Structural investigation of the *Arabidopsis thaliana* circadian clock

Thesis submitted for the degree of Doctor of Philosophy

Oliver Durrant B. Sc. (Hons.)

Department of Biological sciences, University of Warwick

September 2009

Table of Figures

Tables

Declaration

I hereby declare that the research submitted in this thesis was conducted by myself under the supervision of Prof. Vilmos Fülöp and Dr. Isabelle Carré at the Department of Biological Sciences, University of Warwick. No material in this thesis has been presented previously for any qualification or publication.

Acknowledgments

I would like to start by thanking my supervisors Vilmos and Isabelle for their time and support throughout this thesis. It has been a challenging project, so advice and scientific direction was greatly appreciated at the time. To this end, I also need to say a huge thank you to Dr Dean Rea (aka the saviour of SBG). The constructs you prepared for initial expression trials were gratefully received. In addition, I have to thank you for all the technical / scientific advice you gave, and for being a good friend throughout. I refuse to thank you for all the whippings you dealt me at squash!

My family have been absolutely wonderful during my undergraduate degree and my PhD. I have a tendency to worry, so thanks for being so understanding when I have been ill tempered, upset, hard work, frustrating (insert fault as appropriate).... Mum (you can stop worrying now!), Corky, William and Maisie, I love you all very much. Special thanks to my Grandma, who helped me out when I was writing my thesis. I may not have been able to complete this work without your assistance.

For my Dad, the biggest THANKS (literally!). You gave me impartial, logical and sensible advice when it was needed. Your constant contact was a real boost. Knowing someone else is thinking of you made it bearable when I was ready to walk. I have to also thank the RBP (Royal Bank of Paul), of which Dad is the CEO. I became a frequent and demanding customer and have yet to deposit any of my new found riches. Perhaps it isn't such a good time to be banking anyway!?! Blair (he does go by another pseudonym) made certain of that wouldn't you say Dad?

Big 'nuff respec' to the Structural massive. What a brilliant environment to work in. Great personalities that made the social side of work an absolute pleasure. I doubt I will ever be lucky enough to have such a wonderful group of people around me at work. You're included in that Vilmos (fancy a pint at the Cottage Inn?) I miss you all.

These acknowledgements would not be complete without the mention of five individuals who kept me going when it all got too much. First I would like to say thanks to Ray. You were the distraction I needed to keep me sane. You never once told me to stop moaning. You let me get it off my chest so I could cope with the pressure of the next day. You're the definition of a true friend.

Much love to my best friend Ian. I'm never happier than when we are together. It is strange to have a life-long friend whom I still admire (faults and all). How are you so damn talented? Mind you, still not talented enough to pull off your foray into cross-dressing!

Darryl and Sam, I truly don't know how this work would have been done without you. You have become my surrogate family. I literally can not summarise all you have done for me. Suffice to say, I love you both very much.

Finally, thanks to my brother Corky. You're one charismatic man! Thanks for Rio's and the raves, the beers, Tony Hark, Reet Petite, Brian Badonde, lending of shorts and t-shirts, screams of "where's the lamb sauce", "well you must be very proud of yourself" and "la de dah", football, cigarettes, disco dancing, and basically everything I enjoy doing. I'm lucky to have you as a brother.

List of abbreviations

Amp	Ampicilin
ARR	*Arabidopsis* Response Receiver
BLASTp	Basic Local Alignment Sequence Tool (protein)
CAB	CHLOROPHYLL A/B BINDING PROTEIN
CBS	CCA1 binding site
CCA1	CIRCADIAN CLOCK ASSOCIATED 1
CCR2	COLD AND CIRCADIAN REGULATED 2
CCT	CONSTANS, CONSTANS-LIKE and TOC1
CD	Circular Dichroism
CDF1	Cycling of Dof Factor 1
CESG	Center for Eukaryotic structural Genomics
CHE	CCA1 HIKING EXPEDITION
Chlor	Chloramphenicol
CK2	Caesin kinase 2
CO	CONSTANS
Col	Columbia
COP1	CONSTITUTIVELY PHOTOMORPHOGENIC 1
CRY	CRYPTOCHROMES
CT	Circadian time
DD	Constant darkness
DET1	DE-ETIOLATED 1
EE	Evening element
ELF3	EARLY FLOWERING 3
ELF4	EARLY FLOWERING 4
EMS	Ethyl methanesulfonate
FBP	F-box protein
FKF1	FLAVIN-BINDING, KELCH REPEAT, F-BOX 1
FMN	Flavin Mononuclotide
FPLC	Fast protein liquid chromatography
FR	Far-red light
FRQ	FREQUENCY
FT	FLOWERING LOCUS T
GI	GIGANTEA
HK	Histidine Kinase
H-L-H	Helix-loop-helix
HRP	Horseradish peroxidase
IPTG	Isopropyl-beta-thiogalactopyranoside
LB	Luria-Bertani
LCHB	LIGHT HARVESTING CHLOROPHYLL A/B BINDING PROTEIN
LD	Light / Dark
LDP	Long day plant
Ler	Landsberg *erecta*
LHY	LATE ELONGATED HYPOCOTYL
LIP1	LIGHT INSENSITIVE PERIOD 1
LKP2	LOV, KELCH PROTEIN 2
LL	Constant light
LOV	Light, oxygen, voltage
LRR	Leucine rich repeats
LUC	LUCIFERASE
LUX	LUX ARRHYHTMO
MBP	MALTOSE BINDING PROTEIN
MCP	Methyl-accepting chemotaxis protein

