurred marginally later than that of -957p*LHY::luc*$^+$.

3.2.4 Linker Scan Analysis of -847p*LHY::luc*$^+$

Previous attempts have been made to identify the *cis*-element responsible for mediating the rhythmic input to -847p*LHY::luc*$^+$ (Taylor, 2004). A set of reporter constructs in which sequential 10 base-pair regions of -847p*LHY::luc*$^+$ were mutated to the sequence AAGGGCCCAA, scanning the region from position -847 to -747, covering the span of the promoter region to the 5′ border of the CT-rich region. Under white light, each of these constructs was shown to be capable of driving rhythmic expression with the same profile as the -847p*LHY::luc*$^+$ construct. These constructs were reanalysed under red and blue light, giving similar results (data not shown).

The failure of this analysis to identify a locus required for rhythmicity of -847p*LHY::luc*$^+$ presents two possible explanations. Firstly, it is possible that the necessary *cis*-element for rhythmic expression is present in multiple copies, even within this short promoter region, and that no single mutation in Taylor's series rendered all copies of this element non-functional. Visual inspection of this region revealed 3 copies of the sequence AAAAATT and two copies of the sequence GCTTCT. For both of these repeated motifs, none of Taylor's mutations would have removed all copies. The second possibility is that such an element is present downstream of the most 3′ mutation, within the CT-rich region of the 5′ UTR. This possibility remains untested as this region was not altered in the LUC5 construct tested above.

The suggestion that rhythmic expression of the -847p*LHY::luc*$^+$ construct might be conferred by the CT-rich region of the 5′UTR would again raise the possibility that the rhythmicity of -847p*LHY::luc*$^+$ might be a product of translational control. If this were the case, one would expect the -847p*LHY::luc*$^+$ construct to drive constitutive expression of *luc*$^+$ mRNA. To test this possibility, northern analysis of luc$^+$ mRNA was carried out. As shown in Figure 3.5, -847p*LHY::luc*$^+$ is still expressed rhythmically at the level of mRNA and expressed with a very similar pattern to endogenous *LHY* mRNA, demonstrating that the rhythmic expression of -847p*LHY::luc*$^+$ is mediated at the level of mRNA abundance.

3.2.5 Negative Feedback Regulation

Since -847p*LHY::luc*$^+$ is able to respond to a circadian input, driving rhythmic expression of the reporter, we asked whether this circadian signal arose by a negative feedback mechanism between *LHY* and other oscillator components. To determine whether this was so, the *LHY*-overexpressing TN104 transgenic line was transformed with the -847p*LHY::luc*$^+$ construct. In TN104, vari-

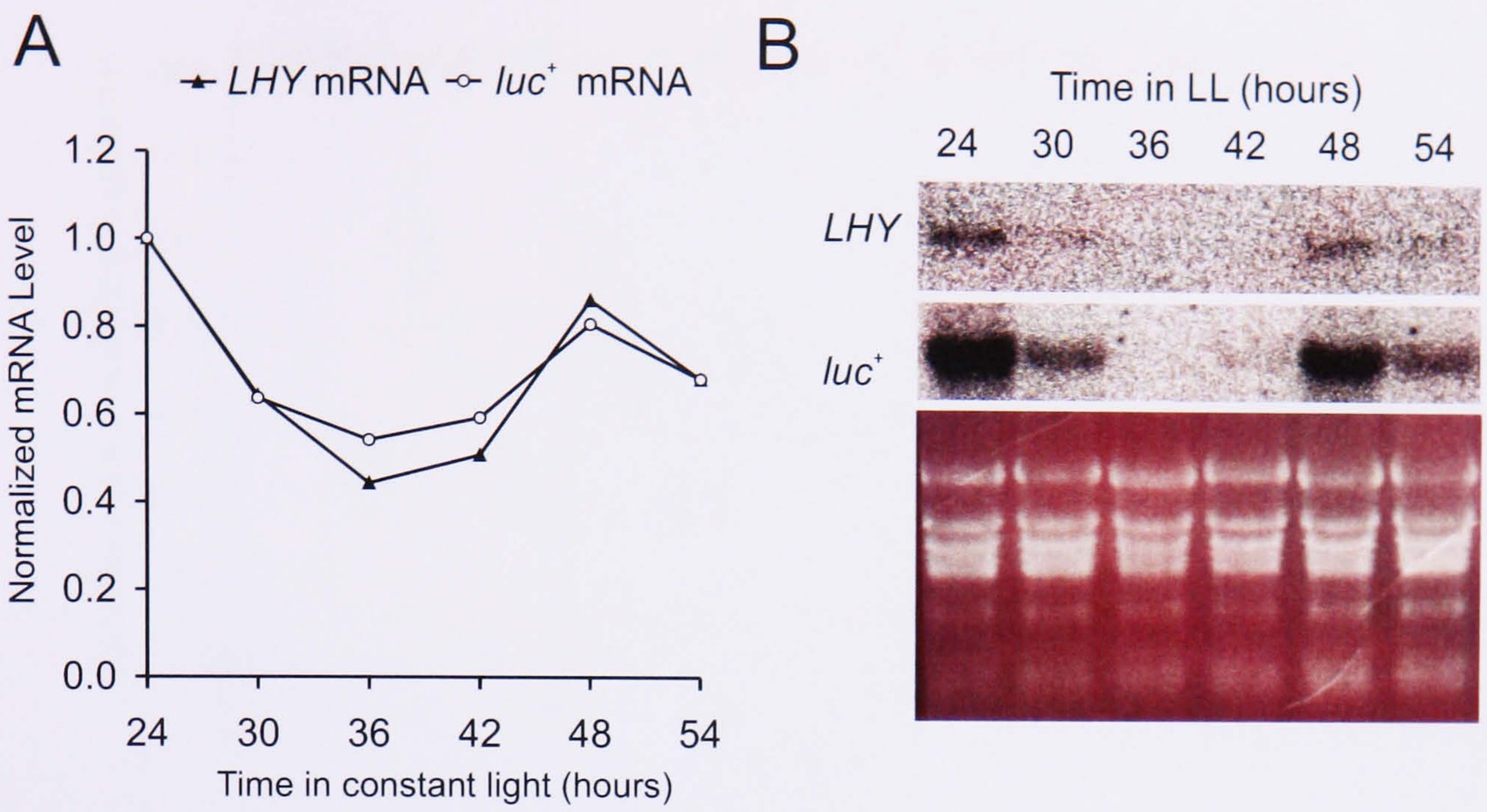


Figure 3.5: Northern analysis of endogenous *LHY* mRNA and *luc*$^+$ mRNA from the -847p*LHY::luc*$^+$ construct under constant light . A) mRNA levels quantified by densitometry. Data have been normalized to the levels at 24 hours in constant light. B) Images of a representative blot showing hybridization of *LHY* and *luc*$^+$ probes. Ethidium bromide stained total RNA is shown in the lower panel as a loading control. Plants of a representative -847p*LHY::luc*$^+$ line were grown for 7 days under 12:12 white light:dark cycles before transfer to constant white light (90μmol/m^2/s). Tissue was harvested every 6 hours and frozen in liquid nitrogen until needed.

ous circadian rhythms are disrupted and the endogenous *LHY* transcipy is expressed arrhythmically at depressed levels (Schaffer *et al.*, 1998). If the -847p*LHY::luc*$^+$ construct is able to mediate negative feedback, one would expect the expression level of this construct to be depressed in TN104 plants.

Figure 3.6 shows that, in the TN104 genetic background, expression of -847p*LHY::luc*$^+$ is arrhythmic and expressed at a level lower than trough levels of expression in the wildtype background, indicating that the rhythmic signal to this construct arises by a negative feedback mechanism employing LHY.

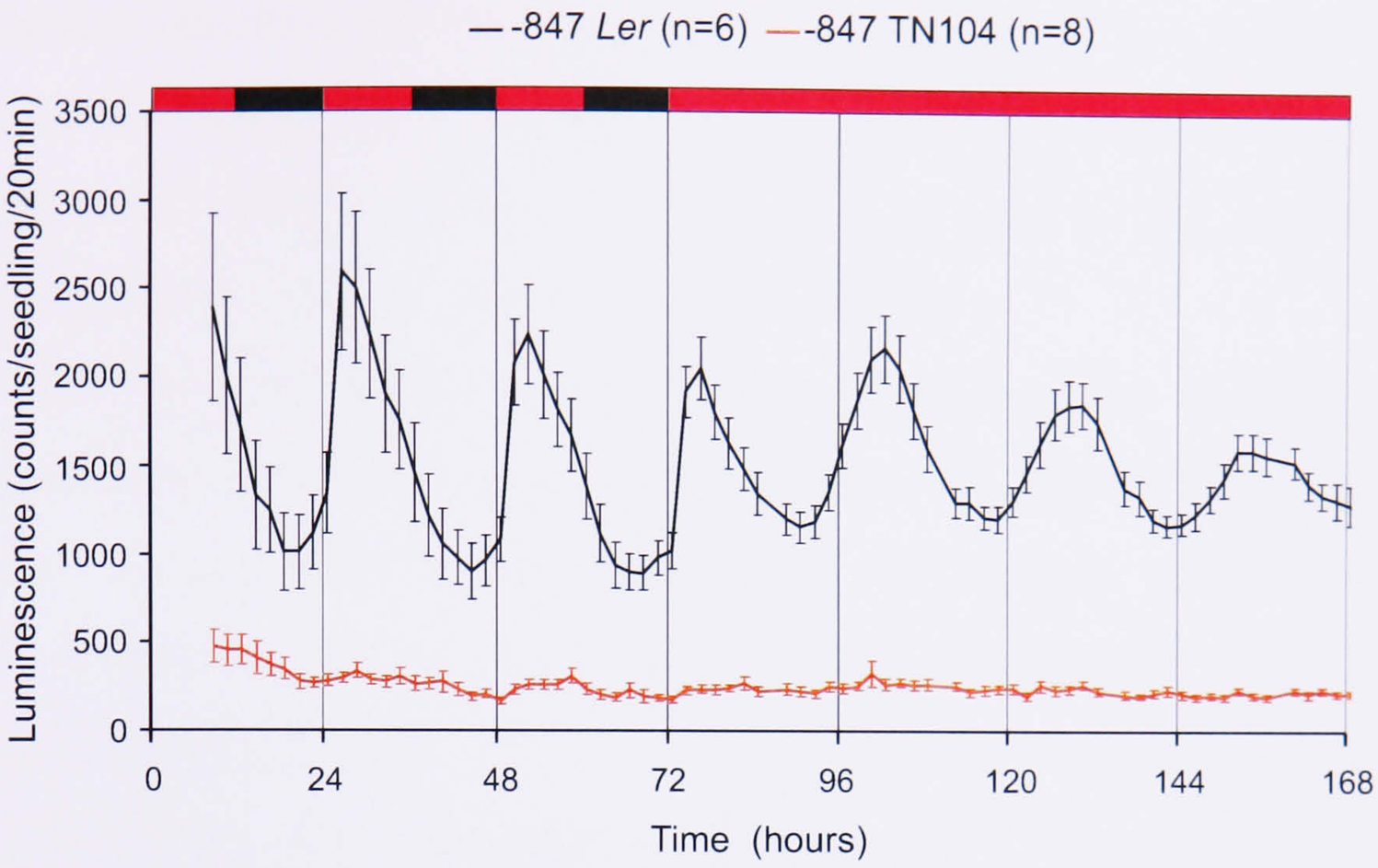


Figure 3.6: The -847 promoter is able to mediate negative feedback. Plants were grown under the same conditions as described for Figure 3.2 before transfer to the imaging chamber. Plants were then subjected to a further 3 light dark cycles of red light before release into constant light conditions. Data are averages of n transgenic lines and error bars represent standard errors of the means.

3.3 Further Investigation of Circadian Regulation of *LHY*

The differing phase relationship between -957p*LHY::luc*$^+$ and -847p*LHY::luc*$^+$ under constant light compared to constant darkness (Figure 3.4) suggests that whatever signal acts between positions -957 and -847 to determine the correct phase of expression of the -957p*LHY::luc*$^+$ construct cannot simply be a static repressive signal that delays the onset of expression. This raises the possibility that this phase modifying signal might be expressed rhythmically, constituting a second rhythmic signal to the *LHY* promoter. For clarity, the rhythmic signal that drives the rhythmic expression of the -847p*LHY::luc*$^+$ construct will hereafter be referred to as R_A, while the hypothetical rhythmic signal acting between positions -957 and -847 will be referred to as R_B.

The notion of two rhythmic inputs coupled at the *LHY* promoter is

an attractive one since the longstanding single loop model of the *Arabidopsis* oscillator, in which *CCA1* and *LHY* form a negative feedback loop with *TOC1* (Alabadí *et al.*, 2001), is inconsistent both with more recent experimental evidence and with certain theoretical considerations.

From the experimental perspective, no allele of *TOC1* has as yet been identified which completely abolishes circadian rhythmicity. instead. all identified *toc1* mutant alleles display short period phenotypes. While severe reduction of *TOC1* expression by RNA interference (RNAi) has been reported to cause arrhythmia under conditions of constant darkness and constant red light, *TOC1* RNAi plants nevertheless displayed free-running rhythmicity under constant blue light (Más *et al.*, 2003a).

Two theoretical objections can also be raised to the single feedback loop model. Firstly, mathematical simulation of this model shows that it is unable to account for the phenotypes of certain mutations to clock components (Locke *et al.*, 2005a). For example, this single feedback loop model failed to simulate the short period, rapidly damping phenotype of *lhy;cca1* mutants. predicting instead a long period phenotype. The second failing of any single loop oscillator model is the inability to track both dawn and dusk phases of the entraining photocycle (Pittendrigh & Daan. 1997).

Pittendrigh and Daan (1976) suggested that this photoperiodic entrainment of the clock could be accounted for by a model of two mutually. but imperfectly, coupled oscillators. In this model, a morning oscillator (M-oscillator) is entrained to the timing of dawn, whereas an evening oscillator (E-oscillator) is entrained to the timing of dusk with these oscillators respectively determining the time of morning-phased and evening phased outputs.

In *Arabidopsis*, several alternative oscillator models have been proposed in which multiple feedback loops and multiple light inputs to the oscillator would account for the observed ability of the oscillator to track both dawn

and dusk phases of the entraining cycle (Locke *et al.*, 2005b, 2006; Zeilinger *et al.*, 2006), though explicit evidence confirming the mechanisms employed in these models is still lacking. The most experimentally characterized of these models contains two loops converging on *LHY*, the *LHY/CCA1-TOC1* loop and the *LHY/CCA1-PRR7/PRR9* loop (Farré *et al.*, 2005), supporting our hypothesis that at least two rhythmic inputs converge on *LHY*.

3.3.1 Experimental Design

We reasoned that, if two rhythmic signals are indeed coupled at *LHY*, then these might arise separately from "morning" and "evening" oscillator mechanisms, as proposed by Pittendrigh and Daan (1976). If this were the case, one would expect these signals, R_A and R_B, to display different phase relationships to the timings of the dawn and dusk cues of the entraining photocycle. Any resultant change in the phase relationship of R_A and R_B to one another would then be manifested in the phases of expression of the -847p*LHY::luc*$^+$ construct, which is determined by R_A, and the -957p*LHY::luc*$^+$ construct, which is proposed to be determined by both R_A and R_B.

To test this hypothesis, we first examined the phase relationship of the -957p*LHY::luc*$^+$ and -847p*LHY::luc*$^+$ constructs to changes in the relative timing of dusk cues. In order to do this, we entrained plants to photocycles of 8L 16D, 12L 12D and 16L 8D before release into constant light. Under these conditions, the last dusk cue available to the clock occurs respectively 16, 12 and 8 hours prior to the onset of constant conditions. Since the phase of *LHY* expression under light:dark cycles is strongly influenced by acute induction at dawn by light, we examined the phase of the first free-running rhythm.

3.3.2 Regulation of the Phase of *LHY*

After entrainment to 8L 16D cycles the peak of expression of the -847p*LHY::luc*$^+$ construct in constant light was advanced by 3 to 4 hours relative to that of the -957p*LHY::luc*$^+$ construct (Figure 3.7 A). This is in contrast to the smaller advance of 1 to 2 hours observed after entrainment to 12L 12D photocycles (Figure 3.7 B). After entrainment to 16L 8D, the phase difference between the -957p*LHY::luc*$^+$ and -847p*LHY::luc*$^+$ constructs diminished to such an extent that no difference was observed in either the phase of onset or peak expression of these constructs (Figure 3.7 C).

A similar set of experiments was performed to examine the phase relationships of the -957p*LHY::luc*$^+$ and -847p*LHY::luc*$^+$ constructs to the timing of dawn cues. In these experiments, following entrainment to 8L 16D, 12L 12 D and 16L 8D photocycles, plants were released into constant darkness.

After entrainment to 8L 16D cycles, on release to constant darkness the -847p*LHY::luc*$^+$ construct displayed a delayed peak phase relative to the -957p*LHY::luc*$^+$ construct, occurring around 2 hours later that that of the -957p*LHY::luc*$^+$ construct (Figure 3.8 A). This is in contrast to the effect described above on release to constant light. A similar phase relationship was observed in constant darkness after entrainment to 12L 12D photocycles, where the peak of expression of the -847p*LHY::luc*$^+$ construct occurred around 2 hours later than that of the -957p*LHY::luc*$^+$ construct (Figure 3.8 B). No difference in timing between the two constructs was observed in constant darkness after entrainment to 16L 8D photocycles (Figure 3.8 C).

From these data, it is clear that the phase relationship between the -957p*LHY::luc*$^+$ and -847p*LHY::luc*$^+$ constructs is dependant on the photoperiod of the entraining cycle. For this shift in phase relationship to occur, the phase modifying signal that acts between positions -957 and -847 must respond to photoperiodic information in the entraining cycle. In other words,

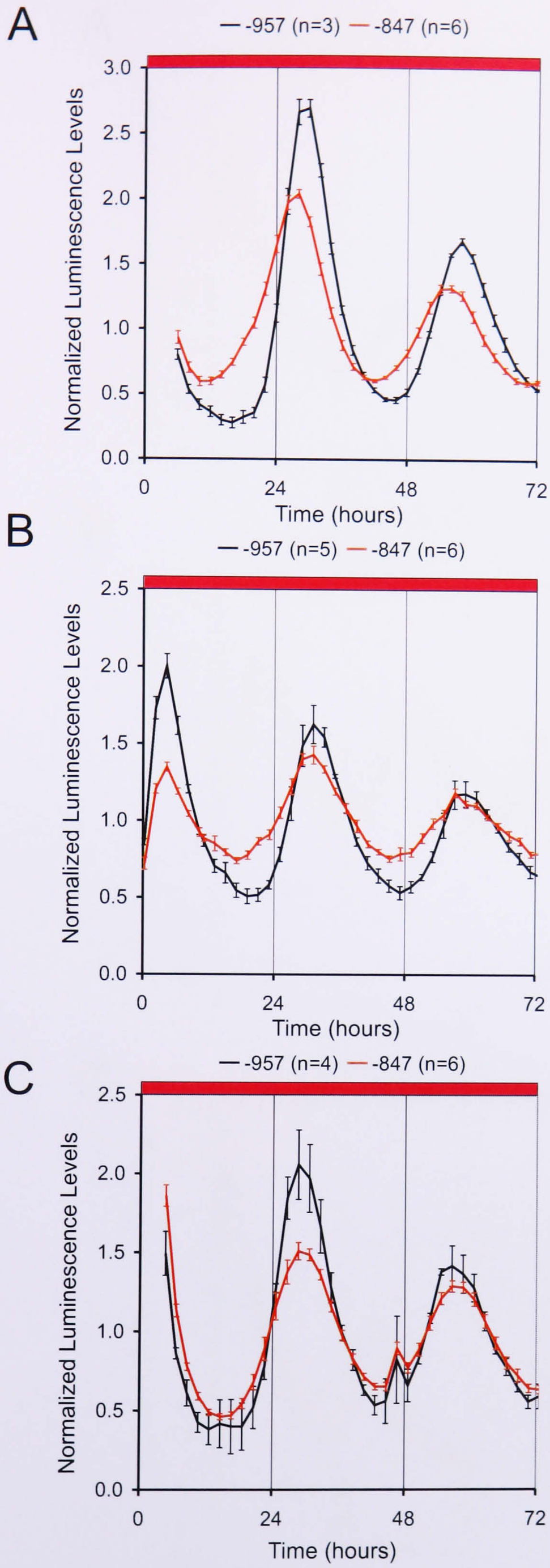


Figure 3.7: Regulation of *LHY* by dusk cues. Plants were grown under cycles of either A) 8L 16D B) 12L 12D or C) 16L 8D for 7 days before transfer to the imaging chamber. Plants were then subjected to a further 3 photocycles using red light before release into constant red light at time 0. Data are averages of *n* transgenic lines and error bars represent standard errors of the means.

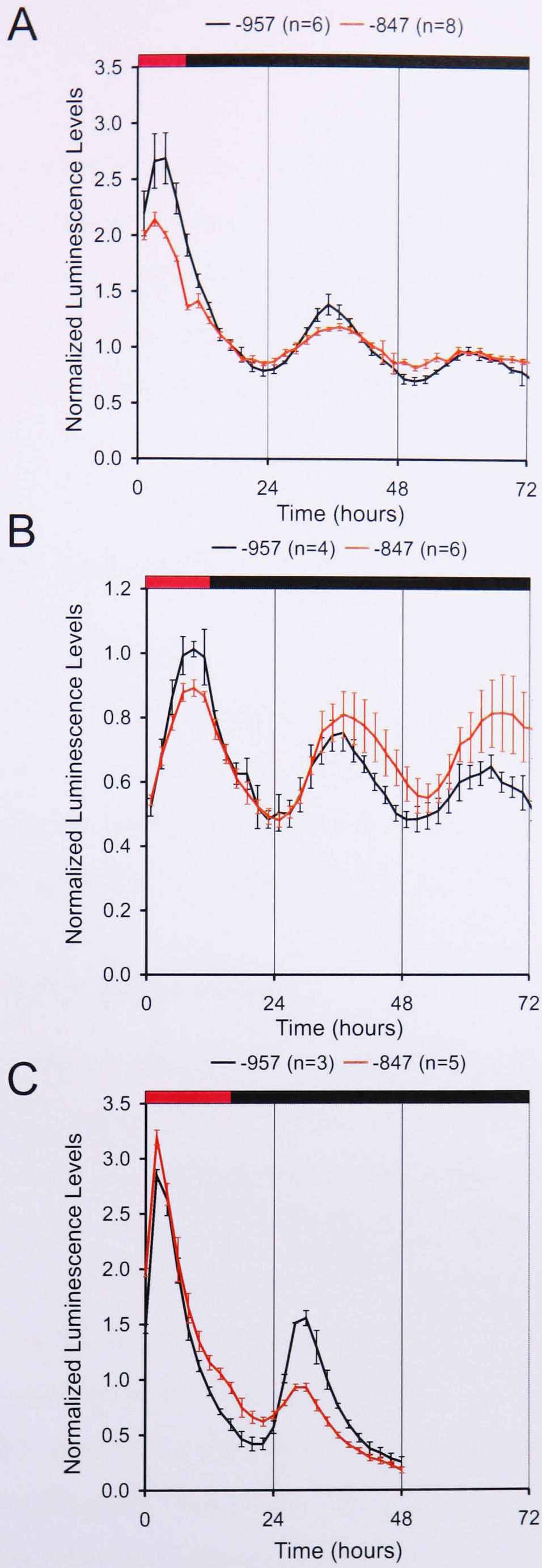


Figure 3.8: Regulation of *LHY* by dusk cues. Plants were grown under cycles of either A) 8L 16D B) 12L 12D or C) 16L 8D for 7 days before transfer to the imaging chamber. Plants were then subjected to a further 3 photocyles using red light before release into constant darkness at time 8, 12 and 16 repsectively. Data are averages of *n* transgenic lines and error bars represent standard errors of the means..

this signal is entrainable. Interestingly, the various phase relationships persist in constant conditions for at least one free-running cycle. Taken together, these properties are strongly suggestive of the phase modifying activity between -957 and -847 being a circadian signal. These data are, therefore, consistent with the hypothesis that at least two rhtyhmic inputs, R_A and R_B, converge on the *LHY* promoter.

3.4 Light-induction phenotypes

Above, it was demonstrated that the -847p*LHY::luc*$^+$ construct, while still displaying a degree of light induction after dawn under light:dark cycles, was less responsive to diurnal light signals than the LUC1 control construct. To further investigate the induction of *LHY* expression by light, the deletion series constructs were examined in a light induction assay in etiolated seedlings.