MCS	Multiple cloning site
MDL	Molecular dimensions Ltd.
Mw	Molecular weights
OD	Optical Density
OGT	*O*-linked *β*-*N*-acetylglucosamine
PAGE	Polyacrylaminde gel electrophoresis
PAS	PERIOD CIRCDAIN PROTEIN, Ah receptor nuclear transolcator protein, SINGLE MINDED PROTEIN
PBS	Phosphate buffered saline
PRC	Phase Response Curve
PCR	Polyermase chain reaction
PDB	Protein Database
PER	PERIOD
PHOT	PHOTOTROPIN
PHY	PHYTOCLOCK
PHY A-E	PHYTOCHROME A-E
PIF3	PHYTOCHROME INTERACTING FACTOR 3
PRR	PSUEDO-RESPONSE REGULATOR
R	Red light
RR	Response Reciever
SCF	Skp / Cullin / F-box
SCN	Suprachiasmatic nucleus
SDP	Short day plant
SDS	Sodium dodecyl sulfate
SPY	SPINDLY
SRR1	SENSITIVITY TO RED LIGHT REDUCED 1
TBS	TCP1 binding site
TBS	Tris-buffered saline
TCP	TB1, CYC, PCFs
TCS	Two-component systems
T-DNA	Transfer DNA
TEV	Tobacco Etch Virus
TIC	TIME FOR COFFEE
TOC1	TIMING OF CAB EXPRESSION 1
WNK1	WITH NO LYSINE 1
Ws	Wassilewskija
ZT	Zeitgeber
ZTL	ZEITLUPE

Summary

Plants, like most organisms, have developed elaborate mechanisms for anticipating periodic environmental changes. The circadian clock allows an organism to adapt its metabolic, developmental and physiological processes to coincide with favourable environmental conditions. At the centre of the *Arabidopsis thaliana* clock, linking environmental inputs and driving the overt biological rhythm is a central oscillator that consists of multiple interlocked transcriptional / translational negative feedback loops. What is known about the structure of the central oscillator comes primarily from genetic analysis. Less clear, is how putative oscillator proteins perform their perceived functions in circadian rhythm maintenance. Described are the cloning, expression and purification of clock-associated proteins; TOC1-PRR, ELF4, LUX and LIP1. Purified ELF4 was subjected to unsuccessful crystallisation trials, probably due to its intrinsically unstructured nature. A truncated form of LIP1 was shown to be an active GTPase, representing the first example of an active GTPase in the *plant* clock. In addition, a protocol for the production of ssDNA aptamers has been developed (against SRR1), which can be used to replace antibody-based experimentation. The work presented discusses the difficulty in obtaining the novel, *plant*-specific proteins in quantities required for crystallisation, and suggests alternative methods for structural and biochemical analysis of these proteins. Moreover, this thesis combines experimental data with a range of Bioinformatic tools to aid design for subsequent biochemical expression, purification and crystallisation trials.

Chapter 1. Introduction

Almost all living organisms have developed elaborate ways to coordinate their activities with the Earth's rotation. Both daily and seasonal fluctuations in the environment alter fundamental life processes, from whole organism seasonal migration to cellular events including gene transcription and cell proliferation. However, subsets of these alterations persist even in the absence of environmental cues. Biological rhythms with a periodicity of approximately 24 hours (h) under constant conditions are termed circadian rhythms (Fig 1.1). Although oscillations persist in constant conditions, the clock is set (entrained) to rhythmic cues of the day/night cycle, most notably light and temperature. Without entrainment the clock would have to be completely accurate or it would gradually drift out of phase with the environment. One further property of circadian clocks is that they are temperature compensated. When an organism is placed at different temperatures within its physiological range, the period remains relatively stable (Pittendrigh, 1960a). This important feature allows the phase relationship between physiology of an organism and the environment to remain robust during unpredictable changes in temperature. Having an internal molecular timekeeper allows an organism to predict environmental changes conferring a selective advantage, as they alter their developmental, physiological and biochemical processes to coincide with favourable conditions, which increase the chance of reproductive success and survival.

1.1. Molecular mechanism of circadian clocks

The core clock components vary between organisms, with very little conservation of the molecules involved, suggesting independent evolution of the circadian system on several occasions (Millar, 2004; Young and Kay, 2001). Despite this, conceptually, the molecular mechanisms of circadian clocks show remarkable similarity between Kingdoms. The simplest model highlights three basic components; input pathways, the central oscillator and output pathways (Fig 1.2.) (Dunlap *et al.*, 1999). Input pathways perceive environmental cues, notably light and temperature, and transmit the information by signal transduction to the central oscillator. This allows the entrainment of an organism to its environment. The central oscillator then generates and maintains rhythmicity utilising positive / negative feedback loops to drive output

pathways. The output pathways connect the oscillator to the metabolism and physiology of the organism. The importance of the circadian clock has been highlighted by microarray experiments which show that 8 % of mammalian genes and as many as 15 % of *plant* genes are under rhythmic control (Edwards *et al.*, 2006).

Although conceptually appealing, this model is an oversimplification of how circadian systems work. Lines of evidence suggest that rhythmically controlled outputs feed back to the central oscillator (Gardner *et al.*, 2006; McClung, 2006) and input pathways are themselves outputs of the clock (Tóth *et al.*, 2001). Adding further complexity is the fact that the oscillator may consist of several feedback loops that contribute independently of each other in maintaining rhythmicity. Furthermore, the core oscillator in *Drosophila melanogaster*, *Neurospora crassa* and *Arbidopsis thaliana* is thought to comprise of several interlocking feedback loops. These feedback loops are thought to be cell and tissue specific. For example, the mammalian clock comprises of a core oscillator in specialised cells of the suprachiasmatic nucleus (SCN) located in the brain. This oscillator acts as the central oscillator of a subset of 'slave' oscillators located in peripheral tissues; liver, kidney, oesophagus and skin. Until recently all *plant* cells were thought to contain the same central oscillator in every cell. It now appears that the oscillator found in shoots differs from the oscillator located in roots (James *et al.*, 2008), indicating a degree of tissue specificity and also another level of complexity.

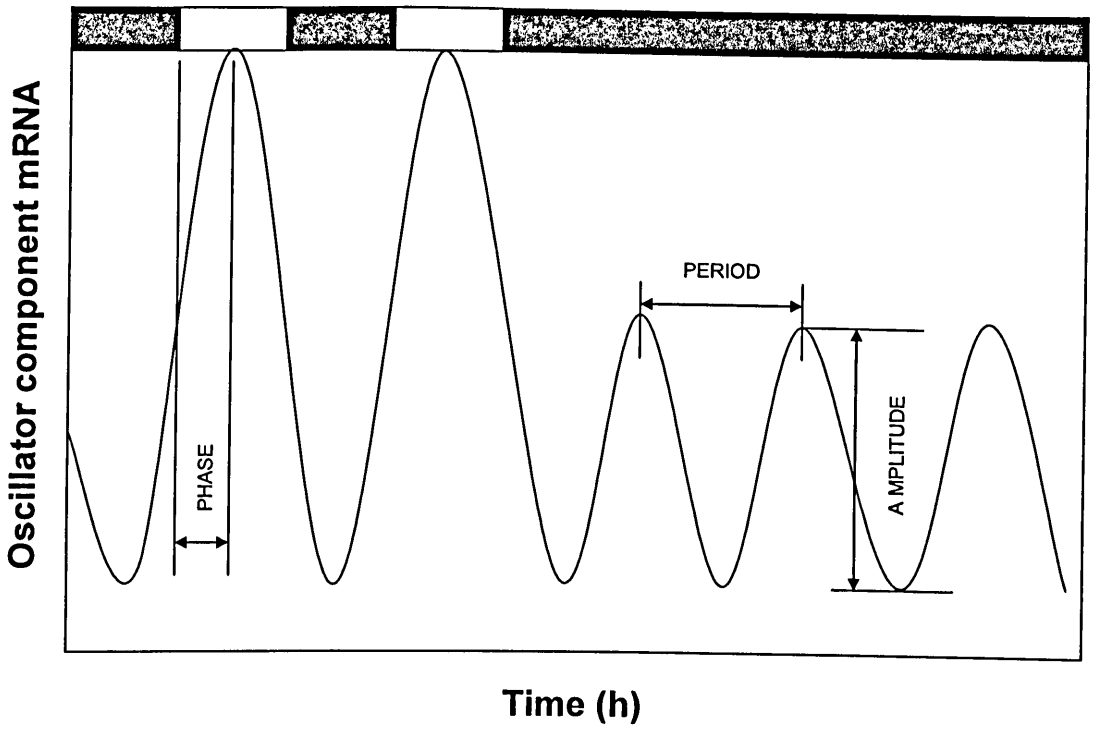


Fig 1.1. Terms used to describe circadian rhythms. Period is defined as the time it takes to complete one full cycle. It can be measured from any phase marker, although it is usually measured from peak to peak. Phase is the time of day that an event occurs. For example, if the peak of a rhythm occurs at mid-day, its phase would be 12 h. Phase is often expressed in terms of Zeitgeber time (ZT). Zeitgeber literally translates as time-giver in German, with dawn defined as ZT0 and dusk as ZT12. The amplitude of a rhythm is one half the peak-trough distance.
White boxes represent periods of light, dark boxes represent periods of darkness. In this example, a clock component is entrained to two light-dark (LD) cycles before being transferred to constant darkness. The clock component has a period of 24 h under LD cycles. Under free-running conditions, the component's period shortens and the amplitude of the rhythm decreases.