3.4.1 Experimental design

In light grown seedlings, photoreceptor levels exhibit a strong circadian rhythm of expression (Toth *et al.*, 2001), which can complicate studies of light signalling. In etiolated seedlings, however, photoreceptor proteins accumulate to much higher levels, minimising the effects of circadian regulation on light signalling. This accumulation of photoreceptors has the additional effect of increasing the sensitivity of etiolated seedlings to light signals, producing a more sensitive assay for acute induction of gene expression by light.

Both *LHY* and *CCA1* are known be acutely induced by red light in etiolated seedlings (Wang & Tobin, 1998; Viczián *et al.*, 2005) but the responses of these genes to blue light have not been characterized elsewhere. As described in Chapter 1, in *Arabidopsis*, detection of red and blue light signals is performed by distinct families of photoreceptors, the phytochromes for red and

far-red light and the cryptochromes and phototropins for blue light. It is not known whether these photoreceptors transduce signals to the nucleus through common or distinct pathways. Therefore, induction of expression from the *LHY* promoter by both red and blue light was investigated.

To assay light induction of the deletion series reporter constructs, luciferase activity in 7 day old etiolated seedlings was monitored for 2 to 3 days to establish a baseline before exposure to 10μmol/m^2/s of red or blue light for 1 hour. Post-pulse expression of the reporter constructs was then monitored for a further 2 days.

3.4.2 Induction of expression by red light

The baseline expression of the control construct, LUC1 was rhythmic at a very low amplitude (Figure 3.9 A and B). In some experiments, these rhythms were synchronized whereas in others, each seedling expressed the reporter rhythmically but with different phases such that no discernible rhythm can be picked out from averaged data. In the latter case, we expect that asynchronous seed germination was responsible for the differing circadian phase of individual seedlings.

Upon exposure to the red light pulse, the LUC1 construct displayed a tripartite response (Figure 3.9 A). The first stage (Stage I) corresponds to the acute induction of expression in response to light. Under the experimental conditions employed here, LUC1 showed 1.5-fold induction over the pre-pulse baseline. During Stage II, expression fell below the baseline before rising again for Stage III, which corresponds to the first synchronized circadian peak after the light pulse. The form of this response is consistent with that seen for *LHY* mRNA levels (Viczián *et al.*, 2005).

As with circadian regulation, under light induction the -957p*LHY*::*luc*$^+$ construct behaved in the same way as the full length construct, demonstrating

that this construct contains sufficient sequence for correct induction by red light.

Interestingly, despite the reduced light induction of the -847p*LHY::luc*$^+$ construct under light:dark cycles, this construct consistently displayed the same 1.5-fold acute induction observed for the longer constructs in this assay. However, during stage II of the response expression of this construct failed to drop below the level of the baseline. As suggested on the basis of the circadian expression profile of this construct, this result again hints that a repressive activity directed to the *LHY* promoter is unable to act through a construct lacking the -957 to -847 region. During Stage III, the amplitude of the first post-pulse circadian rhythm was reduced for the -847p*LHY::luc*$^+$ construct, relative to the LUC1 and -957p*LHY::luc*$^+$ constructs.

3.4.3 Induction of expression by blue light

A similar expression profile was seen following induction by blue light (Figure 3.9 C and D). Again, the three stages of the response to the one hour light pulse are clearly visible. As for induction by red light, -957p*LHY::luc*$^+$ contains sufficient sequence to direct the same response to the blue light pulse as the full length construct.

The magnitude of acute induction by blue-light was similar to that produced by red light at around 1.5-fold increase. The same fall in expression below pre-induction levels was again seen in stage II of the response for LUC1 and -957p*LHY::luc*$^+$. During Stages II and III, the -847p*LHY::luc*$^+$ construct showed the same altered expression profile as after induction by red light, failing to drop below the pre-induction levels of expression during Stage II and displaying a reduced amplitude circadian peak during Stage III.

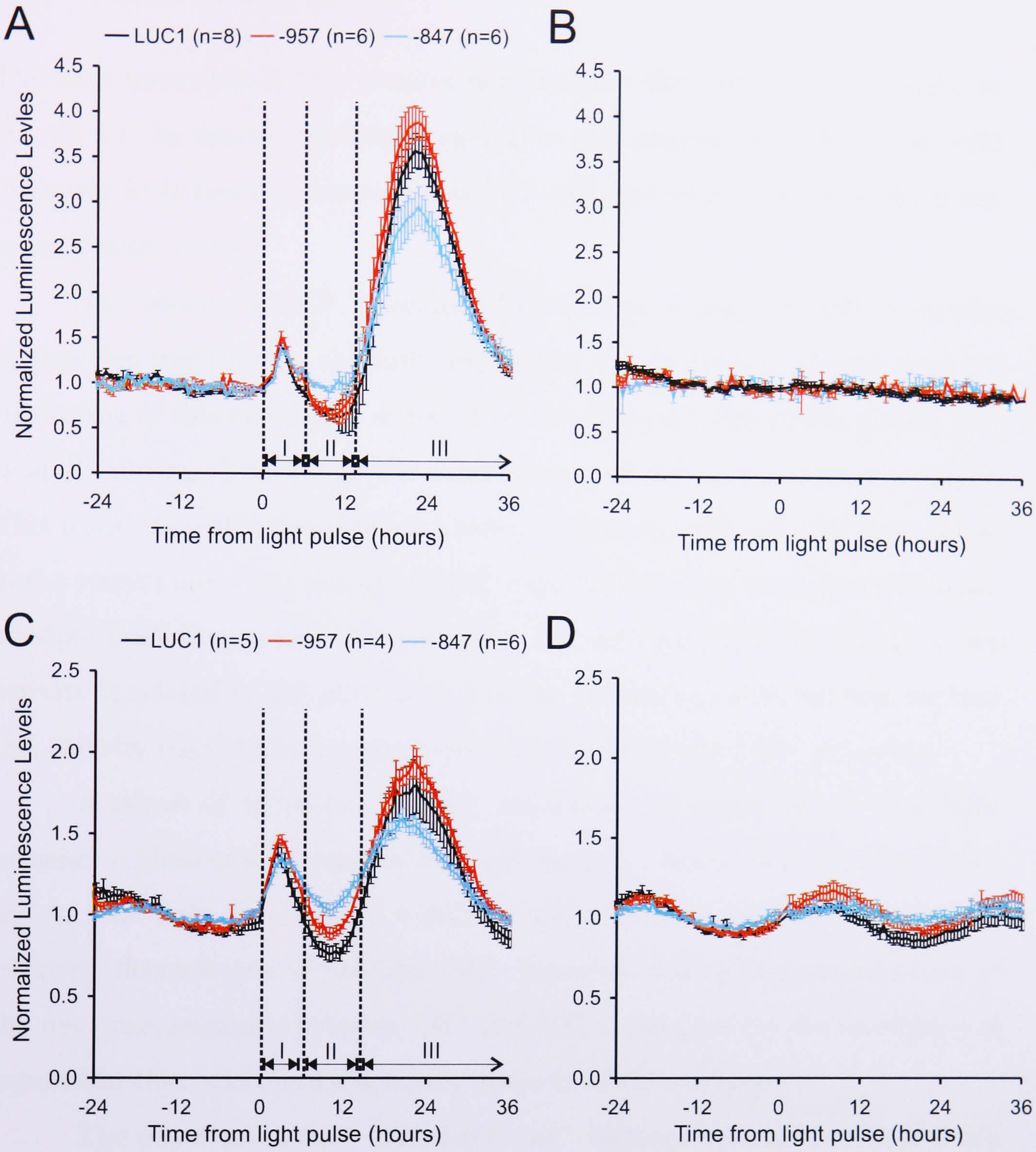


Figure 3.9: Light induction phenotypes of deletion series constructs. For each line, 30 to 60 seeds were sown in collars on MS3 agar. Seeds were stratified for 7 days in the dark at 4 °C before exposure to cool white light ($90\mu mol/m^2/s$) for 2 hours to encourage germination. Plants were grown for 7 days in darkness before transfer to the imaging chamber. After 36 hours, plants were given a 1 hour flash of either red light (A) or blue light (C) at $10\mu mol/m^2/s$. Dark controls for each experiment are shown in B and D. Data are averages of several independent lines normalized to average expression levels over the 24 hours preceding the light pulse. Error bars represent standard errors of the means.

3.5 Discussion

The data presented in this chapter demonstrate that sufficient regulatory elements for the correct circadian regulation and response to light of the *LHY* promoter lie between nucleotide positions -957 and -638, relative to the translational start codon.

The region of *LHY* promoter downstream of position -847 is capable of directing free-running rhythmic expression of a *luciferase* reporter gene but the timing of this expression differs from the wild-type expression profile, particularly during the onset of the expression peak during the subjective night. This result indicates that further elements between -957 and -847 have a role in the correct circadian timing of *LHY* expression during the subjective night, possibly mediating a repressive activity. On the basis that the timing of this activity is related to the photoperiod of the entraining cycle, we propose that this activity constitutes a second rhythmic input to the *LHY* promoter.

Analysis of induction of *LHY* expression by either red or blue light revealed a three-stage response to light signals. Acute induction by either red or blue light in etiolated seedlings appears to be entirely mediated by sequence downstream of position -847. However, during the second phase of the response, sequence between -957 and -847 is required for the repression of expression that was observed during stage II of the response.

The observation that -847p*LHY::luc*$^{+}$ displays the same level of acute induction by light as the longer constructs in etiolated seedlings is an interesting one given that this construct lacks the G-box motif present in the longer constructs. As described in chapter one, the current model of red light signalling describes the interaction of activated PHYB protein with the DNA-binding factor PIF3, which binds to G-box elements in the promoters of target genes. Given the absence of a G-box motif in -847p*LHY::luc*$^{+}$, this model fails to account for the light induction of this construct.

Using a *PIF3* antisense line, Martínez-García *et al.* (2000) showed that PIF3 was necessary for light induction of LHY expression in etiolated seedlings. That paper was later contradicted by work by Viczián *et al.* (2005), who showed that light induction of LHY was unaffected in the PIF3 loss-of function mutant *poc1*. These disparate observations could, perhaps, be reconciled with the hypothesis hat one of more of the closely related bHLH factors acts redundantly with PIF3 and that the *PIF3* antisense line employed by Martínez-García *et al.* affected multiple redundant factors. The data presented here show clearly that, under the conditions employed, the G-box element, and presumably any bHLH factors that bind there, are unnecessary for the acute induction of *LHY* to wild-type levels by either red or blue light in etiolated seedlings. This is strong evidence that the model proposed by Martínez-García *et al.* is insufficient to describe the light-responsive expression of *LHY*.

A further complication in understanding the regulation of *LHY* by light signals arises from the disparity between light induction in response to dawn during light:dark cycles, where the level of -847p*LHY::luc*$^+$ is induced to around half the level of the LUC1 control construct, and acute light induction in etiolated seedlings, where the -847p*LHY::luc*$^+$ construct displayed the same degree of induction as LUC1. These data can be reconciled by two different hypotheses. Firstly, light signalling networks are known to significantly differ between different developmental stages to function in different ways in etiolated seedlings and light grown seedlings (Ma *et al.*, 2001). However, it is also possible that the circadian regulation of the *LHY* promoter influences its ability to respond to light signals. In this hypothesis, the phase advance of the -847p*LHY::luc*$^+$ construct relative to LUC1 would alter the relationship between optimally light sensitive state of the promoter and the timing of dawn, resulting in reduced levels of induction.

Chapter 4

Identification of Candidate Transcription Factor Binding Sites in the *LHY* Upstream Regulatory Region

4.1 Introduction

The results described in Chapter 3 demonstrate that sufficient regulatory elements to define circadian and light responsive transcription of *LHY* lie within a region within 957 basepairs upstream of the translational start site of *LHY* and that at least two rhythmic inputs are integrated by the *LHY* promoter. This chapter describes approaches to localize the transcription factor binding sites within this region that might be responsible for these transcriptional responses.

4.2 A Bioinformatic Approach to Identify *LHY cis*-Elements

Several databases of transcription factor binding sites are available on the worldwide web, which can be used to scan a sequence for candidate transcription factor binding sites, for example, the PLACE and PlantCARE databases, which were designed specifically for plant promoters (Lescot *et al.*, 2001; Higo *et al.*, 1999). This approach quickly proved inappropriate for identifying transcription factor binding sites on the *LHY* promoter. Two properties of many described binding sites, short length and variability at some nucleotide positions, led to the identification of an unmanageable number of binding site candidates in sequences of the length considered here. For example, between positions -957 and -847, the PLACE and PlantCARE databases together returned over 30 potential transcription factor binding sites (Table 4.1).

A second approach to identify regulatory elements in promoter sequences is "phylogenetic footprinting". This approach applies the assumption that, among orthologues, regulatory elements will be better conserved than the surrounding sequence. Applying this approach to the *LHY* promoter, two problems quickly became evident. Firstly, while a number of plant genome projects have generated large amount of genomic genomic sequence data, as of January 2007, genomic sequence encompassing the promoter regions of putative *LHY* or *CCA1* orthologues could be identified for only 4 plant species. Secondly, differentiating between orthologues of *LHY* or *CCA1* proved difficult as the proteins encoded by these genes share a great deal of similarity to one another, such that LHY and CCA1 proteins are comparably similar to one another as to any potential orthologue. Furthermore, in three of the four species, only one *LHY/CCA1* homologue could be identified, raising the possibility that oscillator architecture is slightly different in these species, con-

Factor or Site Name	Location	Strand	Signal Sequence	Database
5UTR Py-rich stretch	10	-	TTTCTTCTCT	PLantCARE
GATABOX	27	+	GATA	PLACE
DOFCOREZM	37	+	AAAG	PLACE
CACTFTPPCA1	39	-	YACT	PLACE
CACTFTPPCA1	44	-	YACT	PLACE
CAT-box	50	+	GCCACT	PlantCARE
SORLIP1AT	52	+	GCCAC	PLACE
UPRMOTIFIIAT	53	+	$CC(N)_{12}CCACG$	PLACE
CACTFTPPCA1	54	+	YACT	PLACE
CAAT-box	57	+	CAAT	PlantCARE
CAATBOX1	59	+	CAAT	PLACE
ROOTMOTIFTAPOX1	60	-	ATATT	PLACE
GATABOX	62	-	GATA	PLACE
GTGANTG10	64	-	GTGA	PLACE
ABREATRD22	65	-	RYACGTGGYR	PLACE
G-box	67	+	CACGTG	PlantCARE
IRO2OS	67	-	CACGTGG	PLACE
ABRERATCAL	68	+	MACGYGB	PLACE
CACGTGMOTIF	68	+	CACGTG	PLACE
DPBFCOREDCDC3	68	-	ACACNNG	PLACE
ABRELATERD1	69	+	ACGTG	PLACE
ACGTABREMOTIFA2OSEM	69	+	ACGTGKC	PLACE
ACGTATERD1	69	+	ACGT	PLACE
GADOWNAT	69	+	ACGTGTC	PLACE
CBFHV	73	-	RYCGAC	PLACE
PRECONSCRHSP70A	81	+	$SCGAYNR(N)_{15}HD$	PLACE
Skn-1 motif	83	-	GTCAT	PlantCARE
WBOXHVISO1	85	+	TGACT	PLACE
GARE-motif	89	-	AAACAGA	PlantCARE
PYRIMIDINEBOXHVEPB1	94	+	TTTTTTCC	PLACE
GT1CONSENSUS	95	-	GRWAAW	PLACE
GT1GMSCAM4	95	-	GAAAAA	PLACE
TATABOX2	102	-	TATAAAT	PLACE

Table 4.1: Candidate transcription factor binding sites in the -957 to -847 region of the *LHY* promoter identified in the PlantCARE and PLACE *cis*-element databases

taining only on *LHY/CCA1* homologue where two serve in *Arabidopsis*.

4.3 Identification of protein binding sites by electrophoretic mobility shift assays

In order to identify potential transcription factor binding sites on the *LHY* promoter, electrophoretic mobility shift assays (EMSAs) were carried out using the segments of LHY upstream sequence that were identified in Chapter 2 as having a regulatory role.

The electrophoretic mobility shift assay relies on the reduced electrophoretic mobility of a protein-DNA complex compared to the mobility of the free DNA. A radiolabelled probe is created from the DNA under study and the mobility of this probe under electrophoresis is compared in the presence and absence of putative DNA-binding proteins or, in the case of work presented here, a protein-enriched plant extract. Protein binding sites in the probe sequence can be identified either by systematic mutagenesis of the probe DNA until a specific mutation is identified that abolishes the mobility shift, or, indirectly, using competition assays.

In a competition assay, a large molar excess of unlabelled competitor DNA is added to the binding reaction such that, if the competitor DNA contains the same binding site as the labelled probe, protein binding to the probe will be out-competed by the more abundant competitor DNA, resulting in a reduced proportion of the less mobile protein-probe species on electrophoresis. This approach can be used for finer mapping of binding sites on the probe by using a set of partially overlapping competitors that spans the length of the probe.

Electrophoretic mobility shift assays were initially carried out using a protein extract made from 10 day old *Arabidopsis* seedlings entrained to 12:12

light dark cycles before transfer to constant light for 24 hours. Tissue was harvested at this point (ZT24), corresponding to the subjective dawn, in the expectation that factors responsible for the activation of *LHY* expression would be most abundant at the time corresponding to the onset of *LHY* expression.

4.4 Identification of binding sites between positions -957 and -847

In Chapter 3, it was suggested that a rhythmic input, R_B, acts on the *LHY* promoter between -957 and -847 base pairs upstream of the translational start site. Deletion analysis of p*LHY::luc*$^+$ reporter constructs suggested that this input may take the form of a repressive activity that holds *LHY* expression at low levels during the subjective night. This repressive activity may also play a role in the response of *LHY* expression to induction by light signals.

To generate an EMSA probe for this region, the LUC1 plasmid was amplified by PCR using the primers H286D1New and B157R1New and the amplified fragment cloned into the pCR2.1 vector. Digestion of this plasmid with the restriction enzyme NgoMIV yielded an appropriate fragment for radiolabelling and use in EMSAs.

To identify binding sites within this probe, a set of 30 base-pair overlapping competitors was designed, covering the entire length of the probe (Figure 4.1 A). These competitors were added to the binding reactions in 100-fold excess of the probe.

Compared to the single, high mobility band produced by the free probe, EMSAs carried out using this probe and ZT24 protein extracts produced a relatively complex pattern of mobility-shifted bands, consisting of at least four reduced mobility complexes (Figure 4.1 B).

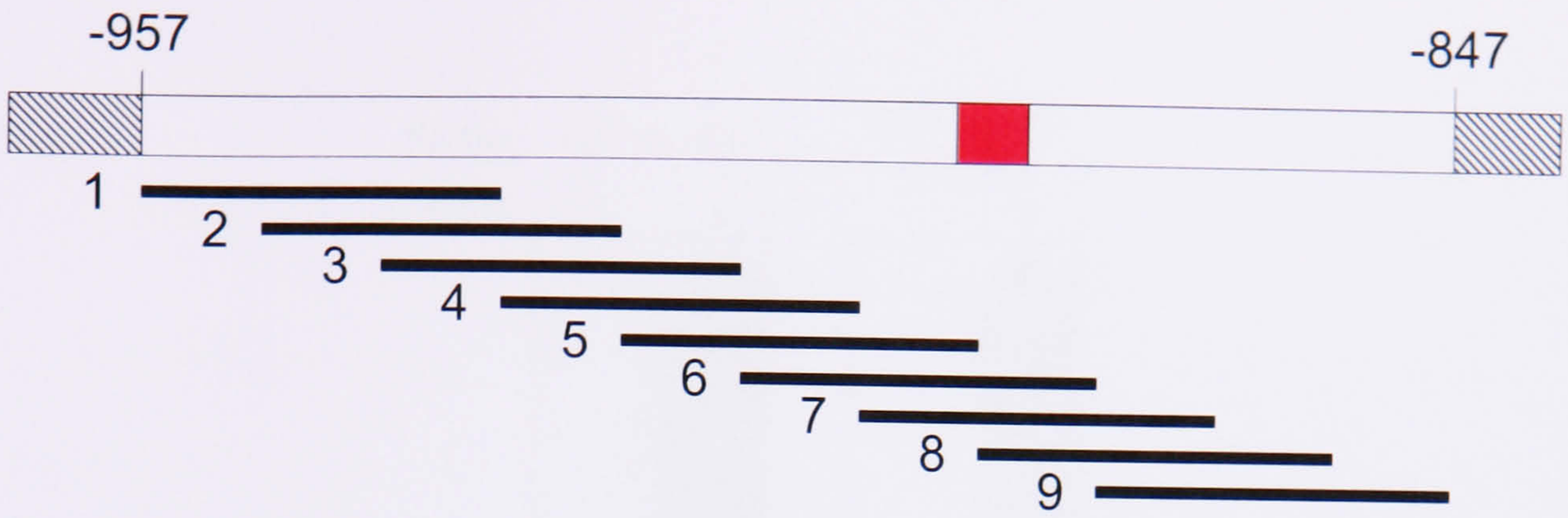
In competition assays, complexes I and II were consistently co-competed

by competitors 6, 7, 8 and 9, suggesting that these complexes represent different species of a common underlying binding activity (Figure 4.1B). For convenience, this binding activity will hereafter be referred to as activity "A". Similarly, complexes III and IV were co-competed by competitors 3, 4, 8 and 9, suggesting that these two complexes also arise from a common binding activity, activity "B".

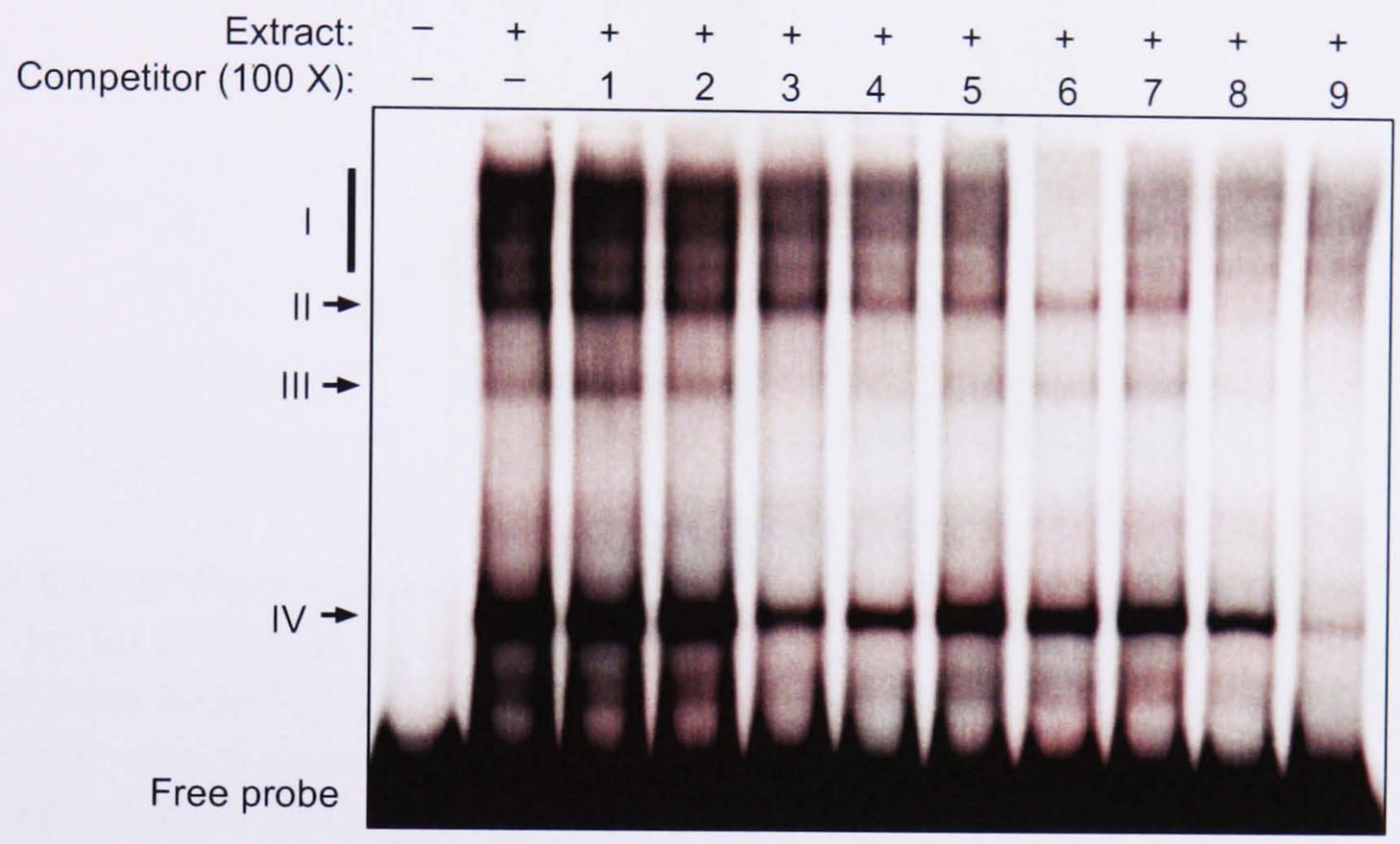
To identify any potential binding sites for activity "B" (complexes III and IV), the sequences of competitors 3, 4, 8 and 9 were examined for any common sequence motifs. Visual inspection of these sequences yielded two instances of an AAAAA sequence motif. This motif is present on the forward strand of competitors 3 and 4 and on the complement strand of competitors 8 and 9 (Figure 4.1 C). Neither of the PLACE or PlantCARE databases contained any description of this sequence motif, however, neither did these databases contain any other previously described *cis*-elements common to all of the competitors. The AAAAA sequence motif is, therefore, a strong candidate binding site for the activity represented by complexes III and IV.

Inspection of competitors 6, 7, 8 and 9, which all compete for activity "A" (complexes I and II), revealed the presence of the G-box motif in competitors 6 and 7, though no obvious sequence motif was common to all four competitors (Figure 4.1 C). To investigate the contribution of the G-box motif to the formation of complexes I and II, a second probe to the -957 to -847 region was created by amplification and cloning of the -957 to -847 region of the -957GboxMpLHY:luc+ construct (Dr Jae-Yean Kim) in the same manner as the wild-type probe was created. In this construct, the G-box motif was mutated from AC<u>CACGTG</u>TC (where the underlined sequence is the canonical G-box core hexamer) to AC<u>CACCCGG</u>GG. This mutation had previously been shown to prevent binding of the bHLH transcription factor PIF3 to a G-box target sequence *in vitro* (Martínez-García *et al.*, 2000). Compared to

A



B



C

Competitor 3: acttgagatataccaaaaagtgcagtagac
Competitor 9: gcgatgacttctgttttttccatttatacc

Competitor 6: agccactacaatatcaccacgtgtcgatct
Competitor 7: atatcaccacgtgtcgatctgcgatgactt

Figure 4.1: Competition EMSA to identify protein binding sites within a -957 to -847 probe.
A) Schematic of the -957 to -847 EMSA probe. Hatched boxes represent extensions incorporating NgoMIV restriction sites for radiolabelling. The position of the G-box motif is shaded in red. Competitor oligonucleotides are shown below as black bars.
B) Competition EMSA with the -957 to -847 probe and a ZT24 extract from 8 day old seedlings. Binding reactions were prepared as described in Chapter 2 before electrophoresis on a 6% acrylamide:bis-acrylamide, 0.25 x TBE gel at 200V. Where present, competitor oligonucleotides were added in 100-fold molar excess of the probe.
C) Nucleotide sequences of competitors 3 and 9, and competitors 6 and 7. Common sequence motifs between each pair are underlined.

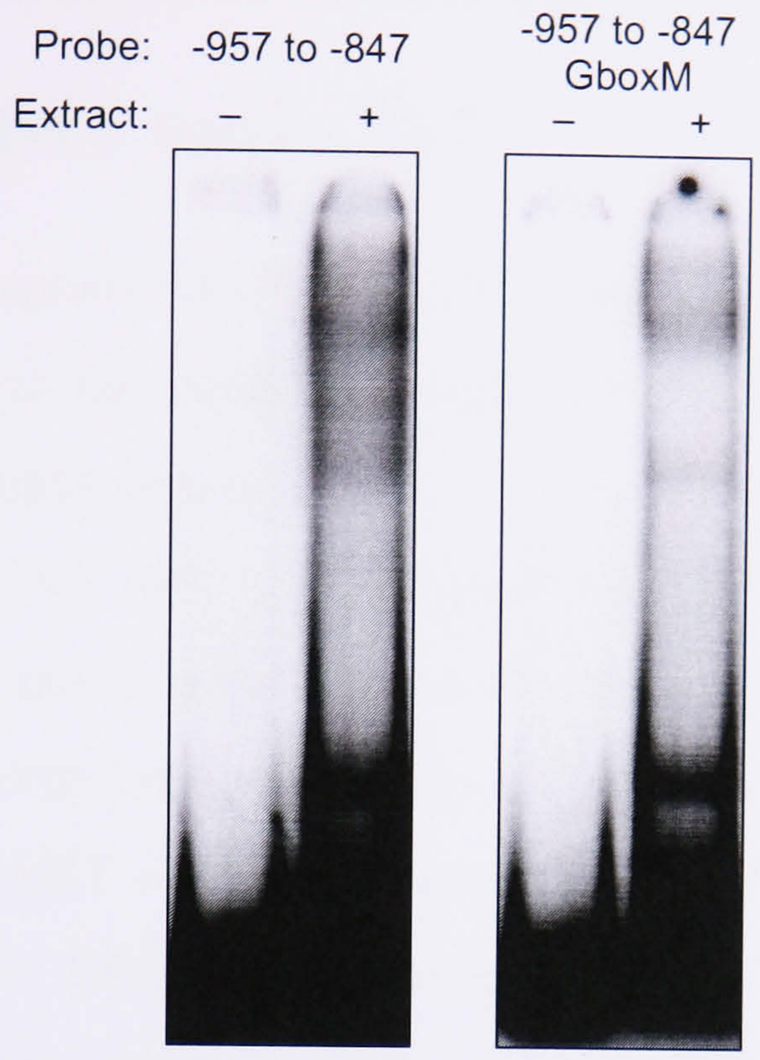


Figure 4.2: EMSAs using a ZT24 protein extract from 8 day old seedlings The -957 to -847 probe is compared to an otherwise identical probe in which the G-box was mutated from to ACCACGTGTC to ACCACCCGGG (-957 to -847 GboxM). Binding reaction were prepared as described and electrophoresed on a 6% acrylamide:bis-acrylamide, 0.25 x TBE gel at 200V.

the wild-type probe, complex I and II formation was greatly reduced on the -957 to -847 GboxM probe, demonstrating that complexes I and II represent a G-box binding activity (Figure 4.2).

Why competitors 8 and 9, which contain no G-box motif, should also compete for complexes I and II is unclear. One explanation may be that, while competitors 8 and 9 do not contain a binding site that directly competes for complex I and II formation, the protein complexes that do form on competitors 8 and 9 (and, with a lesser affinity, on competitors 3 and 4), are able to interact with the proteins of the G-box-binding complex. If these protein-protein interactions were strong enough, they may be able out-compete the DNA-protein interaction of the G-box dependent complex.

4.5 Identification of binding sites between positions -847 and -757

The -847p*LHY::luc*$^+$ reporter described in Chapter 3 was shown to be capable of responding to a rhythmic input, R_A, and to be capable of acute induction by both red and blue light signals.

Since analysis of the *LHY* 5′ UTR showed that sequence between positions -641 and +1 was not required for the rhythmic expression of *LHY*, analysis of potential transcription factor binding sites that may be implicated in the regulation of -847p*LHY::luc*$^+$ expression was limited to the region between positions -847 and -641. Due to the technical difficulty of PCR amplification of the repetitive CT-rich region between positions -757 and -641, this region was analysed separately, as described below.

A probe corresponding to the region between positions -847 and -757 was generated by PCR amplification of the LUC1 plasmid using the primers H163D1New and B94R1New. Cloning of the PCR product and preparation of the probe were as described for the -957 to -847 probes. Again, a set of 30 base-pair overlapping competitors were designed to span the length of the probe (Figure 4.3 A).

Electrophoretic mobility shift assays carried out using this probe and ZT24 protein extract again revealed a number of different reduced mobility complexes (Figure 4.3 B). Since all of these complexes were competed for by competitor 14, and to a slightly lesser extent, by competitor 15, these complexes are, again, likely to represent a single DNA binding activity.

Interestingly, the sequences of competitors 14 and 15 contain the same AAAAA motif identified as a candidate binding site on the -957 to -847 probe (Figure 4.3 C). Moreover, the stronger competitor, competitor 14, contains three instances of this motif compared to one instance in the weaker competitor

15. If the AAAAA motif does constitute a binding site, then one would expect that competitors 14 and 15 would out-compete the -957 to -847 probe for binding of activity "B". Conversely, one would expect that competitors 3, 4, 8 and 9 would out-compete the -847 to -757 probe for complex formation.

To test this hypothesis, EMSAs were performed using a short probe derived from competitor 14, referred to hereafter as probe 14, rather than the longer PCR generated probes. Competitors 3 and 9 were both able to compete with this probe for complex formation, demonstrating that complexes formed on the -847 to -757 probe rely on the same underlying DNA-binding property as activity "B" on the -957 to -847 probe (Figure 4.4 A). Because the AAAAA motif is represented in all competitors that interact with this binding activity, this is also strong circumstantial evidence that this motif may be the DNA target sequence for activity "B".

If the AAAAA motif is responsible for the binding of activity "B", this would mean that activity "B" has five binding sites on the *LHY* promoter. Interestingly, three of these binding sites would lie in the region between positions -847 to -747 that was investigated in Bethan Taylor's linker scan analysis. Since no single linker scan mutation would have affected all three binding sites, this may well explain why none of Taylor's linker scan constructs displayed any change in circadian expression.

While all sequences capable of competing for DNA-binding activity "B" contain the AAAAA motif on either the forward or reverse strands, it should be noted that not all competitors containing this sequence were able to compete for activity "B". Specifically, competitor 12 contains one instance of this motif on the forward strand and competitor 13 contains two instances, one on each strand though no competing activity was observed for either sequence. Nonetheless, the AAAAA motif remains the strongest candidate binding site for activity "B".

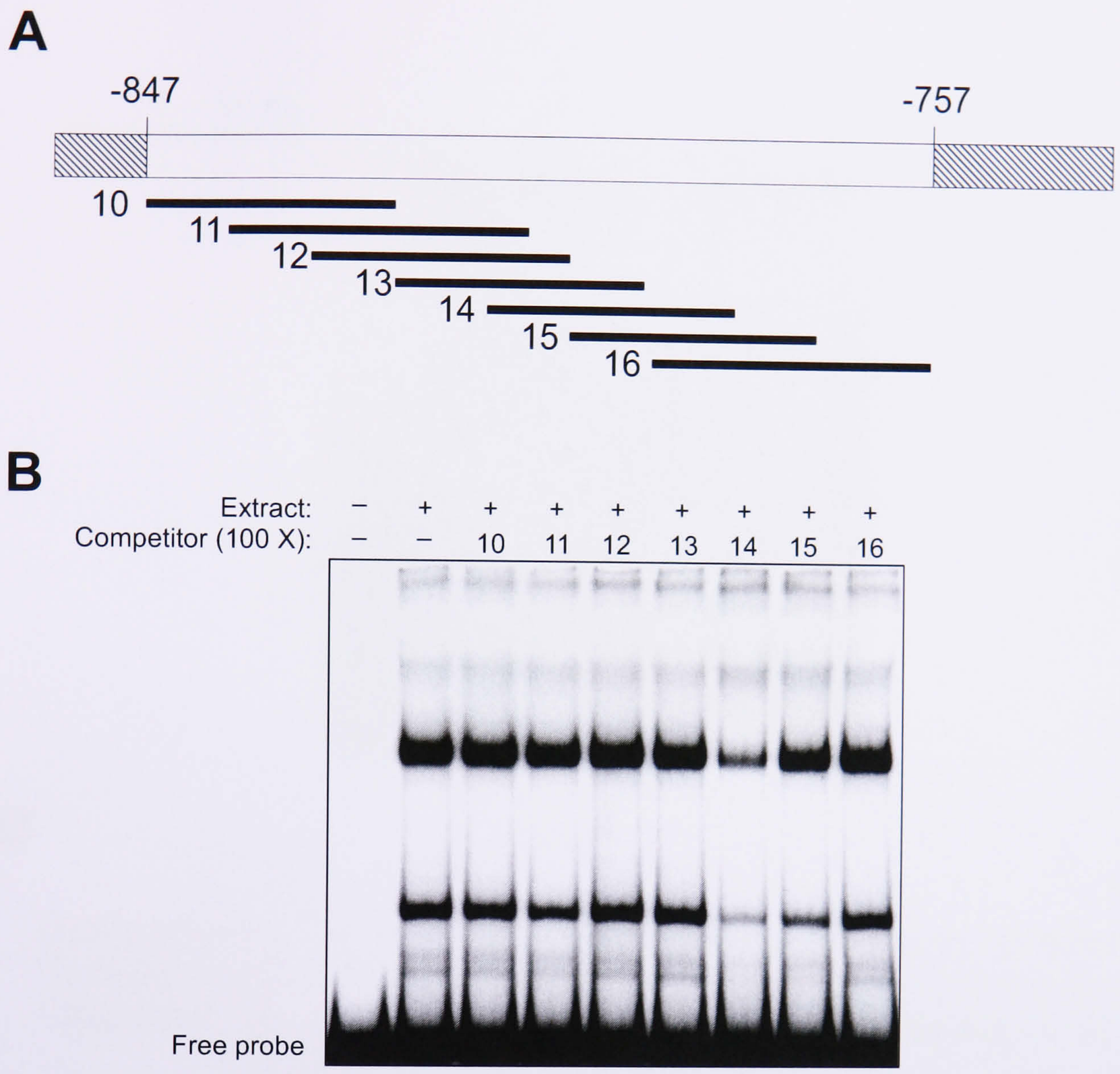
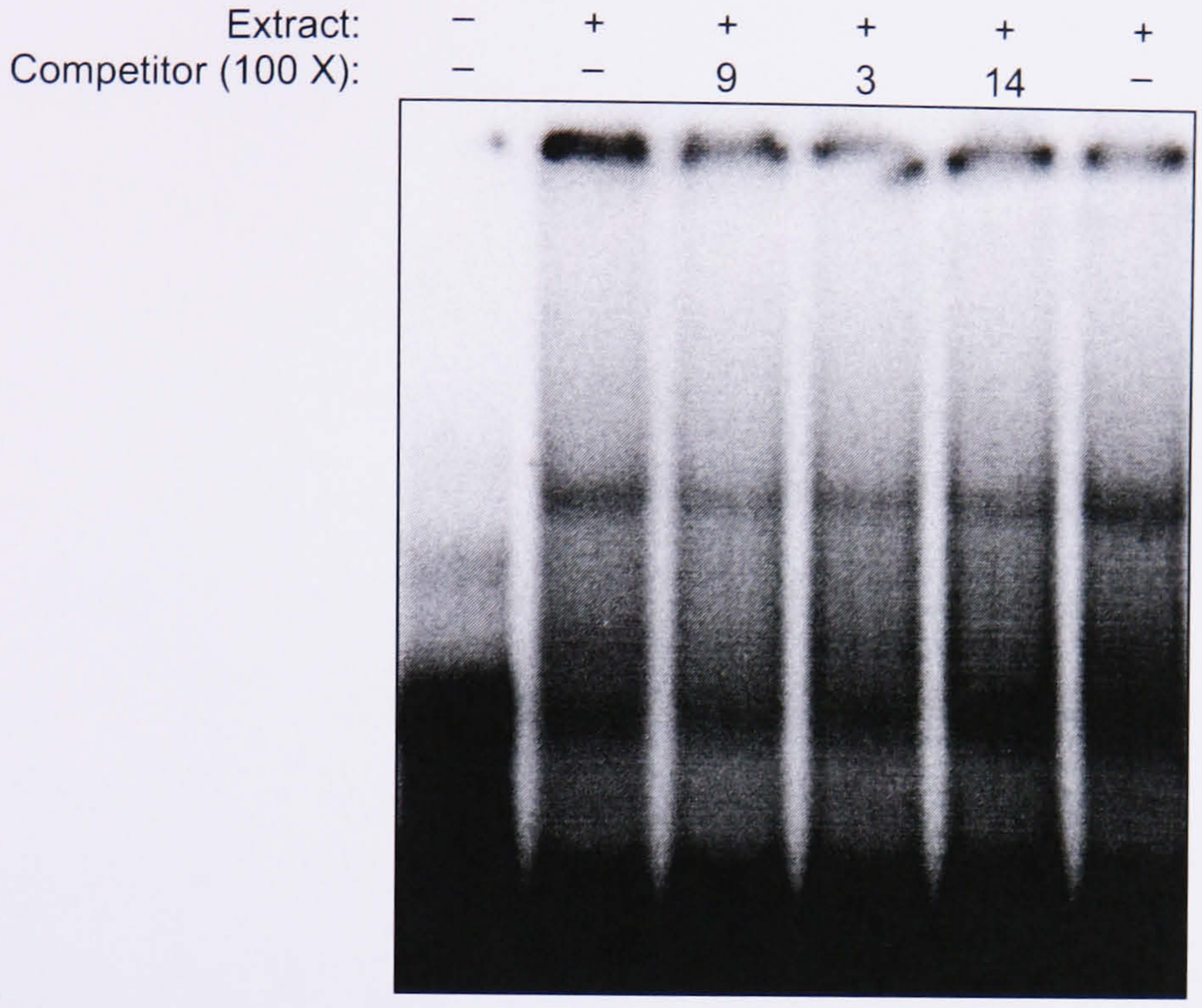


Figure 4.3: A) Schematic of the -847 to -757 EMSA probe. Hatched boxes represent extensions incorporating NgoMIV restriction sites for radiolabelling. Competitor oligonucleotides are shown below as black bars. B) EMSA competition assay with the -847 to -757 probe and a ZT24 extract from 8 day old seedlings. Binding reactions were prepared as described in Chapter 2 before electrophoresis on a 6% acrylamide:bis-acrylamide, 0.25 x TBE gel at 200V. Where present, competitors were added in 100-fold molar excess of the probe. C) Nucleotide sequences of competitors 14 and 15. Common sequence motifs between these competitors are underlined.

A



B

Competitor 9: gcgatgacttctg<u>tttttt</u>ccatttatacc
Competitor 3: acttgagatatacc<u>aaaaa</u>gtgcagtagac
Competitor 14: a<u>ttttt</u>cc<u>aaaaa</u>ttaggg<u>aaaaa</u>ttgtt

Figure 4.4: Cross competition between competitors 9, 3 and 14 A) EMSA competition assay with Probe 14, derived from competitor 14, and a ZT24 extract from 8 day old seedlings. Binding reactions were prepared as described in Chapter 2 before electrophoresis on an 8% acrylamide:bis-acrylamide, HEPES-EDTA gel at 200V. Competitors were added in 100-fold molar excess of the probe. B) Nucleotide sequences of competitors 9, 3 and 14 with the common sequence motifs underlined.

4.6 Identification of candidate binding sites within the CT-rich region, -756 to -641

The repetitive nature of the CT-rich region of the *LHY* 5′ UTR necessitated a slightly different approach to the design of EMSA probes as PCR of repetitive regions is technically challenging due to the greater risk of mis-priming of the reaction. Instead, four short, overlapping probes spanning this region were generated from synthetic oligonucleotides (Figure 4.5 A).

While the ctr1 probe did not show any detectable mobility shift in the presence of ZT24 protein extract, the remaining probes, ctr2, ctr3 and ctr4 were all capable of complex formation with this extract (Figure 4.5 B). As these probes have the repeated CTT motif in common (Figure 4.6 A), cross-competition assays were performed to determine whether the reduced-mobility complexes formed on each of these probes represent common or distinct DNA-binding activities. In these experiments, binding to all three complex-forming ctr probes was out-competed by oligonucleotide competitors ctr3 and ctr4, suggesting a common binding activity between the three probes (Figure 4.6 B). However, neither ctr2 nor ctr3 were able to out-compete the ctr4 probe for the lower complex, suggesting that this complex may represent a ctr4-specific binding activity.

Examination of the ctr2, ctr3 and ctr4 sequences reveals the only common sequence motif to be the repeated-CTT motif, suggesting that this repeated motif represents a protein binding site. Indeed, the ctr3 competitor consists almost entirely of the repeated ctt motif and is able to form protein-DNA complexes and to act as a competitor against ctr2 and ctr4 probes. The only candidate binding site for the ctr4-specific complex is the CAGTTAT motif, as this is clearly distinct from sequence found in other competitors. This motif has also previously been identified as a binding site for members of the

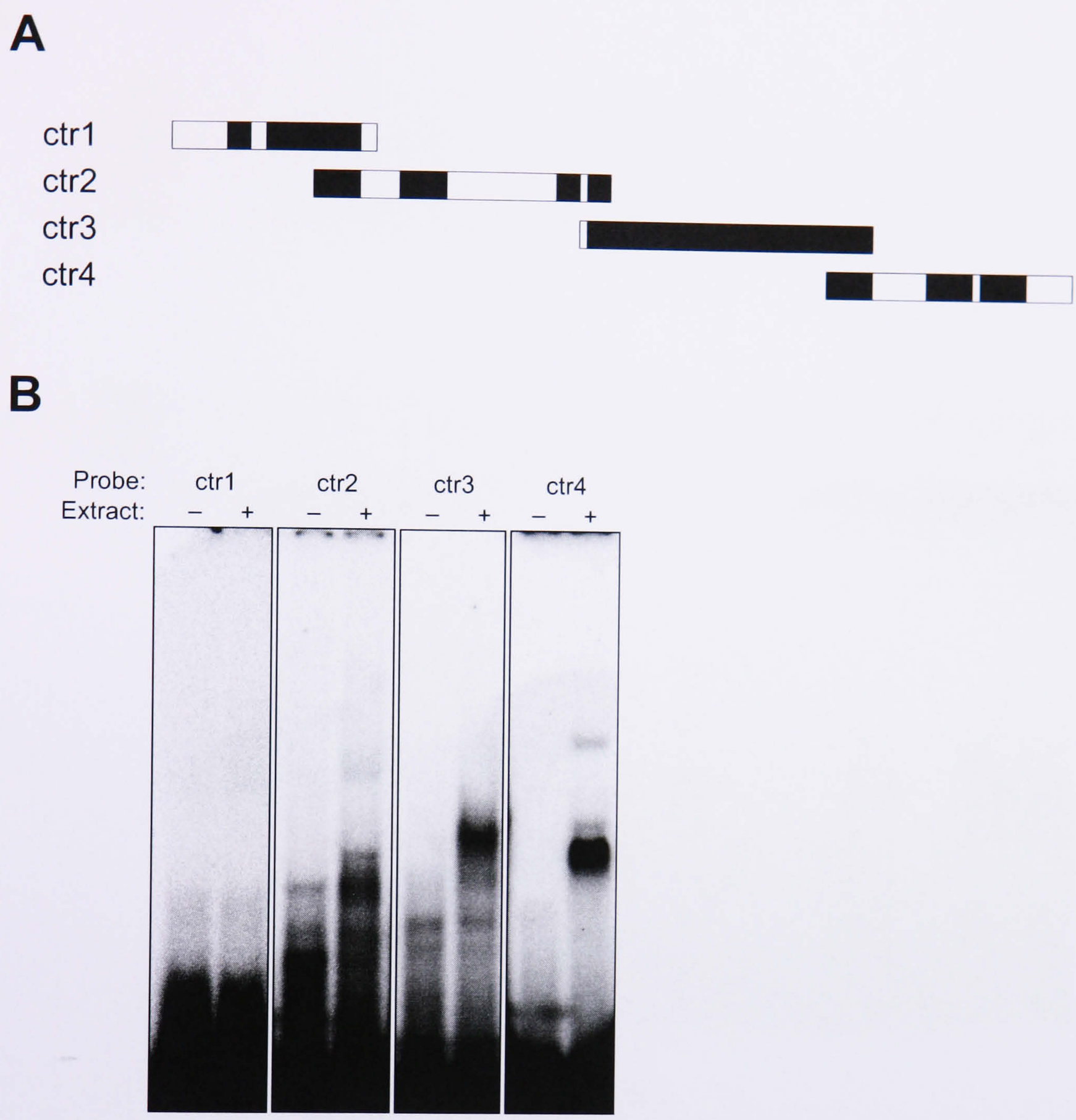


Figure 4.5: EMSA to identify protein binding sites within a -957 to -847 CT-rich region A) Schematic of probes to the CT-rich region of the LHY upsteam regulatory region. Black shaded areas represent regions comprised of the sequence motif CTT B) EMSA competition assay with probes ctr1, ctr2, ctr3 and ctr4 using a ZT24 extract from 8 day old seedlings. Binding reactions were prepared as described in Chapter 2 before electrophoresis on an 8% acrylamide:bis-acrylamide, HEPES-EDTA gel at 200V.