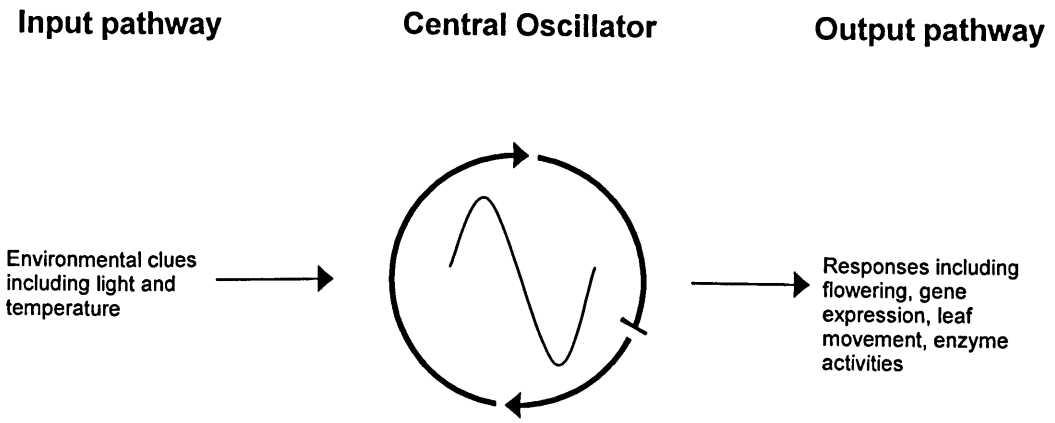


Fig 1.2. Schematic representation of the circadian clock. The clock has been classically divided into input pathways, a central oscillator and output pathways. Within the central oscillator, the positive / negative feedback loop is depicted with arrows for the positive arm and perpendicular lines representing the negative arm.

1.2. Rhythmicity in Plants

Diurnal rhythms have been noted from as far back as the fourth century BC. Androsthenes observed daily leaf movements of the tamarind tree on Tylos, an island in the Persian Gulf (Reviewed in McClung, 2006). These movements were not attributed to any internal rhythm at the time and it was not until the 18th century that the French astronomer de Marian reported daily movements of leaves persisted in constant darkness (de Mairan, 1729), suggesting an endogenous timekeeping mechanism. Almost 100 years later, the period of these leaf movements were deduced to be ~24 h (but not exactly 24 h), further indicating that these rhythms were more than a mere response to varying light levels, but were in fact circadian in nature. Not surprisingly these findings were greeted with scepticism. In 1873, German Botanist Pfeffer (reviewed in van Doorn and van Meeteren, 2003) set out to prove that observed leaf movements were due to flawed experimental procedure, suggesting that light was leaking into the caves and wine cellars thus disrupting the constant conditions. In doing so, Pfeffer recorded many examples of *plant*s with free-running periods and these served as the greatest evidence that the rhythms were endogenous rather than due to some subtle effect of a geophysical cue. Despite all the evidence, circadian rhythms were not generally accepted until the beginning of the 20th century (Sweeny, 1969).

The circadian clock is now known to control many aspects of *plant* physiology including leaf movement, hypocotyl elongation, odour production, stomatal opening, calcium levels, carbon dioxide fixation and photosynthesis (Engelmann and Johnson, 1998; Dowson-day and Millar, 1999; Overland, 1960; Webb *et al.*, 1998; Johnson *et al.*, 1995; Dodd *et al.*, 2005; Harmer *et al.*, 2001). The clock is also able to measure photoperiod (day-length), allowing the detection of seasonal changes (Sweeny, 1987) which is fundamental with regard to flowering time.

Although the connection between the circadian clock and fitness is not fully understood, there is increasing evidence that it provides the *plant* with a competitive advantage. *Plant*s with clock periods in sync with the environment fix more carbon, contain more chlorophyll, grow quicker and survive better than short or long period mutants in *Arabidopsis* (Dodd *et al.*, 2005). This indicates that the circadian clock allows prediction of environmental changes, allowing the best use of resources available.

1.2.1. *Arabidopsis* as a tool for studying the *Plant* circadian clock

Arabidopsis thaliana has become an important tool in the elucidation of *plant* circadian biology. It has a relatively small genome, is small in size and has a short generation time (6-8 weeks). Perhaps even more significant is its genetic similarity with the world's most agriculturally important crops including corn, wheat and soybean. *Arabidopsis thaliana* is therefore a perfect system for identification of genes that may be components of the circadian clock. Wild-type *Arabidopsis* backgrounds Colombia (col), Wassilewskija (Ws) and Landsberg *erecta* (Ler) with defective circadian associated phenotypes are used to elucidate potential components of the circadian clock.

1.2.2. Screening for molecular components

Identification of clock components was primarily based on single mutant backgrounds which show aberrant circadian phenotypes. Mutants are commonly made using the chemical ethyl methanesulfonate (EMS) or by transferred DNA (T-DNA) of an engineered Ti plasmid that is incorporated into the *plant* genome by transfection of *Agrobacterium tumefacians*. The mutational analysis of *plant*s with

abnormal circadian phenotypes could only suggest at possible genetic involvement, as many of the circadian phenotypes, including hypocotyl length and flowering time are only partially controlled by the circadian clock. However, many of the components of the input pathways have been discovered in this way.

The elongation of hypocotyls is under circadian control from the moment of germination, with rapid cell expansion in the evening and arrest of growth in the morning (Dowson-day and Millar, 1999). As light has a profound effect on the length of hypocotyls, mutants displaying unusual elongation patterns have been useful in the screening for components of the light signalling pathways, as well as highlighting components downstream that may control hypocotyl elongation.

One of the best studied developmental processes under the control of the circadian clock is the measurement of photoperiod which results in the switch from vegetative to reproductive growth. The initial hypothesis assumed that an accumulation of a chemical in the organism had to reach a threshold in order to trigger a physiological response, for example flowering. This is only achieved if the product is not degraded. Indeed, it may only accumulate under certain conditions (light) and be degraded under others (dark). The threshold is only achieved if given enough time under the positive condition. This model is referred to as the Hourglass model, but is known to be naïve as it does not account for endogenous rhythmicity and the light is required to turn over the cycle every day. The model was extended to propose the existence of a circadian rhythm of photoperiodic photosensitivity. In spring, light illuminates the photosensitive phase of the circadian clock oscillator and triggers a response. This is called the External coincidence model and it has two effects; it entrains the rhythm of the photosensitivity and also acts as the stimulus. According to the external coincidence model, the circadian clock controls a light-sensitive component that accumulates in the leaves. In *Arabidopsis*, the *CONSTANS* (*CO*) gene is the light-sensitive component (Searle and Coupland, 2004). During vegetative growth, *CO* mRNA accumulates at peak levels during dusk; however, the *CO* mRNA is very unstable in the dark. When the days become longer, the high levels of *CO* that are present, overlap with the light phase of the photoperiod. Far-red and blue light stabilise the CO protein and it translocates to the nucleus (Valverde *et al.*, 2004) where it activates the transcription of *FLOWERING LOCUS T* (*FT*) (Samach *et al.*,

2000). FT moves from the leaves to the shoot apex to promote flowering (Huang *et al.*, 2005). Therefore, mutants identified due to their defective photoperiod of flowering are often representative of defects in the circadian clock.

Genetic studies of central clock components in *Arabidopsis thaliana* was hampered by the lack of convenient, non-invasive rhythmic markers. To this end, reporter constructs utilising the clock controlled output gene promoter of *CHLOROPHYLL A/B BINDING PROTEIN* (*CAB*) fused to the firefly *LUCIFERASE* (*LUC*) gene were constructed. Supplying luciferin to transgenic *plant*s containing the *CAB:LUC* reporter construct results in bioluminescence that is rhythmic in nature and therefore reflects the activity of circadian clock (Millar *et al.*, 1995a). This luminescence is easily detected with a photon-scintillation camera and can monitor circadian rhythms in an automated fashion for as long as required. The system has been extended to include other circadian controlled promoters fused to *LUC*, allowing the successful isolation and characterisation of mutants fundamental to the control of output pathways.

Many mutants with circadian phenotypes were not initially isolated using the luciferase reporter assay. Monitoring the circadian rhythms of leaf movements in *Arabidopsis* seedlings by an automated time-lapse imagining system has provided a quicker alternative to the creation of transgenic lines required for the luciferase screen (Millar *et al.*, 1995b). Mutant seedlings are grown in 12 hr light and 12 hr dark cycles for 7 days and then transferred to constant conditions for imaging. In wild-type *Arabidopsis* seedlings, the cotyledons and leaves are maximally raised once per day under the control of the circadian clock. As this method is not as robust as the aforementioned luciferase screen, it is a useful secondary screen for lines enriched in circadian mutants.

1.3. Molecular components of the *Arabidopsis* circadian clock

1.3.1. THE TIMING OF CAB EXPRESSION 1 (TOC1)

TOC1 was the first gene to be implicated in the circadian clock using the *CAB:LUC* reporter system. The clock mutant *toc1-1* shows a period of 21 h rhythms irrespective of the light condition (Millar *et al.*, 1995b; Somers *et al.*, 1998; Strayer *et al.*, 2000) and it also shows no typical light-dependant phenotypes (Somers *et al.*, 1998; Más *et al.*, 2003). In addition, the *toc1-1* mutant allele has an early flowering phenotype insensitive of daylength under 24 h circadian cycles, but not under the 21 h endogenous period of this mutant (Strayer *et al.*, 2000), suggesting that *TOC1* is an important component of the central oscillator rather than an input to the clock. Mutations in *TOC1* shorten the period of all circadian rhythms tested (Millar *et al.*, 1995b; Somers *et al.*, 1998; Strayer *et al.*, 2000; Alabadi *et al.*, 2001). Further evidence of the importance of *TOC1* has been gained by using over-expressing lines. *TOC1* overexpressing lines show a lengthening of periods in a dosage dependant manner (Más *et al.*, 2003) with constitutively over-expressing *TOC1* abolishing rhythmic expression of all putative clock genes tested thus far (Makino *et al.*, 2002).

TOC1 is the first identified member of the PSUEDO-RESPONSE REGULATOR family (*PRR-1*), which has a peak expression at dusk, with slowly decreasing levels throughout the night (Strayer *et al.*, 2000; Mizoguchi *et al.*, 2002). The TOC1 protein levels peak in the middle of the subjective night and is degraded by *ZEITLUPE* (*ZTL*) via the 26S proteasomal pathway as the morning progresses (Más *et al.*, 2003). The nuclear localised TOC1 protein indirectly promotes the expression of two Myb-domain transcription factors *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) and *LATE ELONGATED HYPOCOTYL* (*LHY*), which have peak expression at dawn.

1.3.2 LHY.

LHY was first identified as a late flowering and long hypocotyl mutant, which was later shown to be arrhythmic with regard to circadian phenotypes (Schaffer *et al.*, 1998). *LHY* loss-of -function mutation shortens the period by 1-2 hours (Mizoguchi *et al.*, 2002; Alabadi *et al.*, 2002) and *LHY* overexpression resulted in arrhythmicity of leaf movements and clock controlled genes including *LIGHT HARVESTING CHLOROPHYLL A/B BINDING PROTEIN* (*LCHB*) and *COLD AND CIRCADIAN REGULATED 2* (*CCR2*) (Schaffer *et al.*, 1998). *LHY* mRNA levels peak in the subjective morning and decrease as the day progresses. LHY is post-translationally modified by *DE-ETIOLATED 1* (*DET1*) and degraded via the proteasomal pathway (Song and Carré, 2001).