A

ctr1: gcttctggcttctcttcttcttcttc
ctr2: cttcttccagtcttcttcagcctaaaacagtcttcctt
ctr3: tccttcttcttcttcttcttcttcttcttcttcttctt
ctr4: cttcttcagttatcttcttccttcttctctct

B

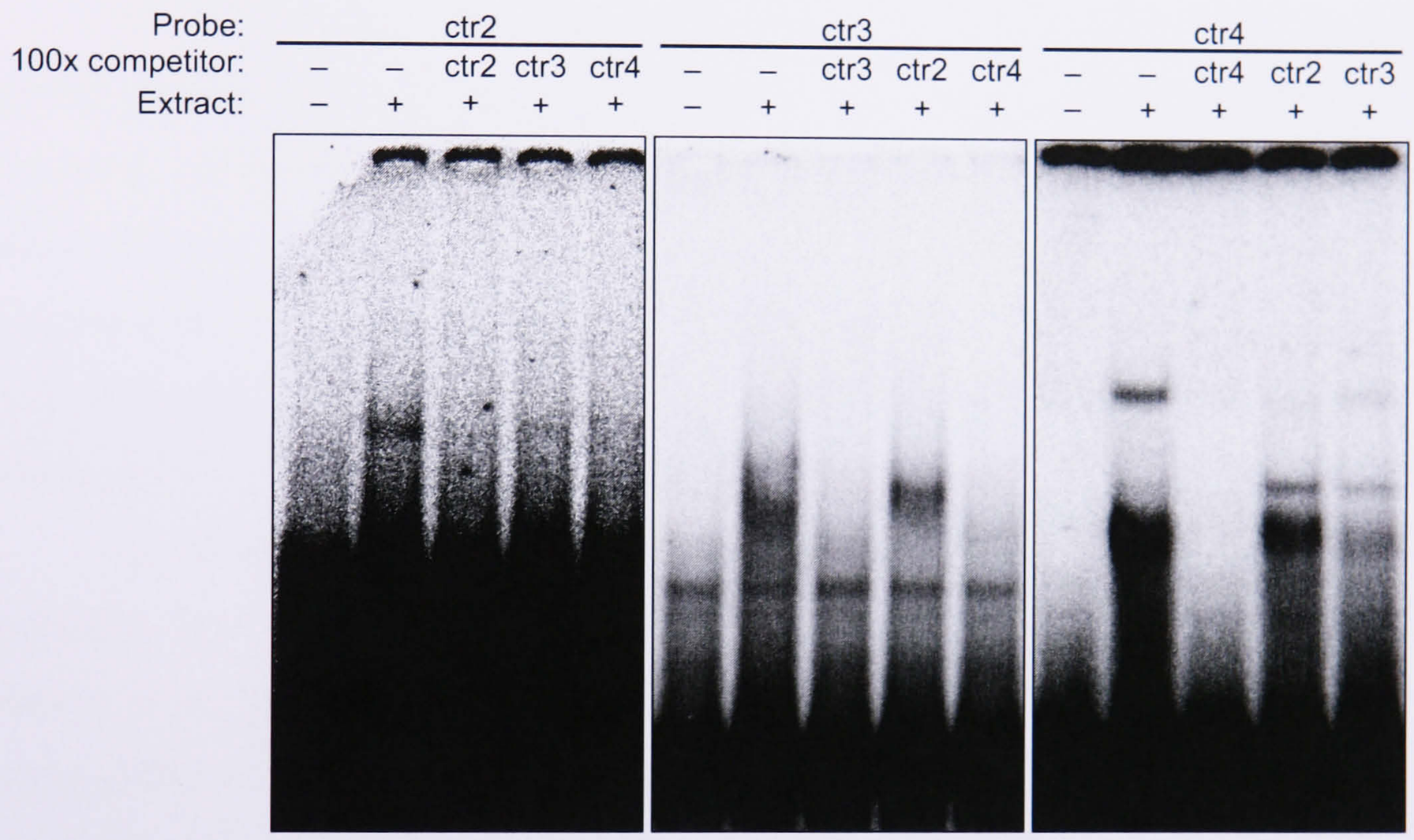


Figure 4.6: Cross competition EMSA using probes to the CT-rich region. A) Nucleotide sequences of probes/competitors to the CT-rich region of the LHY upstream regulatory region. B) EMSA competition assay with probes ctr1, ctr2, ctr3 and ctr4 using a ZT24 extract from 8 day old seedlings. Binding reactions were prepared as described in Chapter 2 before electrophoresis on an 8% acrylamide:bis-acrylamide, HEPES-EDTA gel at 200V. Competitor oligonucleotides were added in 100-fold molar excess of the probes.

Myb family of transcription factors.

4.7 Discussion

In this chapter, electrophoretic mobility shift assays have identified at least four sequence motifs in the *LHY* promoter that mediate protein-DNA interactions *in vivo* (Figure 4.7).

The first of these sequence motifs, the G-box motif (CACGTG), is a specific instance of the E-box motif (CANNTG), the consensus binding site for the basic helix-loop-helix family of transcription factors (Toledo-Ortiz et al., 2003). At least 147 bHLH transcription factors are encoded in the *Arabidopsis* genome and, given that bHLH proteins have been shown to function either as homodimers or heterodimers, regulation by bHLH factors is likely to be highly combinatorial (Toledo-Ortiz *et al.*, 2003).

As discussed in Chapter 3, a G-box motif has previously been suggested to mediate transcriptional responses to red light signals (Martínez-García *et al.*, 2000), however, analysis of the transcriptional responses of the -847p*LHY::luc*$^{+}$ construct, in which the G-box motif was absent, showed that the *LHY* G-box was not required for acute transcriptional induction by red light. Interestingly, however, data presented in Chapter 3 suggested that the region between positions -957 and -847 mediates a second rhythmic input to the *LHY* promoter, so it is possible that the G-box motif that lies within that region may mediate a circadian input to LHY. Several members of the PIF3-like subfamily of bHLH transcription factors (PIF3, PIF4, PIL1 and PIL6), which recognise the G-box motif, have been shown to physically interact with TOC1 *in vivo* (Makino *et al.*, 2002; Yamashino *et al.*, 2003), suggesting that these proteins may function as co-factors with TOC1 as TOC1 itself contains no known DNA-binding domain (Strayer *et al.*, 2000). Together, these ob-

servations present the hypothesis that a TOC1-dependent rhythmic input to *LHY* may be mediated by the G-box motif.

With this hypothesis in mind, EMSAs were carried out using extracts from *TOC1*ox (Courtesy of Dr Laszlow Kozma-Bognar) and *TOC1* RNAi (Más *et al.*, 2003a) seedlings and their cognate wild-types, however, no change in the proportion of the reduced mobility complex was detected with any of the probes described in this chapter (data not shown). This may reflect the technical limitation of *in vitro* assays, such as the EMSA, alternatively, it is possible that the binding of complexes to the *LHY* promoter is not a rhythmic process and rhythmicity is transduced by altering the activities of constitutively bound transcription factor complexes.

The second sequence motif associated with protein-DNA interactions with the LHY upstream region was the (GG/CC)AAAAA motif. This motif was present in multiple copies in the *LHY* upstream region and was present in both the -967 to -847 and -847 to -757 regions of *LHY*. As this motif could not be reconciled with any of the characterized *cis*-elements described in the PLACE and PlantCARE databases, it is not possible to say whether it represents a novel element or simply a non-consensus form of a previously described element. The presence of this element in the -847p*LHY::luc*$^{+}$ construct places the element as a candidate mediator of the action of rhythmic input R_A on the *LHY* promoter and also as a potential target of light signalling to *LHY*. A mutation to this element that abolishes protein binding allows both these hypotheses to be tested in Chapter 6. Interestingly, there seems to be some interaction between DNA-protein complex formation on this element and complex formation on the G-box element (Figure 4.1 B), though the functional relevance, if any, of this is unclear.

Finally, two further protein-binding sites were identified within the CT-rich region of the *LHY* upstream regulatory region. The first of these motifs

consists of several runs of CTT repeats, which are present in multiple copies between positions -757 and -641. The second binding site conforms to the consensus binding site for Myb transcription factors. Both these elements, like the AAAAA motif, are candidates for mediating the transcriptional responses to light signals or rhythmic input R_A by virtue of their presence in the -847p*LHY::luc*$^+$ construct, although the design of mutations affecting independently affecting binding to either site were not designed due to technical constraints.

In an attempt to identify which of the four candidate transcription factor binding sites might mediate the two rhythmic inputs to the *LHY* promoter, EMSAs were carried out for each probe using a series of protein extracts from seedlings harvested at four hour intervals from ZT24 to ZT52. Disappointingly, no changes in complex formation were observed across the time series for any of the probes (data not shown). Again, this may be because changes in complex activity, rather than complex formation, transduce the rhythmic signals to the *LHY* promoter. An alternative explanation is that the rhythmic binding activity of any given transcription factors is masked by the activities of a larger number of related transcription factors, which may be expressed constitutively or with differing circadian phases.

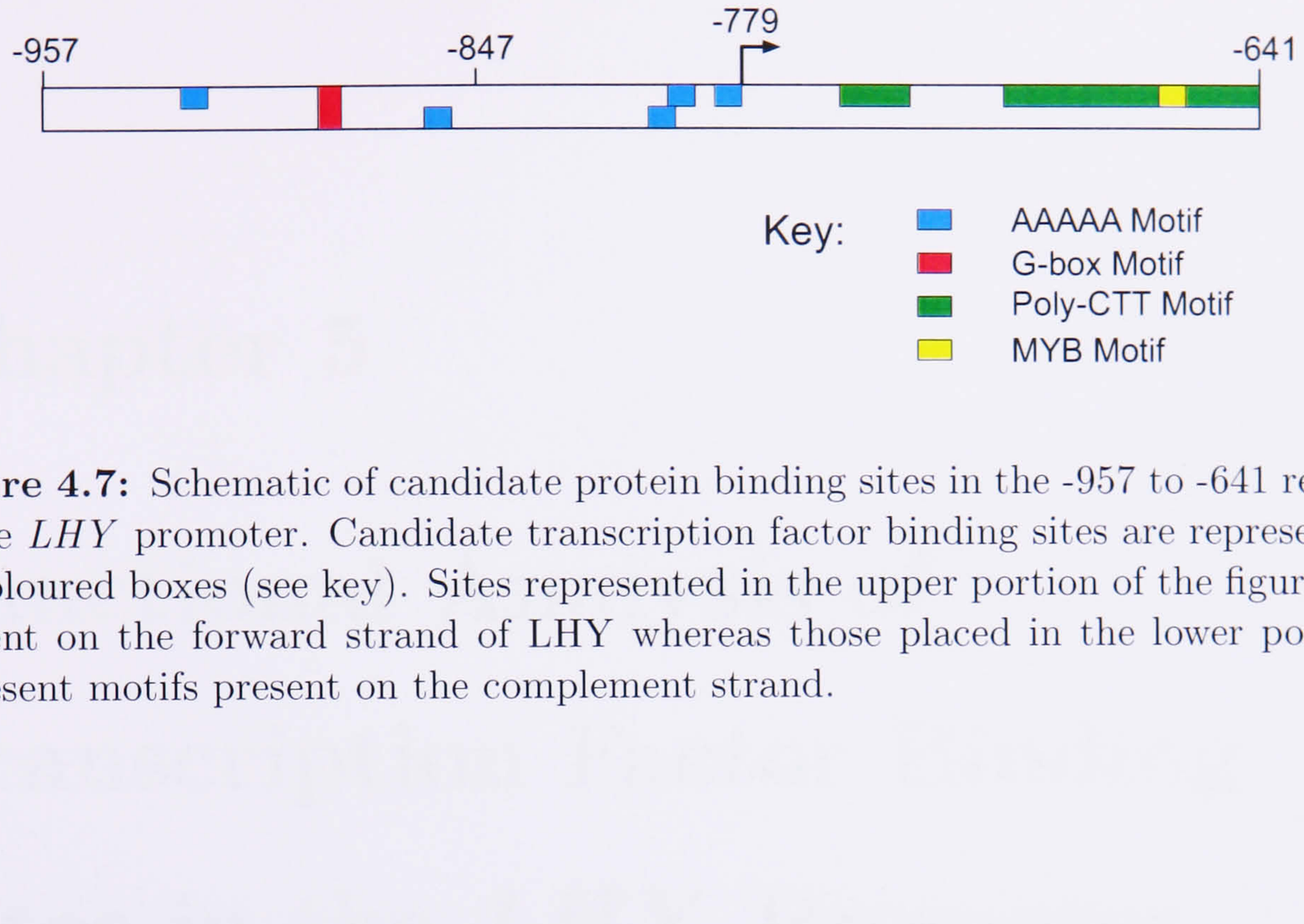


Figure 4.7: Schematic of candidate protein binding sites in the -957 to -641 region of the *LHY* promoter. Candidate transcription factor binding sites are represented by coloured boxes (see key). Sites represented in the upper portion of the figure are present on the forward strand of LHY whereas those placed in the lower portion represent motifs present on the complement strand.

Chapter 5

Functional Analysis of Transcription Factor Binding Sites in the *LHY* Promoter

5.1 Introduction

In Chapter 4, analysis of the *LHY* upstream sequence by electrophoretic mobility shift assays identified four sequence motifs that may mediate transcription factor binding to the LHY promoter including a G-box motif, an AAAAA motif associated with activity "B" binding, a poly-CTT tract and a putative Myb binding motif. Two of these sequence motifs, AAAAA and the CTT repeats, are present in multiple copies within the *LHY* upstream regulatory region.

Though these sequence motifs are associated with DNA-binding activities *in vitro*, it has yet to be shown if these binding sites function as *cis*-regulatory elements *in vivo*. Furthermore, it is not yet known whether these candidate transcription factor binding sites mediate light input or either of the two rhythmic inputs to the *LHY* promoter as described in Chapter 3. To

address these issues, this chapter describes the design of mutations that alter or abolish protein binding to these motifs *in vitro*. The *in vivo* function of these motifs was then determined by analysis of luciferase reporter constructs incorporating these mutations.

5.2 Design of Mutations to Candidate Transcription Factor Binding Sites.

5.2.1 Design of G-box Mutations

As discussed in Chapter 3, the G-box motif is a strong candidate for mediating the rhythmic input R_B on the *LHY* promoter. A mutation that abolishes binding to the G-box would then be expected to remove R_B from the set of rhythmic inputs to the *LHY* promoter leaving only rhythmic input R_A. In that case, a promoter carrying such a mutation would be expected to display the circadian phenotype of the -847p*LHY::luc*$^+$ construct, which lacks the G-box entirely, displaying an early phase of onset under constant light and a late peak phase under constant darkness.

The GboxM mutation of Martínez-García *et al.* (2000) (CACGTG to CACCCG) was shown to abolish binding of PIF3 to a G-box motif identified by random binding site selection and we have shown that this mutation also affects protein binding to the G-box of the *LHY* promoter (Figure 4.3 B). As such, this mutation would appear to be suitable for determining the role of the G-box in *LHY* expression. However, initial analysis of a -957GboxM*pLHY::luc*$^+$ reporter construct in *Arabidopsis* seedlings indicated that this construct was not expressed (Jae-Yean Kim, unpublished data). While this result suggests an important role for the G-box motif in the regulation of basal levels of *LHY* expression, the lack of detectable expression precluded any interpretation of

how that role might relate to circadian regulation. Therefore, several other mutations to the G-box motif were designed in the hope of altering the regulatory function sufficiently to be able to draw conclusions on the function of the G-box without completely abolishing expression of the reporter gene.

There is mounting evidence that the sequences flanking the core CACGTG hexamer of the G-box can further define the regulatory role of a given G-box element. For example, changes in the two nucleotides on either side of the core hexamer have been shown to alter the nature of DNA-protein complex formation on the G-box *in vitro*, affecting not only binding affinity but also which of two classes of electrophoretic mobility species was formed (Williams *et al.*, 1992). Furthermore, analysis of promoter sequences co-regulated by PHYA mediated far red light signals has also shown that certain G box-flanking sequences are over-represented in promoters that are co-induced by far red light, whereas a different set of flanking sequences is overrepresented in promoters that are co-repressed in those conditions (Hudson & Quail, 2003).

In the expectation that mutation of the flanking sequences of the *LHY* G-box would yield a weaker phenotype than mutation of the G-box core hexamer, four mutations were designed to the two basepair regions on either side of the *LHY* G-box (Table 5.1). These mutations were designed to reflect each of the classes of binding activity described by Williams *et al.* (strong and weak variants of class I and class II binding activities).

5.2.2 Design of mutations to the Activity "B"-binding motif

As described in Chapter 4, the AAAAA motif is a strong candidate motif for the Activity "B"-binding site, because this motif is common to all of the competitors that bind Activity "B".

To identify mutations that abolish binding of Activity "B", the candi-

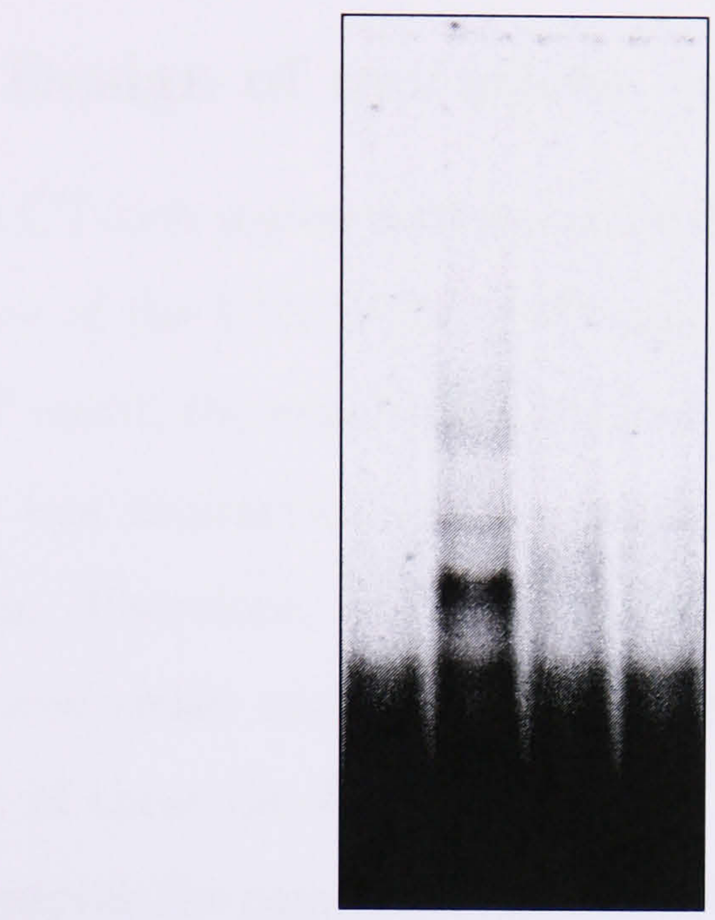
G-box Motif	Sequence
LHY G-box	AC<u>CACGTG</u>TC
Strong Class I	GA<u>CACGTG</u>TC
Weak Class I	AC<u>CACGTG</u>GT
Strong Class II	GT<u>CACGTG</u>AC
Weak Class II	CT<u>CACGTG</u>AG

Table 5.1: Mutations to the LHY G-box motif. Mutations were designed to the two basepair sequences on either side of the G-box core hexamer (underlined). These mutations correspond to those shown by Williams *et al.* (1992) to alter the *in vitro* binding properties of the G-box. The wildtype *LHY* G-box itself corresponds to a weak Class I motif.

date mutations were incorporated into competitor oligonucleotides to assess the effect of the mutation in competition assays. The first mutation assayed was the alteration of the AAAAA motif to AACCG. In the context of Competitor 14, applying this mutation to all three instances of the AAAAA motif failed to abolish competition (Figure 5.1). This indicated that the final three adenosine residues of the AAAAA motif are not critical for defining the activity "B" binding site.

Closer examination of competitors 3, 9, 14 and 15 revealed that the AAAAA motif is always preceded by the sequences GG or CC, so a second mutation of this motif was designed which also altered these preceding residues (Figure 5.2A). This second mutation, made in the context of Competitor 3, successfully abolished competition for binding of activity "B", indicating that this mutation prevented binding of the activity "B" complex (Figure 5.2 B). Three reporter constructs incorporating this mutation were designed to determine the regulatory function of the Activity "B" binding site (Figure 5.3). In the -957 1,2M construct, the two activity "B"-binding sites in the -957 to -847 region were mutated, in the -957 1,2,3,4.5M construct, all five instances of the activity "B" binding site in the -957 construct were mutated and in the -847

A



B

Competitor 14: a<u>ttttt</u>cc<u>aaaaa</u>ttaggg<u>aaaaa</u>ttgtt

Competitor 14M$_A$: a**cgg**<u>tt</u>cc<u>aa**ccg**</u>ttaggg<u>aa**ccg**</u>ttgtt

Figure 5.1: Cross competition EMSA to test a mutation to the AAAAA site. A) EMSA competition assay with probe 14 using a ZT24 extract from 8 day old seedlings. Binding reactions were prepared as described in Chapter 2 before electrophoresis on an 8% acrylamide:bis-acrylamide, HEPES-EDTA gel at 200V. Where present, competitor oligonucleotides were added in 100-fold molar excess of the probe. B) Nucleotide sequences of competitor 14 and competitor 14 incorporating the mutation to the AAAAA site (Competitor 14M_A). The AAAAA site is underlined and the mutation is highlighted by bold type.

3,4,5M construct, the three instances of this site in the -847 construct were mutated.

5.2.3 Design of mutations to the CT-rich region

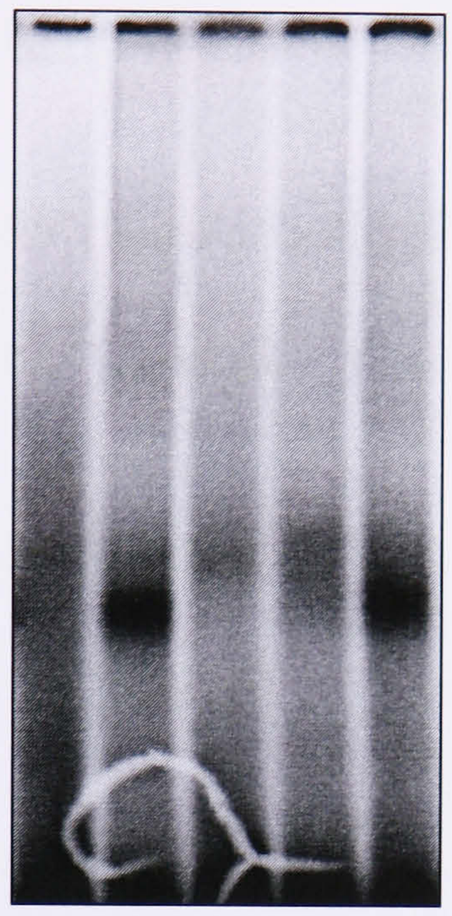
While the CT-rich region contains a candidate MYB binding site (CAGTTAT, an instance of the CNGTTR MYB core sequence) and three instances of the poly-CTT motif, the generation of point mutations in this region of a reporter construct was impractical due to the long, repetitive CTT repeat motifs in this region. Therefore, a series of deletions were designed within this region that removed either one, two or all three copies of the CTT-repeat motif. A schematic of these deletions is shown in Figure 5.4. Of these deletions, only Δ1ctr preserves the candidate Myb motif, allowing the functional contribution of this motif to be ascertained by comparing the phenotypes of Δ1ctr and Δ2ctr constructs. Each of these deletions was made in the context of both -957p*LHY::luc*$^+$ and -847p*LHY::luc*$^+$ constructs.

5.2.4 Generation of Mutant Reporter Constructs and Transformation of *Arabidopsis thaliana*

Mutant reporter constructs were created by QuikChange^{TM} site directed mutagenesis of either -957p*LHY::luc*$^+$ or -847p*LHY::luc*$^+$ constructs in the pGreen vector. These constructs were then transformed into *Agrobacterium tumefaciens* C58 for transformation of *Arabidopsis thaliana*, ecotype Ws. by floral dip. The strategy used to generate each construct and the number of transgenic lines generated are summarized in Table 5.2.

Initially, attempts were made to select Arabidopsis primary transformants by screening for the kanamycin resistance conferred by the *NptII* gene in the pGreen T-DNA but this approach recovered no kanamycin resistant trans-

A



B

Competitor 3: acttgagatataccaaaaagtgcagtagac

Competitor 3M: acttgagatata**tgtc**aaagtgcagtagac

Figure 5.2: Cross competition EMSA to test a second mutation to the AAAAA site. A) EMSA competition assay with probe 14 using a ZT24 extract from 8 day old seedlings. Binding reactions were prepared as described in Chapter 2 before electrophoresis on an 8% acrylamide:bis-acrylamide, HEPES-EDTA gel at 200V. Competitor oligonucleotides were added in 100-fold molar excess of the probe. B) Nucleotide sequences of competitor 3 and competitor 3 incorporating the mutation to the AAAAA site (Competitor 3M). The AAAAA site is underlined and the mutation is highlighted by bold type.

Construct	Template	Primers	Number of Transgenic Lines
-957 1M	-957p*LHY::luc*$^+$	Site 1M F Site 1M R	N/A
-957 1.2M	-957 1M	Site 2M F Site 2M R	6
-957 [345M]1	-957p*LHY::luc*$^+$	Site 345M1 F Site 345M1 R	N/A
-957 3,4,5M	-957 [345M]1	Site 345M2 F Site 345M2 R	4
-957 [1345]M	-957 3,4,5,M	Site 1M F Site 1M R	N/A
-957 allM	-957 [1345]M	Site 2M F Site 2M R	0
-957 Strong Class I	-957p*LHY::luc*$^+$	5' Pal F 5' Pal R	5
-957 Weak Class I	-957p*LHY::luc*$^+$	3' Pal F 3' Pal R	6
-957 Strong Class II	-957p*LHY::luc*$^+$	St Cl II F St Cl II R	6
-957 Weak Class II	-957p*LHY::luc*$^+$	Wk Cl II F Wk Cl II R	4
-847 [345M]1	-847p*LHY::luc*$^+$	Site 345M1 F Site 345M1 R	N/A
-847 3,4,5M	-847 [345M]1	Site 345M2 F Site 345M2 R	6
-847 Δ1ctr	-847p*LHY::luc*$^+$	Δ1ctr F Δ1ctr R	0
-847 Δ2ctr	-847p*LHY::luc*$^+$	Δ2ctr F Δ2ctr R	0
-847 Δ3ctr	-847p*LHY::luc*$^+$	Δ3ctr F Δ3ctr R	6

Table 5.2: Generation of mutant reporter constructs. Mutant reporter constructs were generated by QuikChange site-directed mutagenesis of a template construct. using the primers indicated. Where the number of transgenic lines recovered is shown as N/A (not applicable). this indicates that a construct was merely an intermediate used in the generation of another construct.

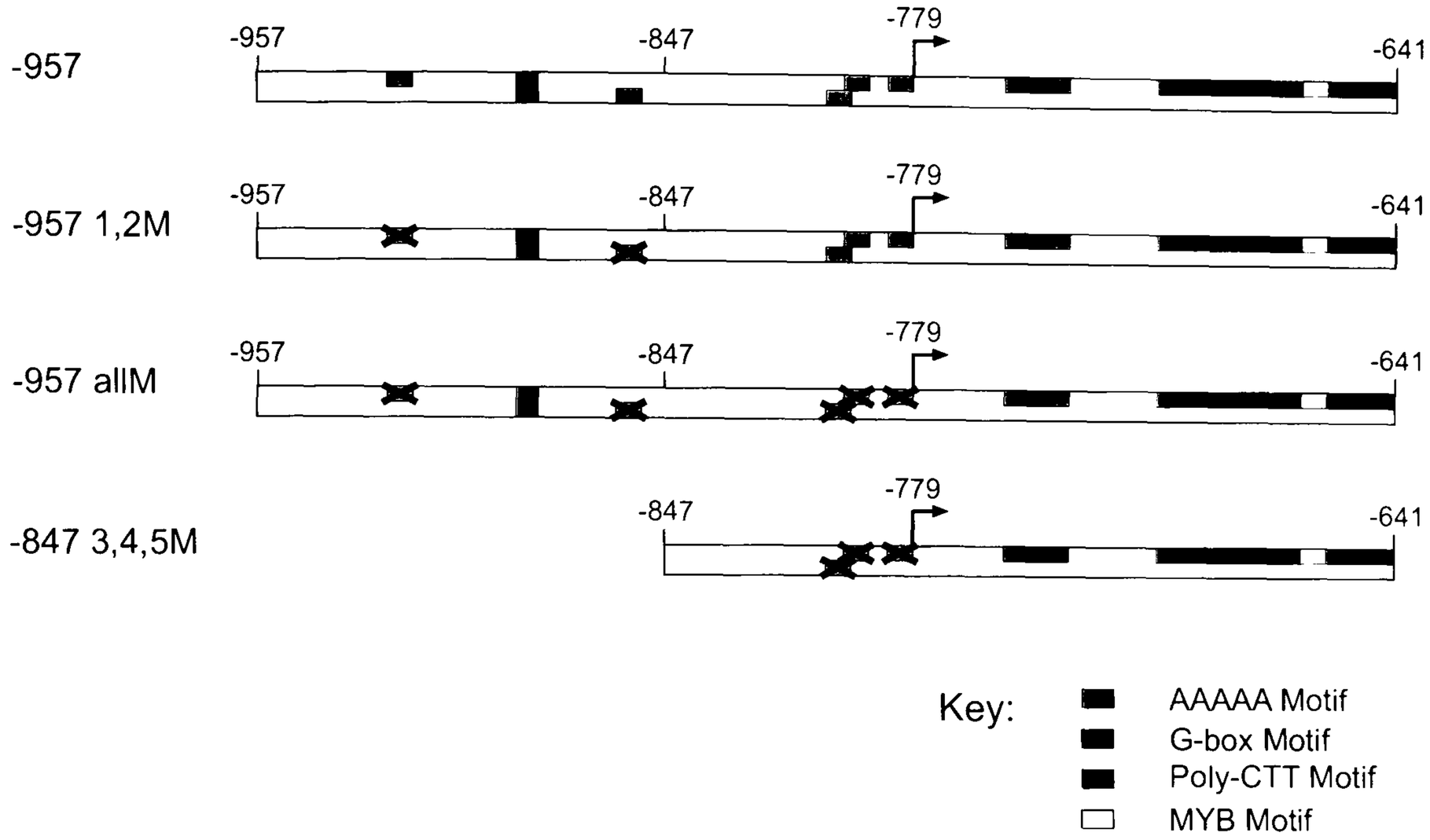


Figure 5.3: Schematic of the *LHY* upstream regions of reporter constructs carrying mutations to the Activity "B"-binding sites. Activity "B"-binding sites are represented by blue boxes and mutation of a site is indicated by a cross.

formants for any of the constructs. However, primary transformants were readily identifiable when screened for luciferase activity. As such, all primary transformants were selected by luciferase activity. Sequencing the DNA flanking the T-DNA borders of the plasmid constructs showed that the pGreen plasmid into which the original -957p*LHY::luc*$^+$ and -847p*LHY::luc*$^+$ constructs had been cloned did not contain the *NptII* gene. Rather, this plasmid contained a tract of exogenous sequence within the ColEI replication origin, which would account for the size of the plasmid approximating that expected if NptII were present. This spontaneous expansion of the pGreen000 plasmid has also been reported by others (see `http://www.pgreen.ac.uk/pGreenII/text2.htm`).

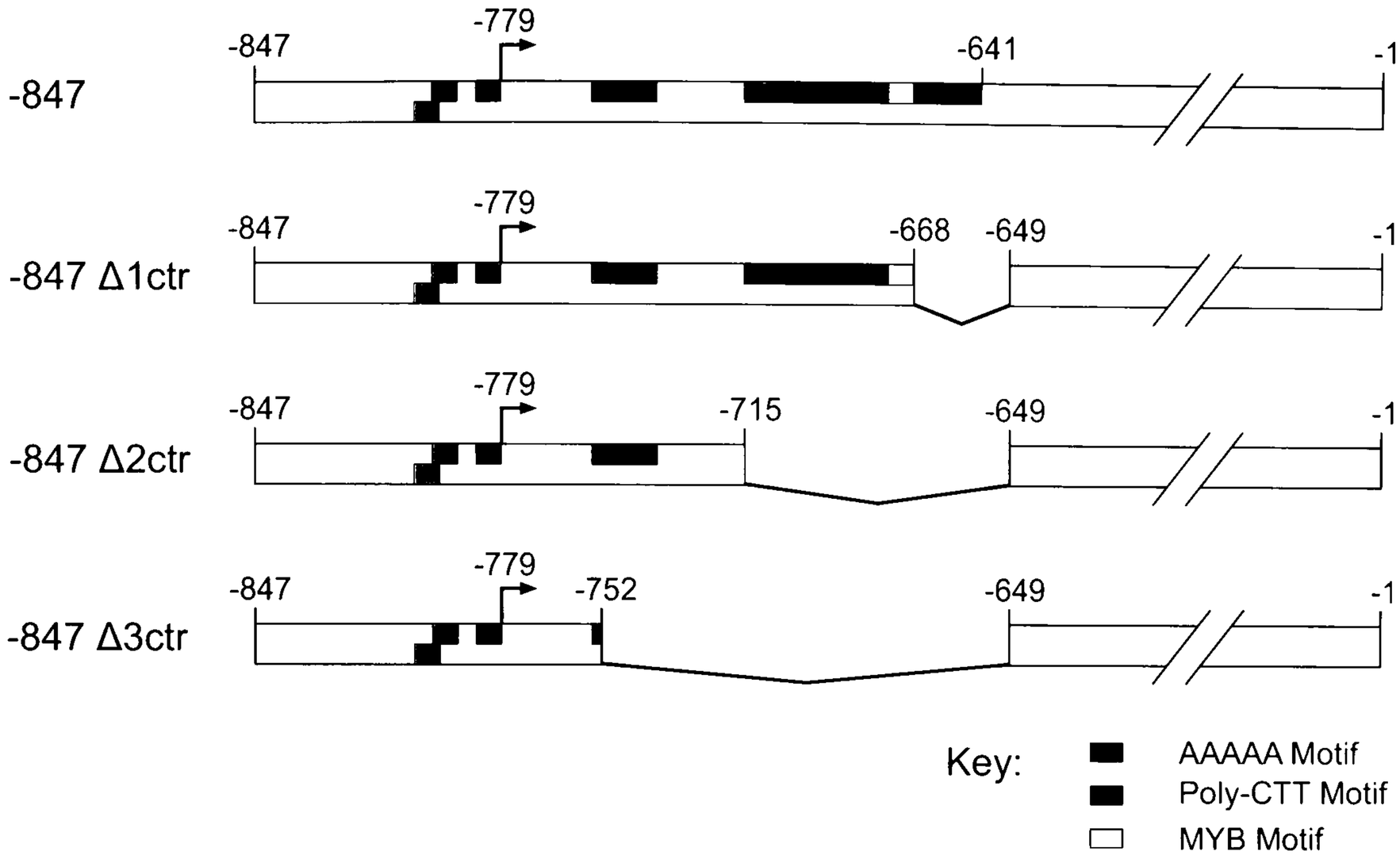


Figure 5.4: Schematic of the *LHY* upstream regions of reporter constructs carrying deletions within the CT-rich region. Regions deleted in each construct are represented by "V" shaped lines linking the remaining regions.

5.3 Functional Analysis of Protein Binding Sites in the -847p*LHY::luc*$^+$ Construct

In Chapter 3, it was shown that the -847p*LHY::luc*$^+$ construct was subject to a rhythmic input, R_A and inputs from both red and blue light signals. Of the candidate transcription factor binding sites identified in Chapter 4, three copies of the activity "B"-binding motif, the poly-CTT motifs and the Myb binding site motif are present in this construct. Any of these sites might therefore mediate the responses of the -847p*LHY::luc*$^+$ construct to R_A, red light signals or blue light signals.

To determine whether the activity "B"-binding sites are involved in mediating any of the various inputs, the expression profile of the -847 3.4,5M construct was examined. As no transgenic lines were recovered for the -857

Δ1ctr and -857 Δ2ctr constructs. the contributions of the poly-CTT motifs and Myb binding site motif to regulation of LHY expression were considered together by examination of the expression profile of the -857 Δ3ctr construct. in which the entire CT-rich region, including the MYB motif and all three poly-CTT motifs had been deleted (Figure 5.4).

The regulation of the -847 3,4,5M and -847 Δ3ctr constructs were compared to the -847 construct in 7 day old seedlings grown under cycles of 12 hours of white light and 12 hours of darkness. The constructs were assayed for driven rhythmicity under a further two 12L 12D cycles of red light. To analyse the circadian regulation of these constructs. plants were then transferred to conditions of constant red light.

However, before these aspects of regulation were considered, the mean expression levels of each reporter construct were determined to ascertain whether these sites might also play a more general role in the regulation of *LHY* expression.

5.3.1 Expression levels of the -847 3,4,5M and -847 Δ3ctr constructs

Expression levels of each construct were determined by calculating the mean expression level of each construct in 7 day old seedlings over 5 days in constant red light after entrainment to 12:12 cycles of white light and darkness. These values were normalized to the mean expression level of the -847p*LHY::luc*$^{+}$ construct.

Mutation of the activity "B"-binding sites or deletion of the CT-rich region both had clear effects on the mean expression levels of the respective reporter constructs under free-running conditions (Figure 5.5). Expression of the -847 3,4.5M construct was reduced to around half that of the -847 construct whereas expression of the -847 Δ3ctr construct was lower still and around one

quarter the level of the -847 construct. The clear effect of these mutations on expression levels confirms a functional role for the activity "B"-binding sites and the CT-rich region.

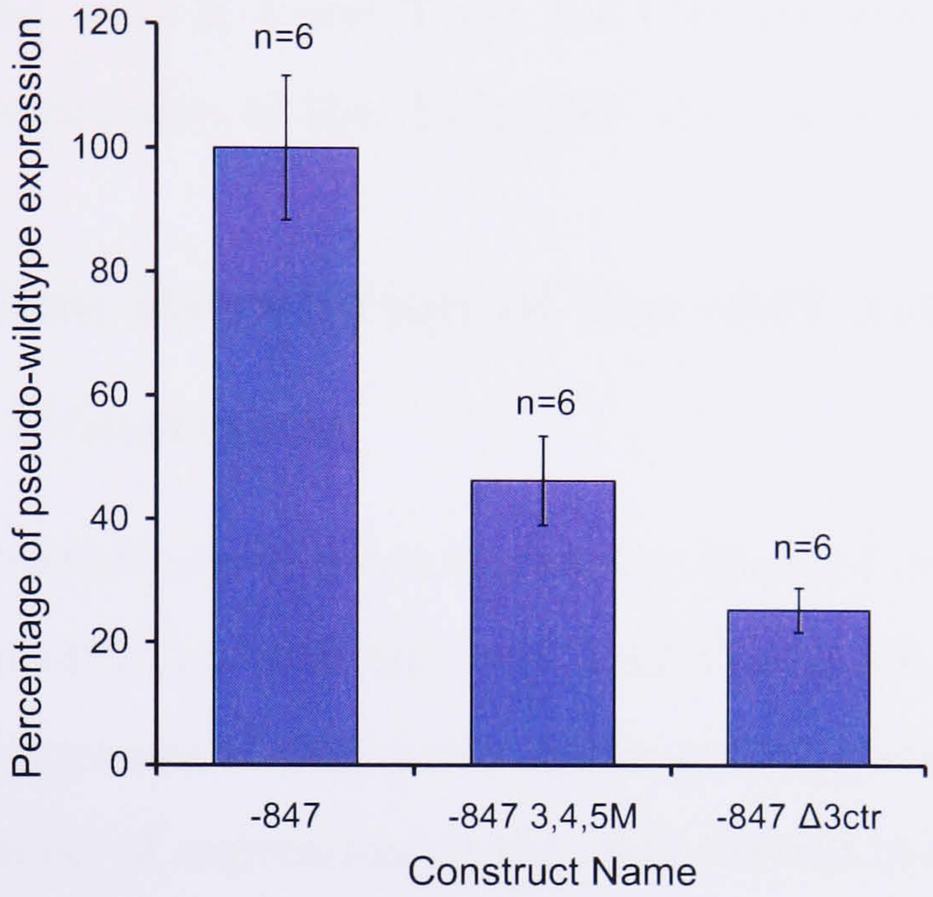


Figure 5.5: Effects on reporter expression levels of mutations to candidate transcription factor-binding sites in the -847p*LHY::luc*$^+$ construct under free-running conditions. Average expression levels were determined for each construct over a period of 120 hours in constant red light and normalized to the mean expression level of the -847p*LHY::luc*$^+$ construct

5.3.2 Diurnal Regulation of the -847 3,4,5M and -847 Δ3ctr constructs

In Chapter 3 it was shown that the induction of expression of the -847p*LHY::luc*$^+$ construct by light at dawn was reduced compared to that of the -957p*LHY::luc*$^+$ construct. Nevertheless, the -847p*LHY::luc*$^+$ construct still showed a degree of induction by light under light dark cycles. To investigate whether either the CT-rich region or activity "B"-binding sites 3, 4 and 5 might account for this aspect of regulation, expression of the -847 3,4,5M and -847 Δ3ctr con-

structs were examined along with that of the -847p*LHY::luc*$^+$ construct under light:dark cycles (Figure 5.6A).

Neither the -847 3.4,5M construct, nor the -847 Δ3ctr showed any further defect in acute induction of expression during dawn, demonstrating that activity "B"-binding sites 3, 4 and 5 and the CT-rich region are not required for this aspect of expression of the -847p*LHY::luc*$^+$ construct.

5.3.3 Circadian Regulation of the -847 3,4,5M and -847 Δ3ctr Constructs

Despite the severe reductions in mean expression levels of these constructs during free-running conditions, both the -847 3,4,5M and -847 Δ3ctr constructs were nevertheless expressed rhythmically under constant conditions and maintained the same phase of expression as the -847 control construct (Figure 5.6 B). It can therefore be concluded that neither the Activity "B"-binding sites, nor the CT-rich region are required for rhythmicity of the -847p*LHY::luc*$^+$ construct.

5.3.4 Acute Light Induction in Etiolated Seedlings

To determine whether either the Activity "B"-binding sites, or elements within the CT-rich region mediate this light response, the acute light induction phenotypes of the -847 3,4,5M and -847 Δ3ctr constructs were assayed alongside the -847p*LHY::luc*$^+$ construct. Figure 5.6 shows the expression profiles of the -847, -847 3,4,5M and -847 Δ3ctr on exposure to one hour of either red light (Figure 5.7 A) or blue light (Figure 5.7 B) at a fluence rate of 10μmol m^{-2} s^{-1}.

The expression profile of the -847 construct was consistent with the observations described in Chapter 3. Following exposure to either red or blue

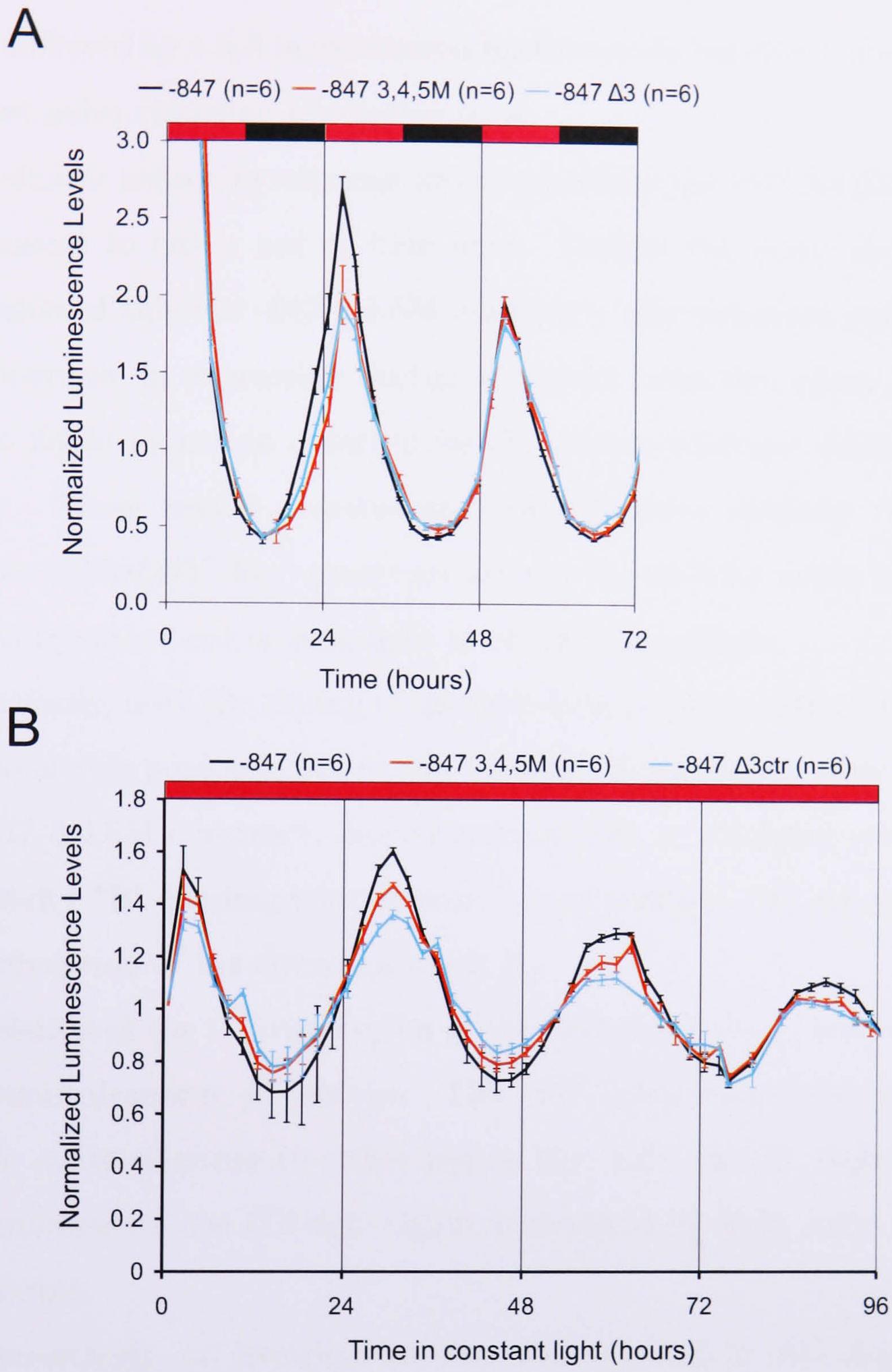


Figure 5.6: The contribution of the CT-rich region and Activity "B"-binding sites 3, 4 and 5 to diurnal and circadian regulation of the -847p*LHY::luc*$^{+}$ construct. Reporter expression was assayed in 7 day old seedlings entrained to 12l 12D cycles under white light before transfered to A) 12L 12D cycles of red light, to assay diurnal regulation, or B) constant red light, to assay circadian regulation. Data are averages of n independent transgenic lines normalized to the mean level of expression of each line over the course of the experiment. Error bars are standard errors of the means

light, a tripartite response was observed, composed of acute induction of expression, followed by a fall in expression back towards baseline before the onset of the post-pulse circadian expression peak.

A similar tripartite response was observed for the -847 3.4.5M construct after exposure to either red or blue light. Despite the lower absolute level of expression of this the -847 3.4.5M construct, this construct still displayed a 2-fold increase in expression during the acute induction phase, similar to the 1.5 to 2-fold induction observed for the pseudo-wildtype -847p*LHY::luc*$^+$ construct. These results demonstrate that the three Activity "B"-binding sites in the -847p*LHY::luc*$^+$ construct are not required for acute induction of expression by either red or blue light in etiolated seedlings.

Similarly, both the timing of the post-light pulse circadian peak and the amplitude of this peak relative to the baseline of expression, were unaffected in the -847 3,4,5M construct, demonstrating that, in etiolated seedlings, the three Activity "B"-binding sites downstream of position -847 are not required for the perception of the circadian input R_A.

Deletion of the CT-rich region of the -847p*LHY::luc*$^+$, however, caused a much more dramatic phenotype. The -847 Δ3ctr construct exhibited no detectable acute response to either red or blue light signals, suggesting that some component of the CT-rich region is required for light induction of the -847 construct.

Interestingly, no circadian rhythm was detected in -847 Δ3ctr expression either before or after the light pulse indicating that the CT-rich region of the *LHY* promoter is required to mediate rhythmic expression of the -847p*LHY::luc*$^+$ construct in etiolated seedlings. This is unlikely to be an artefact of assay sensitivity since the assay was sufficiently sensitive to detect changes in expression of the -847 3.4.5M construct, which was expressed at similar basal levels prior to the light pulse thus indicating that the CT-rich

region of the *LHY* promoter is required to mediate rhythmic expression of the -847p*LHY::luc*$^+$ construct in etiolated seedlings.