1.3.3. CCA1

CCA1 was first identified as a protein that binds to cytosine and adenine rich sequences in the *CAB* promoter (Wang and Tobin, 1998). It is a DNA binding protein that recognises an asymmetric DNA sequence AAAATCT (the CCA1 binding site (CBS)) and also AAATATCT (the evening element (EE)), which is found in the promoters of several important clock genes. As with *LHY*, overexpression of *CCA1* results in arrhythmia of clock controlled genes (Wang and Tobin, 1998) and the *cca1* loss-of-function mutation causes a period shortening of 1-2 h (Alabadi *et al.*, 2002). Furthermore, CCA1 also exhibits peak expression at dusk.

1.3.4. The TOC1- LHY / CCA1 feedback loop

The simplest and best understood model of the *Arabidopsis* central oscillator is composed of *LHY* / *CCA1* and *TOC1*, forming a single negative and positive transcriptional / translational feedback loop (Alabadi *et al.*, 2001) (Fig. 1.3). *LHY* and *CCA1* encode Myb-domain transcription factors that have a large degree of sequence similarity and are partially redundant with respect to the clock (Schaffer *et al.*, 1998; Mizoguchi *et al.*, 2002). The fact that neither the *LHY* or *CCA1* loss-of-function mutants cause general arrhythmia for genes tested, but instead shorten the period by

1-2 h is evidence for this. However, overexpression of *LHY* and / or *CCA1* still cause large disruption to rhythms tested and the double *lhy:cca1* mutant results in a severally disrupted clock (arrhythmia after release from entrainment) (Alabadi *et al.*, 2002), indicating that they are important for proper clock functioning.

The feedback loop is proposed to function as follows: *LHY* and *CCA1* are expressed rhythmically with peak expression in the subjective morning. The cognate proteins reach maximal abundance 2-3 h later where they inhibit expression of *TOC1* by binding to the EE in the *TOC1* promoter. As the levels of LHY and CCA1 increase, levels of *TOC1* decrease and they begin to repress expression of themselves and each other. Consequently, as the day progresses, the levels of *LHY* and *CCA1* transcripts and protein begin to decrease. TOC1 becomes alleviated from its suppression and begins to accumulate with peak *TOC1* expression during the subjective evening. TOC1 then up-regulates the expression of *LHY* and *CCA1*, closing the feedback loop and re-initiating the cycle (Fig 1.3a and b).

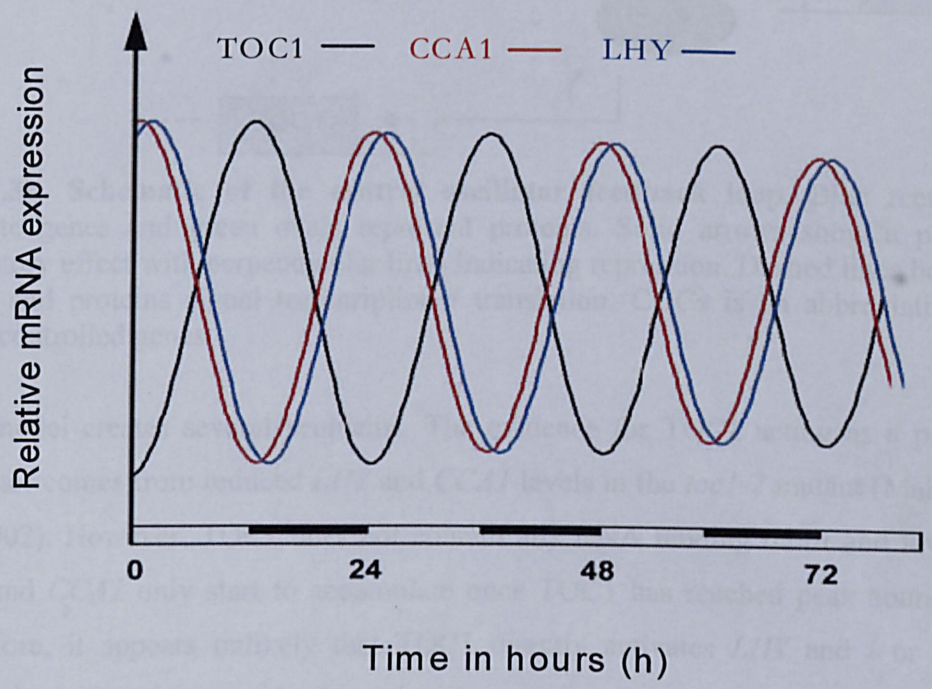


Fig 1.3a. Expression of core oscillator genes LHY, CCA1 and TOC1. White boxes represent subjective day and dark boxes represent subjective night. *LHY* and *CCA1* expression peaks during the subjective morning. As LHY and CCA1 repress the expression of *TOC1*, the trough of *TOC1* is opposite the peak of *LHY* and *CCA1*. As the day progresses, LHY and CCA1 negatively regulate themselves and each other resulting in trough expression during the subjective evening. Free from repression, TOC1 is able to accumulate, with peak expression during the subjective morning. TOC1 positively regulates *LHY* and *CCA1*, restarting the cycle. Figure reproduced from Schöning and Staiger, 2005.