5.4 Functional Analysis of Protein Binding Sites in the -957 to -847 Region.

The region of the *LHY* promoter between positions -957 and -847 upstream of the translational start site was shown in Chapter 3 to be important in setting the phase of *LHY* expression. The -847p*LHY::luc*$^+$ construct, which lacked this region, was expressed with an early phase of onset under constant light and a late peak phase under constant darkness. Moreover, in Chapter 3 it was shown that these effects might be attributable to a second rhythmic input to *LHY*, R_B , acting through the -957 to -847 region.

Analysis of the -957 to -847 region by electrophoretic mobility shift assays identified three candidate transcription factor binding sites that may mediate the action of R_B: a G-box motif and two copies of the activity "B"-binding site. It was also found that the contribution of R_B to the timing of reporter gene expression was most evident after entrainment to short day photoperiods.

5.4.1 The Contribution of the G-box to Expression Levels of the -957p*LHY::luc*$^+$ Construct

As discussed above, a mutation to the core hexamer of the G-box motif in the -957p*LHY::luc*$^+$ construct was found to drastically reduce expression levels of the reporter gene to undetectable levels. To determine whether mutation of the two basepair regions flanking the core hexamer had a lesser effect, the expression levels of constructs in which the *LHY* G-box was mutated to Strong

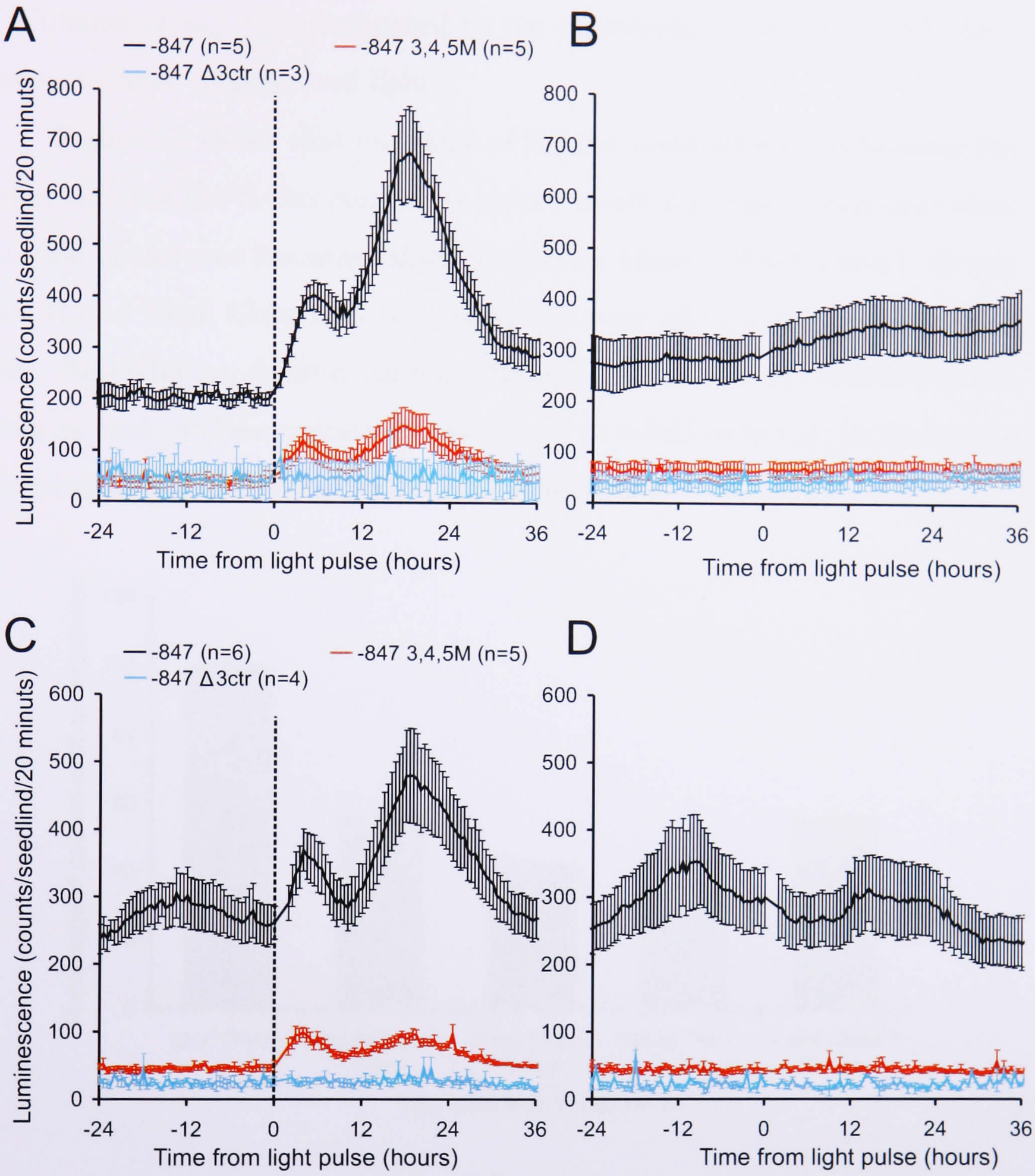


Figure 5.7: The contribution of the CT-rich region and Activity "B"-binding sites 3, 4 and 5 to acute light induction of the -847p*LHY::luc*$^+$ construct in etiolated seedlings. For each line, 30 to 60 seeds were sown in collars on MS3 agar. Seeds were stratified for 7 days in the dark at 4 °C before exposure to cool white light (90μmol/m^2/s) for 2 hours to encourage germination. Plants were grown for 7 days in darkness before transfer to the imaging chamber. After 36 hours, plants were given a 1 hour flash of either red light (A) or blue light (C) at 10μmol/m^2/s. Dark controls for each experiment are shown in B and D. Data are averages of several independent lines. Error bars represent standard errors of the means.

Class I, Weak Class I, Strong Class II or Weak Class II G-box (as defined by Williams *et al.*) were compared to the expression of the -957p*LHY::luc*$^{+}$ construct under constant red light.

Figure 5.8 shows that mutation of the two basepair regions flanking the core hexamer of the G-box motif does indeed result in a milder phenotype than mutation of the core hexamer, since the Strong Class I, Weak Class I, Strong Class II and Weak Class II G-box constructs were all expressed at detectable levels. Nevertheless, these mutations still appear to affect the function of the G-box as each of these mutations caused a 2 to 3-fold reduction in expression compared to the -957p*LHY::luc*$^{+}$ control construct.

Figure 5.8: The effects of the G-box mutations on reporter expression levels. Average expression levels were determined for each construct over a period of 120 hours in constant red light and normalized to the mean expression level of the -957p*LHY::luc*$_{+}$ construct

5.4.2 The Contribution of the G-box to Diurnal Regulation of the -957p*LHY::luc*$^+$ Construct

To determine whether the G-box motif contributes to the correct timing of *LHY* expression, the circadian expression profiles of the four G-box mutant constructs were first compared to the expression profile of the -957 construct under 12L 12D photocycles.

Under light:dark cycles, all G-box mutant constructs were expressed rhythmically, as expected (Figure 5.9 A and B). Under these conditions, the expression profiles of the two Class I G-box constructs was indistinguishable from the profile of the -957p*LHY::luc*$^+$ construct. However, both Class II G-box constructs recapitulated the phenotype of the -847 construct under light:dark cycles, displaying a reduced amplitude of light induction at dawn. These results suggest that the G-box motif is a target of light signalling to the *LHY* promoter which is functionally unaffected by the Class I mutations but is disrupted by the Class II mutations.

5.4.3 The Contribution of the G-box to Circadian Regulation of the -957p*LHY::luc*$^+$ Construct

To determine whether the G-box motif might mediate the action of R_B, the expression profile of the Strong Class II G-box mutant was examined in constant red light, after entrainment to short day (8L 16D) light:dark cycles (Figure 5.10). If the Strong Class II G-box mutation pevents the action of R_B on the *LHY* promoter, one would expect the Strong Class II G-box construct to display the same early phase of expression as the -847p*LHY::luc*$^+$ construct under these conditions. However, in these experiments, the Strong Class II G-box construct was expressed with similar timing to the -957p*LHY::luc*$^+$ control construct, suggesting that mutation of the G-box motif from the wildtype *LHY*

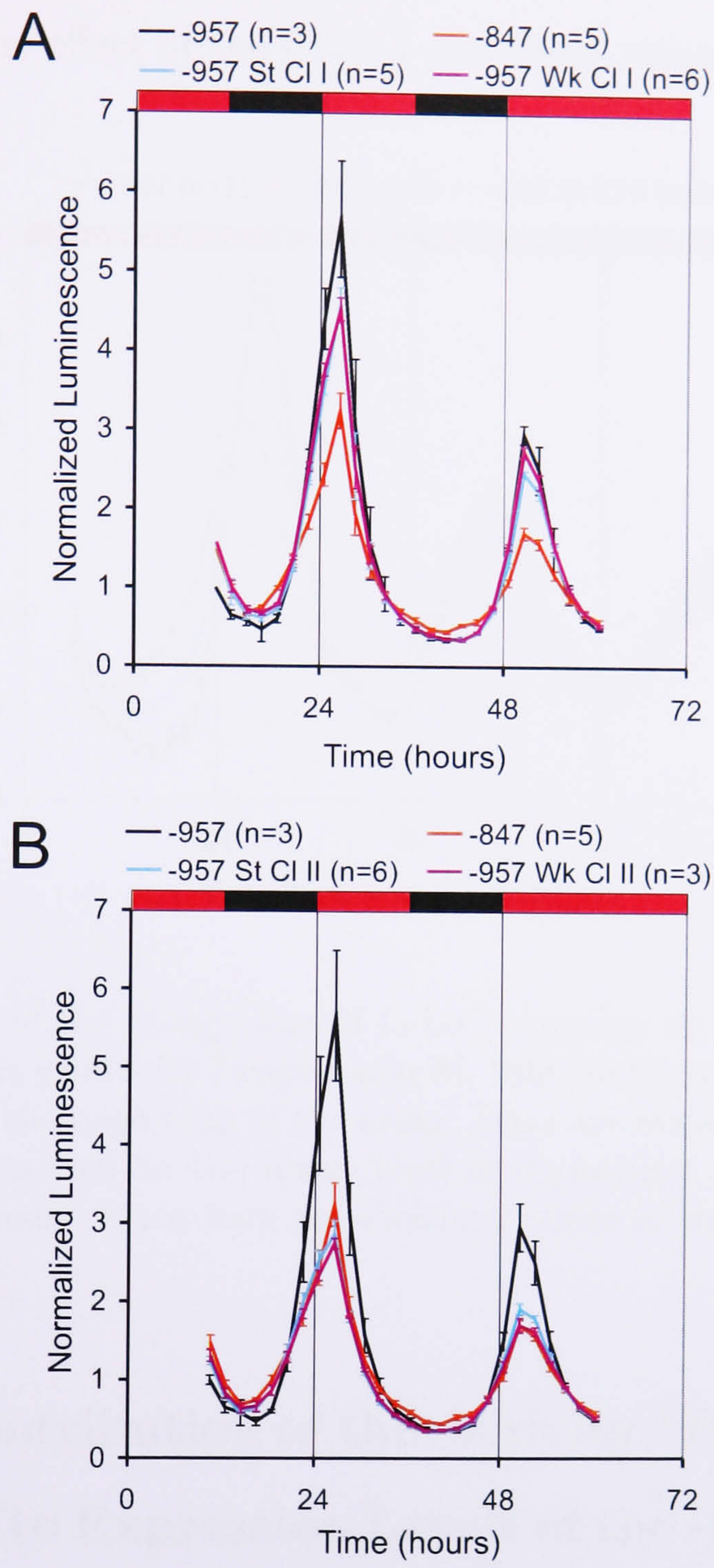


Figure 5.9: Effects of the G-box mutations on diurnal regulation. Seedlings were grown as described for Figure 5.6 before imaging. A) Effects of the Class I mutations. B) Effects of the Class II mutations. Data are averages of n transgenic lines normaized to the average expression level of each line over the course of the experiment. Error bars are standard errors of the means.

sequence to the Strong Class II sequence does not affect the perception of R_B by the reporter. The effect of the Class I mutations remains to be tested.

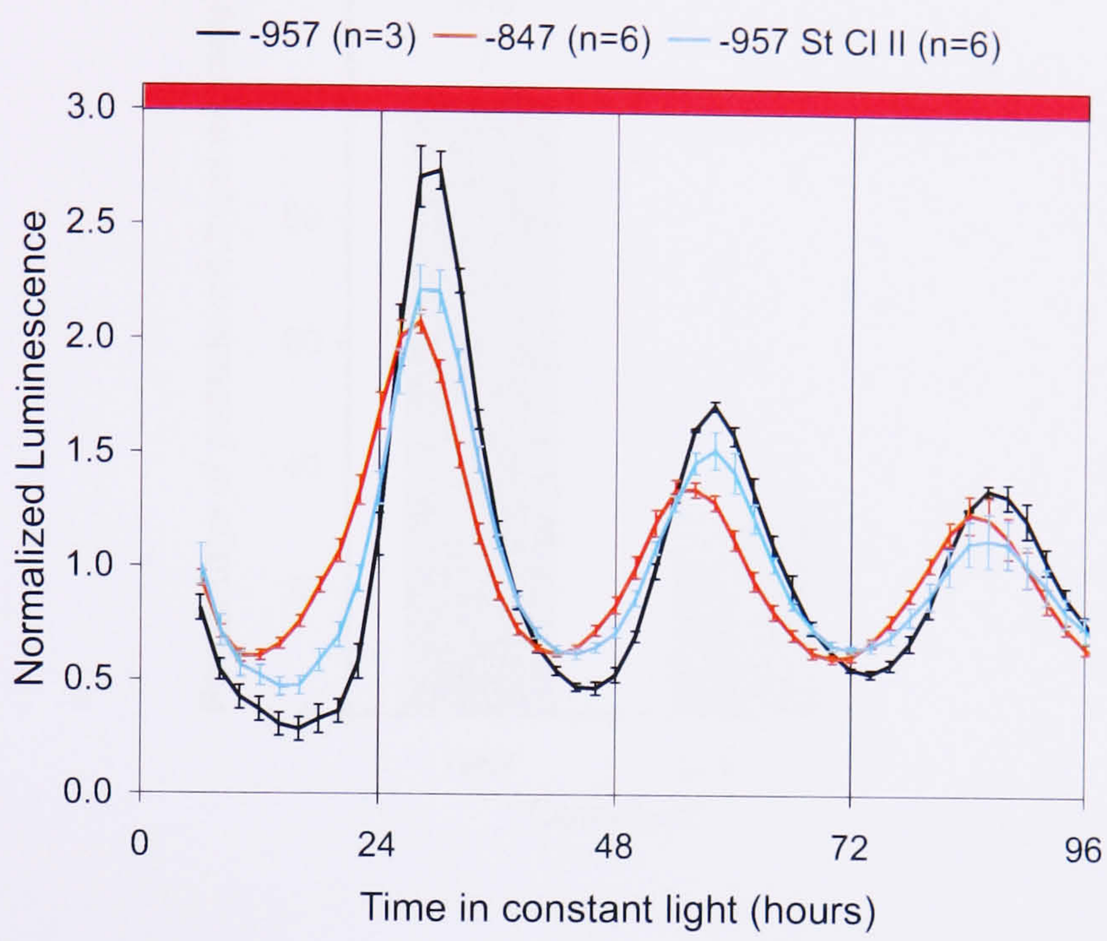


Figure 5.10: Effect of the Stong Class II G-box mutation on the phase of reporter expression. Plant were grown for 7 days under 8L 16D photocycles before release into constant red light at the beginning of the assay. Data are averages of n independent transgenic lines normalized to the mean level of expression of each line over the course of the experiment. Error bars are standard errors of the means.

5.4.4 The Contribution of the Activity "B"-binding Sites 1 and 2 to Expression Levels of the -957p*LHY::luc*$^+$ Construct

Since mutation of the three activity "B"-binding sites in the -847 construct caused a 2-fold reduction in expression levels, the effect of mutation of the two activity "B"-binding sites in the -957 to -847 region was considered. Mutation of these sites in the -957 1,2M construct produced a similar reduction in expression, reducing expression to around one third the levels of the -957 construct, demonstrating that these two sites also play some role in regulating

the expression of *LHY* (Figure 5.11).

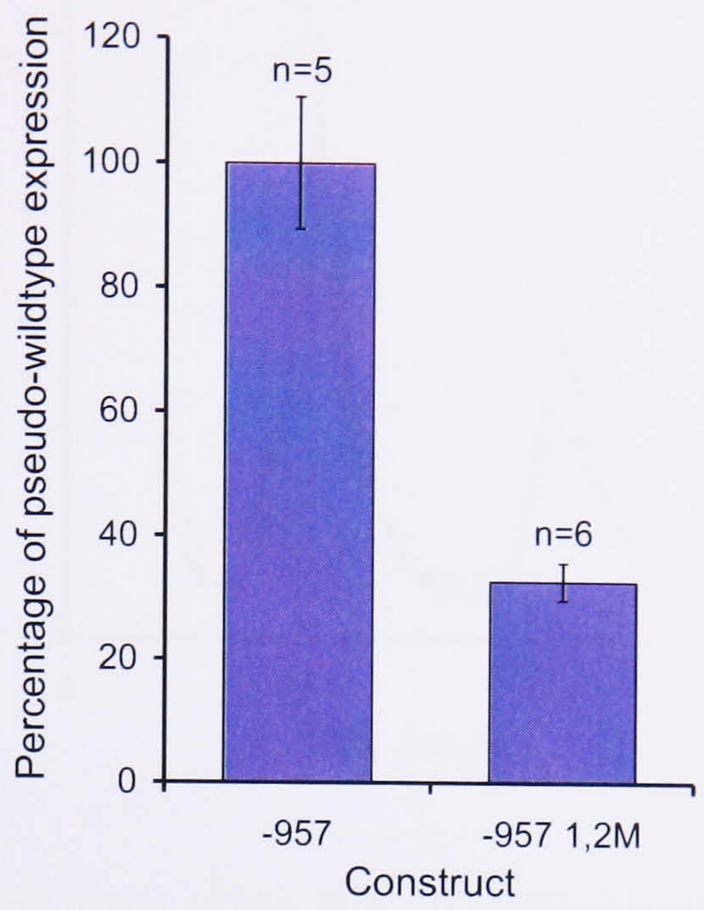


Figure 5.11: The effects mutation of activity "B"-binding sites 1 and 2 on reporter expression levels. Average expression levels were determined for each construct over a period of 120 hours in constant red light and normalized to the mean expression level of the -957p*LHY::luc*$_+$ construct.

5.4.5 The Contribution of the Activity "B"-binding Sites 1 and 2 to Diurnal Regulation of the -957p*LHY::luc*$^+$ Construct

Under light dark cycles, the -957 1,2M construct displayed a similar phenotype to both the -847p*LHY::luc*$^+$ construct and the Class II G-box mutant constructs where the relative amplitude of the expression peak following dawn was reduced to around two thirds that of the -957p*LHY::luc*$^+$ construct, suggesting that along with the G-box motif, activity "B"-binding sites 1 and 2 are also implicated in the induction of *LHY* expression in response to dawn (Figure 5.12).

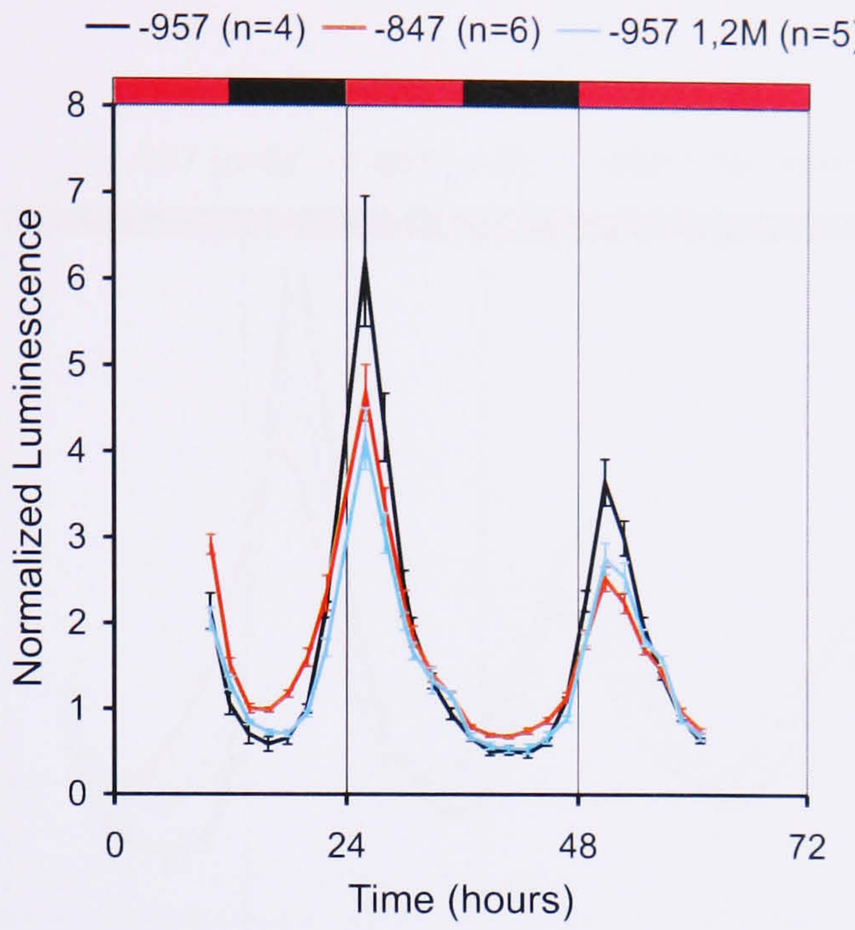


Figure 5.12: The Contribution of the Activity "B"-binding Sites 1 and 2 to Diurnal Regulation of the -957p*LHY::luc*$^+$ Construct. Seedlings were grown as described for Figure 5.6 before imaging. Data are averages of n transgenic lines normalized to the average expression level of each line over the course of the experiment. Error bars are standard errors of the means.

5.4.6 The Contribution of the activity "B"-binding Sites 1 and 2 to Circadian Regulation of the -957p*LHY::luc*$^+$ Construct

Since the Strong Class II G-box mutation did not affect the timing of reporter expression, we asked whether activity "B"-binding sites 1 and 2 might play a role in determining the timing of *LHY* expression in free-running conditions. Again, expression was analysed in constant light after entrainment to short day photocycles of 8 hours of light and 16 hours of darkness per cycle (Figure 5.13). Under these conditions, the -957 1,2M construct was expressed with a similar phase to the -957p*LHY::luc*$^+$ construct, failing to reproduce the early phase phenotype of the -847p*LHY::luc*$^+$ construct and demonstrating that activity "B"-binding sites 1 and 2 are not required for regulation of *LHY* expression

by R_B.