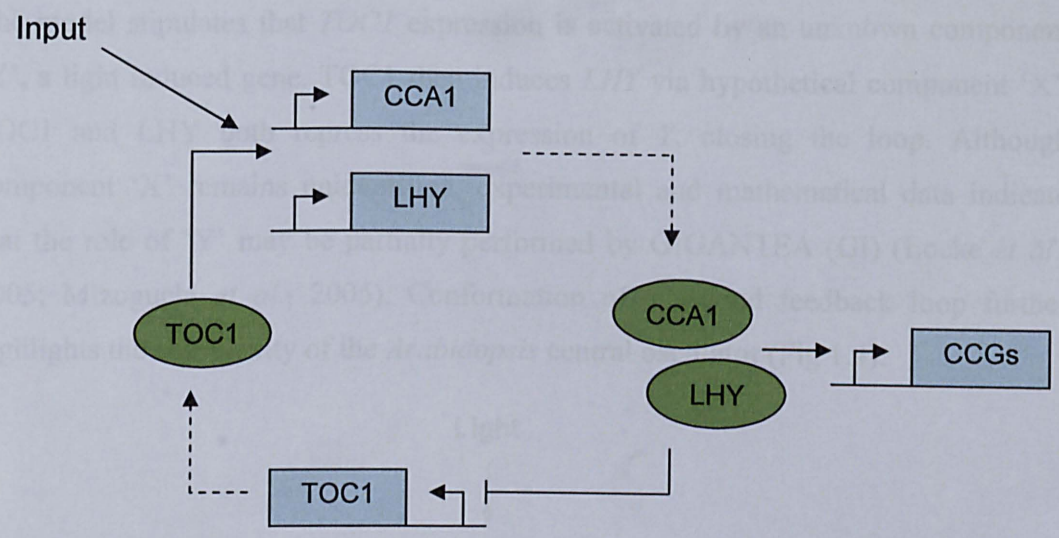


Fig 1.3b. Schematic of the central oscillator feedback loop. Blue rectangles indicate genes and green ovals represent proteins. Solid arrows show a positive regulatory effect with perpendicular lines indicating repression. Dashed lines between genes and proteins signal transcription / translation. CGCs is an abbreviation for clock controlled genes.

This model creates several problems. The evidence for TOC1 acting as a positive regulator comes from reduced *LHY* and *CCA1* levels in the *toc1-2* mutant (Makino *et al.*, 2002). However, TOC1 does not contain any DNA binding motif and levels of *LHY* and *CCA1* only start to accumulate once TOC1 has reached peak abundance. Therefore, it appears unlikely that TOC1 directly activates *LHY* and / or *CCA1* expression. Consistent with this view, overexpression of *TOC1* decreases the amplitude of *LHY* and *CCA1* rather than increasing their rate of induction (Makino *et al.*, 2002). Other proteins are therefore likely to work mutually with TOC1, as part of protein complexes or as individual feedback loops with *LHY* and *CCA1*.

A mathematical model accounting for this has recently been proposed (Locke *et al.*, 2005). The authors devised two separate models for the circadian clock to define the rhythm mathematically. Each model is defined using a set of differential equations, where each equation gives a change of either [protein] or [mRNA] over the circadian time (CT). The goodness of fit of these models with respect to real life is then calculated over many CTs, and each one has a rank assigned. The ranking of the model is given by using a cost-function algorithm. The model fits because no significant difference can be found between the model and experimental data.

The model stipulates that *TOC1* expression is activated by an unknown component 'Y', a light induced gene. TOC1 then induces *LHY* via hypothetical component 'X'. TOC1 and LHY both repress the expression of *Y*, closing the loop. Although component 'X' remains unidentified, experimental and mathematical data indicate that the role of 'Y' may be partially performed by GIGANTEA (GI) (Locke *et al.*, 2005; Mizoguchi *et al.*, 2005). Conformation of a second feedback loop further highlights the complexity of the *Arabidopsis* central oscillator (Fig 1.4).