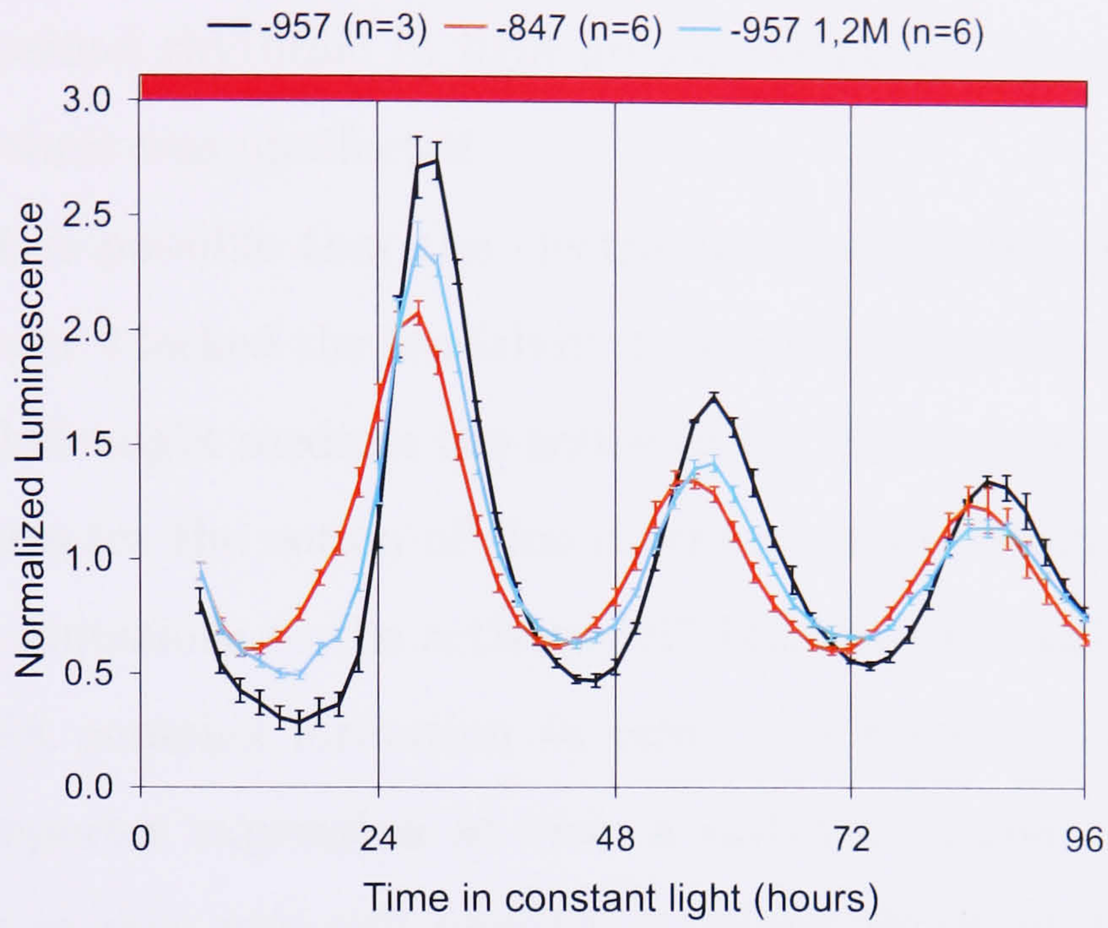


Figure 5.13: The Contribution of the activity "B"-binding Sites 1 and 2 to the phase of reporter expression. Plant were grown for 7 days under 8L 16D photocycles before release into constant red light at the beginning of the assay. Data are averages of n independent transgenic lines normalized to the mean level of expression of each line over the course of the experiment. Error bars are standard errors of the means.

5.5 Discussion

The results presented in this chapter demonstrate that the protein binding sites identified within the upstream sequence of *LHY* in Chapter 4 are indeed functional in the regulation of *LHY* expression, since mutations to each of these sites cause various degrees of reduction in the overall levels of reporter expression.

None of the binding sites identified in Chapter 4 could be assigned a function in mediating either of the two rhythmic inputs to the *LHY* promoter, R_A and R_B. Mutation of the three activity "B"-binding sites in the -847p*LHY::luc*$^+$ construct or deletion of the entire CT-rich region of this con-

struct caused a notable decrease in overall expression levels of the construct but the -847 3,4,5M and -847 Δ3ctr construct, however the expression of these constructs remained rhythmic in light-grown plants and the timing of these expression rhythms was unaffected.

While it is possible that the electrophoretic mobility shift assays employed in Chapter 4 lacked the sensitivity to identify other transcription factor binding sites that might mediate the action of R_A, there are several other possible mechanisms for the action of this rhythmic input. Firstly, it is possible that while the mutations to the activity-"B" binding sites had a strong effect on protein-DNA complex formation *in vitro*, and these mutations caused a reduction in reporter expression *in vivo*, a sufficient degree of transcription factor binding *in vivo* was still possible to direct rhythmic transcription of the reporter constructs. It is also possible that either the CT-rich region and the activity "B" binding sites are able to mediate rhythmic input to the -847p*LHY::luc*$^+$ promoter redundantly or that R_A is in fact composed of two rhythmic inputs, such that loss of either motif alone is insufficient to abolish rhythmicity of reporter expression. Without a reporter construct in which the activity "B"-binding sites were mutated and the CT-rich region deleted, this possibility cannot be ruled out. A final possibility is that the loss of rhythmic transcription is masked by rhythmic translation of the *luciferase* mRNA. Again, this possibility cannot be discounted without further data, examining the expression of luciferase mRNA.

A similar argument can be made regarding regulation of *LHY* by R_B. Neither mutation of the G-box motif nor the two activity "B"-binding sites between positions -957 and -847 was able to reproduce the early phase of expression of the -847p*LHY::luc*$^+$ construct. Again, this could reflect the choice of mutations being inadequate to abolish rhythmicity or the failure to identify other transcription factor binding sites within this region responsible for

regulation by R_B. One interesting possibility is that transcription factor complexes binding the G-box and these two Activity-"B"-binding sites physically interact. Evidence for such an interaction *in vitro* was identified in Figure 5.1. In order to rule this out, a reporter construct incorporating mutations to both the G-box and the two activity "B"-binding sites in this region would be required.

The identification of *cis*-regulatory sequences mediating light input to the *LHY* promoter was more successful. Under light:dark cycles, certain mutations of the G-box motif reduced the light induction of the reporter construct to the same degree as seen for the -847p*LHY::luc*$^+$ construct. This observation is consistent with previous reports that the G-box motif is a target of light signals (Martínez-García *et al.*, 2000). Interestingly, mutation of Activity "B"-binding sites 1 and 2 had the same effect on light induction at dawn as mutation of the G-box motif, suggesting once again a fundamental interaction between these motifs. However, mutation of Activity "B"-binding sites 3, 4 and 5 in the context of the -847p*LHY::luc*$^+$ construct caused no further reduction in the light induction of this construct at dawn, suggesting these sites do not mediate light induction of *LHY* in the absence of a G-box.

In etiolated seedlings, the regulation of *LHY* expression appears to be different than in light-grown seedlings. Despite the finding that the CT-rich region of the *LHY* promoter is not required for diurnal light induction in light-grown plants, in etiolated seedlings this region is required for acute induction of the -847p*LHY::luc*$^+$ construct by both red and blue light pulses. There are many examples of other genes that show differential regulation of light responsive genes between light grown and etiolated seedlings (Ma *et al.*, 2001), so it is likely that there is a degree of plasticity in light signalling networks during different stages of seedling development.

The finding that the -847 Δ3ctr construct showed no acute induction

in response to light signals demonstrated that this sequence motif contains a *cis*-regulatory element that confers light responsiveness in etiolated seedlings. It is not possible to say whether it is the CTT repeat motifs or the candidate MYB binding site which constitutes this light responsive element. It is worth noting, however, that at least one MYB transcription factor, LAF1, has been implicated in light signalling in *Arabidopsis* , though this function has only been tested in far red light (Ballasteros *et al.*, 2001).

Since the -847 Δ3ctr construct did not show a circadian peak of expression after the application of a light pulse, which resets and promotes high amplitude oscillations of the endogenous oscillator mechanism, it can also be concluded that, in etiolated seedlings, the CT-rich region of the *LHY* promoter is also a target of a circadian signal from the circadian oscillator mechanism to the *LHY* promoter. This evidence suggests that, like regulation of *LHY* by light signals, the regulation of *LHY* by the circadian clock also differs between light grown and etiolated seedlings. Since *LHY* itself encodes a clock component, this may suggest that the circadian oscillator network is configured differently during different stages of seedling development, perhaps comprising a smaller set of components in etiolated plants and fewer regulatory feedback loops.

Bibliography

Akashi, M., Tsuchiya, Y., Yoshino, T. & Nishida, E., 2002 "Control of intracellular dynamics of mammalian period proteins by casein kinase I ϵ (CKIϵ) and CKIδ in cultured cells," *Mol Cell Biol* **22**, 1693–1703.

Alabadí, D., Oyama, T., Yanovsky, M., Harmon, F., Más, P. & Kay, S., 2001 "Reciprocal Regulation Between *TOC1* and *LHY/CCA1* Within the *Arabidopsis* Circadian Clock," *Science* **293**, 880–883.

Alabadí, D., Yanovsky, M., Más, P., Harmer, S. & Kay, S., 2002 "Critical Role for CCA1 and LHY in Maintaining Circadian Rhythmicity in *Arabidopsis*," *Curr Biol* **12**, 757–761.

Albrecht, U., Sun, Z. S., Eichele, G. & Lee, C. C., 1997 "A differential response of two putative mammalian circadian regulators, mper1 and mper2, to light," *Cell* **91**, 1055–1064.

Allada, R., White, N. E., So, W. V., Hall, J. C. & Rosbash, M., 1998 "A mutant Drosophila homolog of mammalian Clock disrupts circadian rhythms and transcription of period and timeless," *Cell* **93**, 791–804.

Aronson, B. D., Johnson, K. A., Loros, J. J. & Dunlap, J. C., 1994 "Negative Feedback Defining a Circadian Clock: Autoregulation of the Clock Gene frequency," *Science* **263**, 1578–1584.

Ballasteros, M., Bolle, C., Lois, L., Moore, J., Vielle-Calzada, J.-P., Grossniklaus, U. & Chua, N., 2001 "LAF1, a MYB transcription activator for phytochrome A signalling." *Genes Dev.* **15**, 2613–2625.

Barnes, J. W., Tischkau, S. A., Barnes, J. A., Mitchell, J. W., Burgoon, P. W., Hickok, J. R. & Gillette, M. U., 2003 "Requirement of Mammalian *Timeless* for Circadian Rhythmicity." *Science* **302**, 439–442.

Bell-Pedersen, D. and Cassone, V.M. and Earnest, D.J. and Golden, S.S. and Hardin, P.E. and Thomas, T.L. and Zoran, M.L., 2005 "Circadian rhythms from multiple oscillators: Lessons from diverse organisms," *Nat. Rev. Genet.* **6**, 544–556.

Bunger, M., Wilsbacher, L., Moran, S., Clendenin, C., Radcliffe, L., Hogenesch, J., Simon, M., Takahashi, J. & Bradfield, C., 2000 "*Mop3* is an essential component of the master circadian pacemaker in mammals," *Cell* **103**, 1009–1017.

Casal, J. J. & Yanovsky, M. J., 2005 "Regulation of gene expression by light," *Int J Dev Biol* **49**, 501–511.

Ceriani, M., Darlington, T., Staknis, D., Más, P., Petti, A., Weitz, C. & Kay, S., 1999 "Light-dependent sequestration of TIMELESS by CRYPTOCHROME," *Science* **285**, 553–556.

Cheng, P., He, Q., He, Q., Wang, L. & Liu, Y., 2005 "Regulation of the *Neurospora* circadian clock by an RNA helicase," *Genes Dev.* **19**, 234–241.

Cheng, P., Yang, Y., Gardner, K., Liu, P., Y.Cheng, Yang, Y., Gardner, K. & Liu, Y., 2002 "PAS domain-mediated WC-1/WC-2 interaction is essential for maintaining the steady-state level of WC-1 and

the function of both proteins in circadian clock and light responses of Neurospora." *Mol Cell Biol* **22**, 517–524.

CHENG, P., YANG, Y. & LIU, Y., 2001 "Interlocked feed-back loops contribute to the robustness of the *Neurospora* circadian clock," *Proc Nat Acad Sci USA* **98**, 7408–7413.

CLOUGH, S. J. & BENT, A. F., 1998 "Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*," *Plant J.* **16**, 735–743.

CORREA, A., LEWIS, Z., GREENE, A., MARCH, I., GOMER, R. & BELL-PEDERSEN, D., 2003 "Multiple oscillators regulate circadian gene expression in *Neurospora*." *Proc Nat Acad Sci USA* **100**, 13597–13602.

CROSTHWAITE, S., LOROS, J. & DUNLAP, J., 1995 "Light-induced resetting of a circadian clock is mediated by a rapid increase in *frequency* transcript," *Cell* **81**, 1003–1012.

CURTIN, K. D., HUANG, Z. J. & ROSBASH, M., 1995 "Temporally regulated nuclear entry of the *Drosophila* period protein contributes to the circadian clock," *Neuron* **14**, 365–372.

CYRAN, S. A., BUCHSBAUM, A. M., REDDY, K. L., LIN, M. C., GLOSSOP, N. R., HARDIN, P. E., YOUNG, M. W., STORTI, R. V. & BLAU, J., 2003 "vrille, Pdp1, and dClock form a second feedback loop in the *Drosophila* circadian clock," *Cell* **112**, 329–341.

DAAN, S. & PITTENDRIGH, C., 1978 "Two Coupled Oscillators: Simulation of the Circadian Pacemaker in Mammalian Activity Rhythms," *J Theor Biol* **70**, 297–313.

DANIEL, X., SUGANO, S. & TOBIN, E. M., 2004 "CK2 phosphorylation of CCA1 is necessary for its circadian oscillator function in *Arabidopsis*," *Proc Natl Acad Sci U S A* **101**, 3292–3297.

DARLINGTON, T. K., WAGER-SMITH, K., CERIANI, M., STAKNIS, D., GEKAKIS, N., STEEVES, T., WEITZ, C., TAKAHASHI, J. & KAY, S., 1998 "Closing the circadian loop: CLOCK-induced transcription of its own inhibitors PER and TIM," *Science* **280**, 1599–1603.

DAVIS, S. J., 2002 "Photoperiodism: the coincidental perception of the season," *Curr Biol* **12**, R841–R843.

DODD, A. N., SALATHIA, N., HALL, A., KVEI, E., TTH, R., NAGY, F., HIBBERD, J. M., MILLAR, A. J. & WEBB, A. A., 2005 "Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage," *Science* **309**, 630–633.

DOYLE, M. R., DAVIS, S. J., BASTOW, R. M., MCWATTERS, H. G., KOZMA-BOGNÁR, L., NAGY, F., MILLAR, A. J. & AMASINO, R. M., 2002 "The *ELF4* gene controls circadian rhythms amd flowering time in *Arabidopsis thaliana*," *Nature* **419**, 74–77.

EDWARDS, K. D., ANDERSON, P. E., HALL, A., SALATHIA, N. S., LOCKE, J. C., LYNN, J. R., STRAUME, M., SMITH, J. Q. & MILLAR, A. J., 2006 "FLOWERING LOCUS C mediates natural variation in the high-temperature response of the Arabidopsis circadian clock," *Plant Cell* **18**, 639–650.

EIDE, E., VIELHABER, E., HINZ, W. A. & VIRSHUP, D., 2002 "The circadian regulatory proteins BMAL1 and Cryptochromes are substrates of casein kinase Ie," *J. Biol. Chem.* **277**, 17248–17254.

EMERY, P., SO, W., KANEKO, M., HALL, J. & ROSBASH, M., 1998 "CRY, a *Drosophila* clock and light-regulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity." *Cell* **95**, 669-679.

ERIKSSON, M. E., HANANO, S., SOUTHERN, M. M., HALL, A. & MILLAR, A. J., 2003 "Response regulator homologues have complementary, light-dependent functions in the Arabidopsis circadian clock," *Planta* **218**, 159 162.

FARRÉ, E. M., HARMER, S. L., HARMON, F. G., YANOVSKY, M. J. & KAY, S. A., 2005 "Overlapping and distinct roles of *PRR7* and *PRR9* in the *Arabidopsis* circadian clock," *Curr Biol* **15**, 47 54.

FROEHLICH, A., LIU, Y., LOROS, J. & DUNLAP, J., 2002 "White Collar-1, a circadian blue light photoreceptor, binding to the *frequency* promoter." *Science* **297**, 815–819.

FROEHLICH, A., LOROS, J. & DUNLAP, J., 2003 "Rhythmic binding of a WHITE COLLAR-containing complex to the *frequency* promoter is inhibited by FREQUENCY." *Proc Nat Acad Sci USA* **100**, 5914–5919.

GALLEGO, M. & VIRSHUP, D. M., 2007 "Post-translational modifications regulate the ticking of the circadian clock," *Nat Rev Mol Cell Biol* **8**, 139–148.

GARCEAU, N., LIU, Y., LOROS, J. & DUNLAP, J., 1997 "Alternative initiation of translation and time-specific phosphorylation yield multiple forms of the essential clock protein FREQUENCY," *Cell* **89**, 469–476.

GLOSSOP, N. R., HOUL, J. H., ZHENG, H., NG, F. S., DUDEK, S. M. & HARDIN, P. E., 2003 "VRILLE feeds back to control circadian transcription of Clock in the *Drosophila* circadian oscillator," *Neuron* **37**, 249–261.

GREEN, R. M. & TOBIN, E. M., 1999 "Loss of the circadian clock-associated protein 1 in Arabidopsis results in altered clock-regulated gene expression." *Proc Natl Acad Sci U S A* **96**, 4176–4179.

GRIMA, B., CHLOT, E., XIA, R. & ROUYER, F., 2004 "Morning and evening peaks of activity rely on different clock neurons of the Drosophila brain." *Nature* **431**, 869–873.

HALL, A., BASTOW, R. M., DAVIS, S. J., HANANO, S., MCWATTERS, H. G., HIBBERD, V., DOYLE, M. R., SUNG, S., HALLIDAY, K. J., AMASINO, R. M. & MILLAR, A. J., 2003 "The TIME FOR COFFEE gene maintains the amplitude and timing of Arabidopsis circadian clocks." *Plant Cell* **15**, 2719–2729.

HANAHAN, D., 1985 in D. Glover, (ed.) *DNA Cloning*, vol. 1 (IRL Press Ltd., London).

HARMER, S. L., HOGENESCH, J. B., STRAUME, M., CHANG, H. S., HAN, B., ZHU, T., WANG, X., KREPS, J. A. & KAY, S. A., 2000 "Orchestrated transcription of key pathways in Arabidopsis by the circadian clock," *Science* **290**, 2110–2113.

HAZEN, S. P., SCHULTZ, T. F., PRUNEDA-PAZ, J. L., BOREVITZ, J. O., ECKER, J. R. & KAY, S. A., 2005 "LUX ARRHYTHMO encodes a Myb domain protein essential for circadian rhythms," *Proc Natl Acad Sci U S A* **102**, 10387–10392.

HE, Q., CHENG, P., YANG, Y., WANG, L., GARDNER, K. & LIU, Y., 2002 "White collar-1, a DNA binding transcription factor and a light sensor." *Science* **297**, 840–843.

HELLENS, R. P., EDWARDS, E. A., LEYLAND, N. R., BEAN, S. & MULLINEAUX, P. M., 2000 "pGreen: a versatile and flexible binary Ti

vector for *Agrobacterium*-mediated plant transformation." *Plant Mol Biol* **42**, 819–832.

HICKS, K. A., MILLAR, A. J., CARRÉ, I. A., SOMERS, D. E., STRAUME, M., MEEKS-WAGNER, D. R. & KAY, S. A., 1996 "Conditional circadian dysfunction of the Arabidopsis early-flowering 3 mutant," *Science* **274**, 790–792.

HIGO, K., UGAWA, Y., IWAMOTO, M. & KORENAGA, T., 1999 "Plant *cis*-acting regulatory DNA elements (PLACE) database." *Nucleic Acids Res* **27**, 297–300.

HÖFGEN, R. & WILLMITZER, L., 1988 "Storage of competent cells for *Agrobacterium* transformation," *Nucleic Acids Res* **16**, 9877.

HOGENESCH, J., GU, Y., JAIN, S. & BRADFIELD, C., 1998 "The basic-helix-loop-helix-PAS orphan MOP3 forms transcriptionally active complexes with circadian and hypoxia factors." *Proc Natl Acad Sci U S A* **95**, 5474–5479.

VAN DER HORST, G., MUIJTJENS, M., KOBAYASHI, K., TAKANO, R., KANNO, S., TAKAO, M., DE WIT, J., VERKERK, A., EKER, A., VAN LEENEN, D., BUIJS, R., BOOTSMA, D., J.H., H. & YASUI, A., 1999 "Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms." *Nature* **398**, 627–630.

HUDSON, M. E. & QUAIL, P. H., 2003 "Identification of promoter motifs involved in the network of phytochrome A-regulated gene expression by combined analysis of genomic sequence and microarray data." *Plant Physiol* **133**, 1605–1616.

INAGAKI, N., HONMA, S., ONO, D., TANAHASHI, Y. & K., H., 2007 "Seperate oscillating cell groups in mouse suprachiasmatic nucleus couple

photoperiodically to the onset and end of daily activity." *Proc Nat Acad Sci USA* **104**, 7664–7669.

Ishiura, M., Kutsuna, S., Aoki, S., Iwasaki, H., Anderson, C. R., Tanabe, A., Golden, S., Johnson, C. & Kondo, T., 1998 "Expression of a gene cluster kaiABC as a circadian feed-back process in cyanobacteria." *Science* **281**, 1519–1523.

Iwasaki, H., Nishiwaki, T., Kitayama, Y., Nakajima, M. & Kondo, T., 2002 "KaiA-stimulated KaiC phosphorylation in circadian timing loops in cyanobacteria," *Proc Nat Acad Sci USA* **99**, 15788–15793.

Kageyama, H., Nishiwaki, T., Nakajima, M., Iwasaki, H., Oyama, T. & Kondo, T., 2006 "Cyanobacterial Circadian Pacemaker: Kai Protein Complex Dynamics in the KaiC Phosphorylation Cycle *In Vitro*." *Mol Cell* **23**, 161–171.

Kaneko, M. & Hall, J. C., 2000 "Neuroanatomy of cells expressing clock genes in Drosophila: transgenic manipulation of the period and timeless genes to mark the perikarya of circadian pacemaker neurons and their projections," *J Comp Neurol* **422**, 66–94.

Katayama, M., Kondo, T., Xiong, J. & Golden, S., 2003 "*ldpA* Encodes an Iron-Sulfur Protein Involved in Light-Dependent Modulation of the Circadian Period in the Cyanobacterium *Synechococcus elongatus* PCC 7942," *J Bacteriol* **185**, 1415–1422.

Khanna, R., Kikis, E. & Quail, P., 2003 "EARLY FLOWERING 4 functions in phytochrome B-regulated seedling de-etiolation." *Plant Physiol* **133**. 1530–1538.

KIKIS, E. A., KHANNA, R. & QUAIL, P. H., 2005 "ELF4 is a phytochrome-regulated component of a negative-feedback loop involving the central oscillator components CCA1 and LHY," *Plant J* **44**, 300–313.

KIM, J.-Y., SONG, H.-R., TAYLOR, B. L. & CARRÉ, I. A., 2003 "Light-regulated translation mediates gated induction of the Arabidopsis clock protein LHY," *EMBO J* **22**, 935–944.

KITAYAMA, Y., IWASAKI, H., NISHIWAKI, T. & KONDO, T., 2003 "KaiB functions as an attenuator of KaiC phosphorylation in the cyanobacterial circadian clock system," *EMBO J* **22**, 2127–2134.

KOIKE, N., HIDA, A., NUMANO, R., HIROSE, M., SAKAKI, Y. & TEI, H., 1998 "Identification of the mammalian homologues of the Drosophila timeless gene, Timeless1," *FEBS Lett* **441**, 427–431.

KONOPKA, R. J. & BENZER, S., 1971 "Clock mutants of Drosophila melanogaster," *Proc Natl Acad Sci U S A* **68**, 2112–2116.

KRISHNAN, B., LEVINE, J., LYNCH, M., DOWSE, H., FUNES, P., HALL, J., HARDIN, P. & DRYER, S., 2001 "A new role for cryptochrome in a *Drosophila* circadian oscillator," *Nature* **130**, 313–317.

LEE, K., LOROS, J. & DUNLAP, J., 2000 "Interconnected feed-back loops in the Neurospora circadian system," *Science* **289**, 107–110.

LESCOT, M., DÉHAIS, P., THIJS, G., MARCHAL, K., MOREAU, Y., VAN DE PEER, Y., ROUZÉ, P. & ROMBAUTS, S., 2001 "PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences," *Nucleic Acids Res* **30**, 325–327.

LIN, F., SONG, W., MEYER-BERNSTEIN, E., NAIDOO, N. & SEHGAL,

A., 2001 "Photic signaling by cryptochrome in the *Drosophila* circadian system," *Mol Cell Biol* **21**, 7287–7294.

Liu, X., Covington, M., Fankhauser, C., Chory, J. & Wagner, D., 2001 "*ELF3* Encodes a Circadian Clock-Regulated Nuclear Protein That Functions in an Arabdopsis PHYB Signal Transduction Pathway," *Plant Cell* **13**, 1293–1304.

Liu, Y. & Bell-pedersen, D., 2006 "Circadian rhythms in *Neurospora crassa* and other filamentous fungi." *Eukaryotic Cell* **5**, 1184 1193.

Locke, J., Kozma-Bognár, L., Gould, P., Fehér, B., Kevei, E., Nagy, F., Turner, M., Hall, A. & Millar, A. J., 2006 "Experimental validation of a predicted feedback loop in the multi-oscillator clock of Arabidopsis thaliana," *Mol Syst Biol* **2**, 59.

Locke, J. C., Millar, A. J. & Turner, M. S., 2005a "Modelling genetic networks with noisy and varied experimental data: the circadian clock in Arabidopsis thaliana," *J Theor Biol* **234**, 383–393.

Locke, J. C., Southern, M. M., Kozma-Bognár, L., Hibberd, V., Brown, P. E., Turner, M. S. & Millar, A. J., 2005b "Extension of a genetic network model by iterative experimentation and mathematical analysis," *Mol Syst Biol* **1**, 2005.0013.

Ma, L., Li, J., Qu, L., Hager, J., Chen, Z., Zhao, H. & Deng, X., 2001 "Light control of *Arabidopsis* development entails coordinated regulation of genome expression and cellular pathways." *Plant Cell* **13**, 2589–2607.

Makino, S., Kiba, T., Imamura, A., Hanaki, N., Nakamura, A., Suzuki, T., Taniguchi, M., Ueguchi, C., Sugiyama, T. & Mizuno,

T., 2000 "Genes encoding pseudo-response regulators: insight into His-to-Asp phosphorelay and circadian rhythm in Arabidopsis thaliana." *Plant Cell Physiol* **41**, 791–803.

Makino, S., Matsushika, A., Kojima, M., Yamashino. T. & Mizuno. T., 2002 "The APRR/TOC1 Quintet Implicated in Circadian Rhythms of Arabidopsis I. Characterization with APRR1-Overexpressing Plants." *Plant Cell Physiol* **43**, 59–69.

Martínez-García, J., Huq, E. & Quail, P., 2000 "Direct Targeting of Light Signals to a Promoter-Bound Transcription Factor." *Science* **288**, 859–863.

Más, P., Alabadí, D., Yanovsky, M. J., Oyama, T. & Kay. S. A., 2003a "Dual Role of TOC1 in the Control of Circadian and Photomorphogenic Responses in Arabidopsis," *Plant Cell* **15**, 223–236.

Más, P., Kim, W. Y., Somers, D. E. & Kay, S. A., 2003b "Targeted degradation of TOC1 by ZTL modulates circadian function in Arabidopsis thaliana," *Nature* **426**, 567–570.

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

McClung, C. R., Fox, B. A. & Dunlap, J. C., 1989 "The *Neurospora* clock gene *frequency* shares a sequence element with the *Drosophila* clock gene *period*," *Nature* **339**, 558–562.

McWatters. H., Kolmos. E., Hall. A., Doyle. M., Amasino. R., Gyula. P., Millar, A. J. & Davies. S., 2007 "*ELF3* Is Required for

Oscillatory Properties of the Circadian Clock." *Plant Physiology* **144**, 391-401.

McWatters, H. G., Bastow, R. M., Hall, A. & Millar, A. J., 2000 "The ELF3 *zeitnehmer* regulates light signalling to the circadian clock." *Nature* **408**, 716–720.

Menaker, M., 2003 "CIRCADIAN RHYTHMS: Circadian Photoreception." *Science* **10**, 213–214.

Mignone, F., Gissi, C., Liuni, S. & Pesole, G., 2002 "Untranslated Regions of mRNAs," *Genome Biology* **3**, 0004.1–004.9.

Millar, A. J., Carré, I. A., Strayer, C. A., Chua, N. H. & Kay, S. A., 1995a "Circadian clock mutants in Arabidopsis identified by luciferase imaging," *Science* **267**, 1161–1163.

Millar, A. J., Straume, M., Chory, J., Chua, N. H. & Kay, S. A., 1995b "The regulation of circadian period by phototransduction pathways in Arabidopsis," *Science* **267**, 1163–1166.

Mizoguchi, T., Wheatley, K., Hanzawa, Y., Wright, L., Mizoguchi, M., Song, H.-R., Carré, I. A. & Coupland, G., 2002 "*LHY* and *CCA1* Are Partially Redundant Genes Required to Maintain Circadian Rhythms in *Arabidopsis*," *Developmental Cell* **2**, 629-641.

Mizuno, T., 2005 "Two-component phosphorelay signal transduction systems in plants: from hormone responses to circadian rhythms." *Biosci Biotechnol Biochem* **69**, 2263–2276.

Murakami, M., Yamashino, T. & Mizuno, T., 2004 "Characterization of circadian-associated APRR3 pseudo-response regulator belonging to the

APRR1/TOC1 quintet in Arabidopsis thaliana." *Plant Cell Physiol* **45**, 645–650.

NAKAJIMA, M., IMAI, K., ITO, H., NISHIWAKI, T., MURAYAMA, Y., IWASAKI, H., OYAMA, T. & KONDO, T., 2005 "Reconstitution of Circadian Oscillation of Cyanobacterial KaiC Phosphorylation in Vitro." *Science* **308**, 414–415.

NI, M., TEPPERMAN, J. M. & QUAIL, P. H., 1998 "PIF3, a Phytochrome-Interacting Factor Necessary for Normal Photoinduced Signal Transduction, Is a Novel Basic Helix-Loop-Helix Protein." *Cell* **95**, 657–667.

NISHIWAKI, T., YOSHINORI, S., KITAYAMA, Y., TERAUCHI, K., KIYOHARA, R., TAKAO, T. & KONDO, T., 2007 "A sequential program of dual phosphorylation of KaiC as a basis for circadian rhythm is cyanobacteria," *EMBO J* **26**, 4029–4037.

ONAI, K. & ISHIURA, M., 2005 "PHYTOCLOCK 1 encoding a novel GARP protein essential for the Arabidopsis circadian clock," *Genes Cells* **10**, 963-972.

DE PAULA, R., LEWIS, Z., GREENE, A., SEO, H., MORGAN, L., VITALINI, M., BENNETT, L., GOMER, R. & BELL-PEDERSEN, D., 2006 "Two circadian timing circuits in *Neurospora crassa* share components and regulate distinct rhythmic processes," *J. Biol. Rhythms* **21**, 159–168.

PITTENDRIGH, C., 1979 "Some Functional Aspects of Circadian Pacemakers," in S. Suda, C. Hayaishi & H. Nakagawa, (eds.) *Biological Rhythms and their Central Mechanism* (Elsevier/North-Holland Biomedical Press, Amsterdam).

PITTENDRIGH, C. & DAAN, S., 1997 "A Functional Analysis of Circadian

Pacemakers in Nocturnal Rodents V. Pacemaker Structure: A Clock for All Seasons." *J Comp Physiol A* **106**. 333–355.

Plautz, J. D., Kaneko, M., Hall, J. C. & Kay, S. A., 1997 "Independent photoreceptive circadian clocks throughout Drosophila." *Science* **278**. 1632–1635.

Preitner, N., Damiola, F., Lopez-Molina, L., Zakany, J., Duboule, D., Albrecht, U. & Schibler, U., 2002 "The orphan nuclear receptor REV-ERBalpha controls circadian transcription within the positive limb of the mammalian circadian oscillator," *Cell* **110**, 251-260.

Rand, D., Shulgin, B., Salazar, D. & Millar, A., 2004 "Design principles underlying circadian clocks," *J R Soc Interface* **1**, 119–130.

Reppert, S. & Weaver, D., 2002 "Coordination of circadian timing in mammals," *Nature* **418**, 935–941.

Rosato, E., Codd, V., Mazzotta, G., Piccin, A., Zordan, M., Costa, R. & Kyriacou, C., 2001 "Light-dependent interaction between Drosophila CRY and the clock protein PER mediated by the carboxy terminus of CRY," *Curr Biol* **11**, 909–917.

Rutila, J. E., Suri, V., Le, M., So, W. V., Rosbash, M. & Hall, J. C., 1998 "CYCLE is a second bHLH-PAS clock protein essential for circadian rhythmicity and transcription of *Drosophila period* and *timeless*." *Cell* **93**, 805–814.

Sambrook, J., Fritsch, E. & Maniatis, T., 1989 *Molecular Cloning. A Laboratory Manual.*, second edn. (Cold Spring Harbour Laboratory Press, New York).

Sato, E., Nakamichi, N., Yamashino, T. & Mizuno, T., 2002 "Aberrant expression of the Arabidopsis circadian-regulated APRR5 gene belonging to the APRR1/TOC1 quintet results in early flowering and hypersensitiveness to light in early photomorphogenesis," *Plant Cell Physiol* **43**, 1374–1385.

Sato, T., Panda, S., Miraglia, L., Rudic, R., McNamara, P., Naik, K., Fitzgerald, G., Kay, S. & Hogenesch, J., 2004 "A Functional Genomics Strategy Reveals Rora as a Component of the Mammalian Circadian Clock," *Neuron* **43**, 527–573.

Schaffer, R., Ramsay, N., Samach, A., Corden, S., Putterill, J., Carré, I. A. & Coupland, G., 1998 "The *late elongated hypocotyl* mutation of *Arabidopsis* disrupts circadian rhythms and the photoperiodic control of flowering," *Cell* **93**, 1219–1229.

Schmitz, O., Katayama, M., Williams, S. B., Kondo, T. & Golden, S. S., 2000 "CikA, a Bacteriophytochrome That Resets the Cyanobacterial Circadian Clock," .

Sehgal, A., Price, J. L., Man, B. & Young, M. W., 1994 "Loss of circadian behavioral rhythms and per RNA oscillations in the Drosophila mutant timeless," *Science* **263**, 1603–1606.

Shearman, L. P., Sriram, S., Weaver, D. R., Maywood, E. S., Chaves, I., Zheng, B., Kume, K., Lee, C. C., van der Horst, G. T., Hastings, M. H. & Reppert, S. M., 2000 "Interacting molecular loops in the mammalian circadian clock," *Science* **288**, 1013–1019.

Song, H. R. & Carré, I. A., 2005 "DET1 regulates the proteasomal degradation of LHY, a component of the Arabidopsis circadian clock," *Plant Mol Biol* **57**, 761–771.

STELLING, J., GILLES, E. & DOYLE, F., 2004 "Robustness properties of circadian clock architectures," *Proc Nat Acad Sci USA* **36**, 13210–13215.

STOLERU, D., PENG, Y., AGOSTO, J. & ROSBASH, M., 2004 "Coupled oscillators control morning and evening locomotor behaviour of Drosophila." *Nature* **431**, 862–868.

STOLERU, D., PENG, Y., NAWATHEAN, P. & ROSBASH, M., 2005 "A resetting signal between Drosophila pacemakers synchronizes morning and evening activity," *Nature* **438**, 238–242.

STRAYER, C., OYAMA, T., SCHULTZ, T. F., RAMAN, R., SOMERS, D. E., MÁS, P., PANDA, S., KREPS, J. A. & KAY, S. A., 2000 "Cloning of the Arabidopsis clock gene TOC1, an autoregulatory response regulator homolog," *Science* **289**, 768–771.

TAKUMI, T., MATSUBARA, C., SHIGEYOSHI, Y., TAGUCHI, K., YAGITA, K., MAEBAYASHI, Y., SAKAKIDA, Y., OKUMURA, K., TAKASHIMA, N. & OKAMURA, H., 1998 "A new mammalian period gene predominantly expressed in the suprachiasmatic nucleus," *Genes Cells* **3**, 167-176.

TAYLOR, B., 2004 *The role of the late elongated hypocotyl (LHY) gene in the Arabidopsis circadian clock*, Phd, University of Warwick.

TEI, H., OKAMURA, H., SHIGEYOSHI, Y., FUKUHARA, C., OZAWA, R., HIROSE, M. & SAKAKI, Y., 1997 "Circadian oscillation of a mammalian homologue of the Drosophila period gene," *Nature* **389**, 512–516.

TOLEDO-ORTIZ, G., HUQ, E. & QUAIL, P., 2003 "The Arabidopsis Basic/Helix-Loop-Helix Transcription Factor Family." *Plant Cell* **15**, 1749–1770.

TOMITA, J., NAKAJIMA, M., KONDO, T. & IWASAKI, H., 2005 "No Transcription-Translation Reedback in Circadian Rhythm of KaiC Phosphorylation," *Science* **307**, 251–254.

TOTH, R., KEVEI, E., HALL, A., MILLAR, A., NAGY, F. & KOZMA-BOGNAR, L., 2001 "Circadian clock-regulated expression of phytochrome and cryptochrome genes in Arabidopsis," *Plant Physiology* **127**, 1607–1616.

VICZIÁN, A., KIRCHER, S., FEJES, E., MILLAR, A. J., SCHÄFER, E., KOZMA-BOGNÁR, L. & NAGY, F., 2005 "Functional characterization of phytochrome interacting factor 3 for the Arabidopsis thaliana circadian clockwork," *Plant Cell Physiol* **46**, 1591–1602.

VITATERNA, M., KING, D., CHANG, A.-M., KORNHAUSER, J., LOWREY, P., MCDONALD, J., DOVE, W., PINTO, L., TUREK, F. & TAHAKASHI, J., 1994 "Mutagenesis and mapping of a mouse gene, *Clock*, essential for circadian behavior." *Science* **264**, 719–725.

WANG, W. & MALCOLM, B., 1999 "Two-Stage PCR Protocol Allowing Introduction of Multiple Mutations, Deletions and Insertions Using QuikChange^{TM} Site-Directed Mutagenesis," *BioTechniques* **26**, 680–682.

WANG, Z.-Y., KENIGSBUCH, D., SUN, L., HAREL, E., ONG, M. S. & TOBIN, E., 1997 "A Myb-Related Transcription Factor Is Involved in the Phytochrome Regulation of an Arabidopsis Lhcb Gene," *Plant Cell* **9**, 491–507.

WANG, Z. Y. & TOBIN, E. M., 1998 "Constitutive expression of the *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) gene disrupts circadian rhythms and suppresses its own expression," *Cell* **93**, 1207–1217.

WILLIAMS, M., FOSTER, R. & CHUA, N.-H., 1992 "Sequences Flanking the

Hexameric G-Box Core CACGTG Affect the Specificity of Protein Binding." *Plant Cell* **4**, 485–496.

YAMAGUCHI, R., NAKAMURA, M., MOCHIZUKI, N., KAY, S. & NAGATANI, A., 1999 "Light-dependent translocation of a phytochrome B-GFP fusion protein to the nucleus in transgenic Arabidopsis." *J Cell Biol* **145**, 437–445.

YAMAMOTO, Y., SATO, E., SHIMIZU, T., NAKAMICH, N., SATO, S., KATO, T., TABATA, S., NAGATANI, A., YAMASHINO, T. & MIZUNO, T., 2003 "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.

YAMASHINO, T., MATSUSHIKA, A., FUJIMORI, T., SATO, S., KATO, T., TABATA, S. & MIZUNO, T., 2003 "A Link between circadian-controlled bHLH factors and the APRR1/TOC1 quintet in Arabidopsis thaliana," *Plant Cell Physiol* **44**, 619–629.

ZEILINGER, M. N., FARRÉ, E. M., TAYLOR, S. R., KAY, S. A. & DOYLE, F. J., 2006 "A novel computational model of the circadian clock in Arabidopsis that incorporates PRR7 and PRR9," *Mol Syst Biol* **2**, 58.

ZYLKA, M. J., SHEARMAN, L. P., LEVINE, J. D., JIN, X., WEAVER, D. R. & REPPERT, S. M., 1998 "Molecular analysis of mammalian timeless." *Neuron* **21**, 1115–1122.

Appendix I

The sequence of the *LHY* gene of *Arabidopsis thaliana,* as it was annotated on the Arabidopsis Information Resource website (`http://www.arabidopsis.org`) in January 2007.

LHY genomic sequence (chromosome 1) AT1G01060.1
Red=UTR
Blue=introns
Orange=coding
Bold=not included in reporter construct

attttggaataatttcggttatttcaattagattcgggtagttcagttcttcggttagtaacaaaaactggtctattgttttttgg
t

-1618

taacctagaaccgaaccgaactaaccaaagttctcggtaacctttttgtagtggcttcctgaccgatgaggccgtcaacttcaaaaa
atattgcaactaagctctgctccaacattagagtatctataactatgttaaacgcttctgctttaagcaaaacacagttgtaagct
ggaatctaaaaaaatgagtgtaatgatgtttgctgaattccataaataaatactacatgcttcggttaagacttaagagtaattaa
tgttccttaatttctacaaatgttatataagcaagttgaccaaagttctcgatgataatttgttgaaattttgtataggcattgca
tgatattatatgaaaagatgaagatttttatacagacgcaagttccccgagcagtccaagcttgtcgggtttaattcaactatgtt
aatacgcaaatttatatagaataggcgtaaaagtgaggcccatacaatgtcttattacaagcccaga

-1110

tccagcatagccaatacgtagcagtacaccatcacagctggcaccgtacccacttgtttagtcgtccaagtttgtaccaataatcg
tttacacgtaagcaattgtggaccaccacactcacttttacctacgtgagcttcacattgaagc

-957

aagttctggctcgtagagaagcaacttgagatataccaaaaagtgcagtagacagccactacaatatcaccacgtgtcgatctgcg
atgacttctggtttttccatttatacc

-847

cttggggctgttccagcctcaaataaacttttcaattaaaatttttccaaaaattaggggaaaaattgttgtggctgagattgctt
ctggcttctcttcttcttcttccagtcttcttcagcctaaaa[illegible]
cagttatcttcttccttcttctctctgtttttttaaatttatttttagagatttttttttgttttgcttccgatttgattatttccg
ggaacgatgacttctccggggagttcccggtgagatgataagtcagattgcatacttgtctcctccatggctactctcaaggg[illegible]
aacagtttacattatgagcagtttctaggattcctataacatactaagatctctgtttggctgctgagaaacttatagaagcgatt
aactaaatcttattagctctaaaagttagcataaatgatacgaatctggtgattgattactgatatgaagatttgtgaaggttttg
gctgcggtggattcgtttgggtgaggcttttgtgaataataataaagggaattcttttgagttctgctggagaagcagcgactgtt
tcacggtggactttgaaaagatttctcttttgaatttcgctcatcacttttatcttagtgtttgtggataaatatttctcataaag
tactttctcctttgcagtttctctagaatctaaagaggttatcacaacggctttgcaatttgaaaactttcatgtttggggagatc
aaagatggtttctttttatactttacttgttagagaggatttgaagcagcgaatagctgcaccggtcctgt ATG
[illegible]gtactactactaatgaaataagattggtgttttttgtttgagagatttggactgttgtt
gtgtgaagatttgatttttcttttgggttttcaaatgtttag[illegible]gtcgaaagtttatcttttg
aatgtttagtttgaactctttgagattttatattcctttgtttaggagtgt[illegible]
cttttcattttgtgtgcag[illegible]gtaag
ttgatgatcctttcagatctttttaaaacactttcgttatcttgtgtaaaggt[illegible]
agaagttgcctgtcactgtctggaaaagatgatagaaggttcttctattcttccttttttctaattgcttcaactagaaattagg
cgtctaagtagagtgtgttttgctaagtagagttggttggttccgaggattgcatctcgatcgtttttttatcttttttcatttact
tgttcaactaaacttaacattagattggtctgtccaaacttactgttagaaaaaactctacagagaacttttcttttctgagacca
aagataggtccaaattaacataactaccgttgcatttggatgacacaattgtcccatattcactaaactcgatcaaaataatgggat

tgcatgaattcttttagccatccatatggaaaccaaacctccttgccctataggacaggacatgaagtaactctatcatctctacc
ccgaaaaattcgggtcagtataagacttttttcaaccttggtgatgtggtcaatttcatgaatctcaggatttcacatctgagctg
cattttccatctccagtttggaaaagctcacagttttggttcacatttcagTTGGAGAAAGAGGCTGAAGTTAAAGGCATCCCT
GTTTGCCAAGCTTTGGACATAGAAATTCCGCCTCCTCGTCCTAAACGAAAACCCAATACTCCTTATCCTCGAAAGCCTGGGAACAA
CGGTACATCTTCCTCTCAAGTATCATCAGCAAAAGATGCAAAACTTGTTTCATCGGCCTCTTCTTCACAGTTGAATCAGGCGTTCT
TGGATTTGGAAAAAATGCCGTTCTCTGAGgtttattcctgatctttctcttcttcattttatcattttggtggtcttgtgatagac
ttcttatgttttattttctggttagAAAACATCAACTGGAAAAGAAAATCAAGATGAGAATTGCTCGGGTGTTTCTACTGTGAACA
AGTATCCCTTACCAACGAAAgtaagtctaagaaagctacaaaattgtagtcgaagaatctgatgaatgctttcaagtgcatgacaa
attatttgttccattctgatttttagCAGGTAAGTGGCGACATTGAAACAAGTAAGACCTCAACTGTGGACAACGCGGTTCAAGAT
GTTCCCAAGAAGAACAAAGACAAAGATGGTAACGATGGTACTACTGTGCACAGCATGCAAAACTACCCTTGGCATTTCCACGCAGA
TATTGTGAACGGGAATATAGCAAAATGCCCTCAAAATCATCCCTCAGGTATGGTATCTCAAGACTTCATGTTTCATCCTATGAGAG
AAGAAACTCACGGGCACGCAAATCTTCAAGCTACAACAGCATCTGCTACTACTACAGCTTCTCATCAAGCGTTTCCAGCTTGTCAT
TCATGCTGCAGCTACATTCGCTGCTTCGGTCTGGCCTTATGCGAGTGTCGGGAATTCTGGTGATTCATCAACCCCAATGAGCTCTT
CTCCTCCAAGTATAACTGCCATTGCCGCTGCTACAGTAGCTGCTGCAACTGCTTGGTGGGCTTCTCATGGACTTCTTCCTGTATGC
GCTCCAGCTCCAATAACATGTGTTCCATTCTCAACTGTTGCAGTTCCAACTCCAGCAATGACTGAAATGGATACCGTTGAAAATAC
TCAACCGTTTGAGAAACAAAACACAGCTCTGCAAGATCAAAACTTGGCTTCGAAATCTCCAGCTTCATCATCTGATGATTCAGATG
AGACTGGAGTAACCAAGCTAAATGCCGACTCAAAAACCAATGATGATAAAATTGAGGAGGTTGTTGTTACTGCCGCTGTGCATGAC
TCAAACACTGCCCAGAAGAAAAATCTTGTGGACCGCTCATCCTGTGGCTCAAATACACCTTCAGGGAGTGACGCAGAAACTGATGC
ATTAGATAAAATGGAGAAAGATAAAGAGGATGTGAAGGAGACAGATGAGAATCAGCCAGATGTTATTGAGTTAAATAACCGTAAGA
TTAAAATGAGAGACAACAACAGCAACAACAATGCAACTACTGATTCGTGGAAGGAAGTCTCCGAAGAGgtaaaacaaaaaccaatg
attcatggatattgcagagttcttctgggaactaacactatcaaaatattttcagGGTCGTATAGCGTTTCAGGCTCTCTTTGCAA
GAGAAAGATTGCCTCAAAGCTTTTCGCCTCCTCAAGTGGCAGAGAATGTGAATAGAAAACAAAGTGACACGTCAATGCCATTGGCT
CCTAATTTCAAAAGCCAGGATTCTTGTGCTGCAGACCAAGAAGGAGTAGTAATGATCGGTGTTGGAACATGCAAGAGTCTTAAAAC
GAGACAGACAGGATTTAAGCCATACAAGAGATGTTCAATGGAAGTGAAAGAGAGCCAAGTTGGGAACATAAACAATCAAAGTGATG
AAAAAGTCTGCAAAAGGCTTCGATTGGAAGGAGAAGCTTCTACATGAcagacttggaggtaaaaaaaaaacatccacattttatc
aatatctttaaatctagtgttagtagtttgcttctccaatctttatgaaagagacttttaattttccttccgaacatttctttggt
catgtcaggttctgtaccatattaccccatgtcttgtctcttgtctctgtttgtgtatgctacttgtggtctatatgtcatctgct
actactgttaattaaccattaagcaatggatttgtctttattcttctttttaaaaattgtggatattgattgatgagatcataaat
agtttataatgataaaagatgaactgtacaaaagttacaacttagaaagttttgtcttttggtgtttcttcataatgattctctct
gtttgtttcaaacttgaaggcgctcgttgcccaacgccaaatgcccaaaggtgctaggctgggccttaaagattttactcacacta
cttgggcctacattacaacattttccctattacaagcttaagtaatagaggcctaaagaaggcttatattgtgtagctagttttat
acaatatttcaatctcttaatttggtttactgatgaatgagtagatatacaatttttttgttctcactgaatgagccatatccttg
ctaccaaaccaaattatttgtgaaaaatgtaatgcaagtatcctacatagattataggagtgaccgcaaaaacacaaactatgttt
cgtaataaaaaataaagtatttttaaaaagatgtaaatctttgtaaataattgagggtagtgta