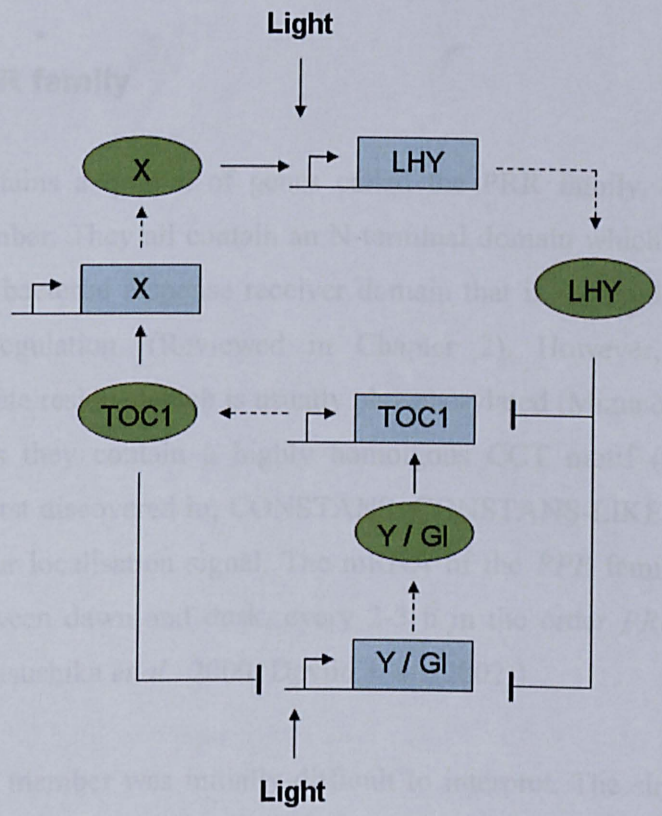


Fig 1.4. Schematic representation of the central oscillator incorporating a mathematical approach. Blue rectangles indicate genes and green ovals represent proteins. Solid arrows show a positive regulatory effect with perpendicular lines indicating repression. Dashed lines between genes and proteins signal transcription. The well-defined LHY / CCA1 – TOC1 loop is extended to include unknown component X, which is positively regulated by TOC1 and results in the up-regulation of *LHY* and *CCA1*. Component Y is a proposed positive regulator of *TOC1* that is negatively regulated by both TOC1 and LHY protein. The nuclear protein GI fulfils the experimental and mathematical criteria required for component Y.

1.4. Other clock related genes

The presence of two feedback loops that still do not account for the robustness of the clock, suggests that there are other components working within the proposed feedback loops or forming new loops themselves. Several other genes and their protein products have been implicated in playing roles within the central oscillator and these are discussed below.

1.4.1. The PPR family

Arabidopsis contains a quintet of genes called the PRR family, of which TOC1 (PRR1) is a member. They all contain an N-terminal domain which shares sequence similarity to the bacterial response receiver domain that is required in 2 component transcriptional regulation. (Reviewed in Chapter 2). However, they lack the conserved aspartate residue which is usually phosphorylated (Mizuno *et al.*, 2005). At their C-terminals they contain a highly homolgous CCT motif (named after the proteins it was first discovered in; CONSTANS, CONSTANS-LIKE and TOC1) that contains a nuclear localisation signal. The mRNA of the *PPR* family are expressed sequentially between dawn and dusk, every 2-3 h in the order *PRR9-PRR7-PRR5-PRR3-PRR1* (Matsushika *et al.*, 2000; Devlin *et al.*, 2002)).

The role of each member was initially difficult to interpret. The single *prr* mutants show a small period variation of 1-1.5 h, with overexpression of individual *PRR9*, *PRR5* and *PRR3* resulting in only small defects in clock output (Matsushika *et al.*, 2002; Mizuno and Nakamichi, 2005). These data suggest minor roles for individual members of the family. However, the triple mutant *prr5 prr7 prr9* has been shown to be arrhythmic (Nakamichi *et al.*, 2005), which shows that they are required for proper clock functioning and that there may be complex interactions between family members with a degree of redundancy present.

Recent experimental data combined with mathematical modelling have proposed another transcriptional feedback loop between PRR7 / PRR9 – LHY / CCA1 (Locke *et al.*, 2006; Zeilinger *et al.*, 2006). CCA1 positively regulates the expression of

PRR7 and *PRR9* by binding to the CBS in their promoters. By contrast, PRR7 and PRR9 negatively regulate the expression of *LHY* and *CCA1* through an unidentified mechanism. This negative regulation is supported by the 4-5 h period delay of *LHY* and *CCA1* in the *prr7 prr9* double mutant (Farré *et al.*, 2005). However, the involvement of the PRR family in light entrainment and the complex nature of the double and triple mutants, have made their role in the central oscillator difficult to deduce.

1.4.2. EARLY FLOWERING 4 (ELF4)

ELF4 is an early flowering mutant. It acts on flowering time by regulation of the floral inducer gene *CO*. *ELF4* mutants have higher activation of the *CO* gene. *ELF4* is expressed at a similar phase to that of *TOC1* and has been shown to be mandatory for normal clock function (Doyle *et al.*, 2002). Loss-of-function *elf4* mutant exhibits attenuation of free-running periods in all the clock outputs tested, including the central oscillator protein CCA1 (Doyle *et al.*, 2002). Levels of *LHY* mRNA are also very low in the knockout *elf4* mutant (Kikis *et al.*, 2005). Expression of *ELF4* is repressed by the action of LHY and CCA1, suggesting another potential feedback loop (Kikis *et al.*, 2005). Interestingly, *ELF4* also appears to be involved in photoperiod perception. The *elf4* mutant seedlings are hyposensitive to red light and *ELF4* mRNA levels are low in the *PHYTOCHROME B* (*phyB*) mutant (Khanna *et al.*, 2003). Taken together, these suggest that ELF4 is important in PHY-mediated light input into the clock. The pleiotropic nature of *ELF4* mutants make its precise role difficult to determine. A further problem is that the putative ELF4 protein appears to be novel in structure, with no known motifs or domains based on primary sequence (highlighted in Chapter 2).

1.4.3. EARLY FLOWERING 3 (ELF3)

Another early flowering mutant (*ELF3*) has been shown to be important for light input into the clock. The *elf3* mutant displayed rhythmic expression of *CCR2* in constant darkness (DD), but rhythmicity was abolished in constant light (Hicks *et al.*, 1996). Further evidence for its role in light input comes from studies using *CAB* induction. The expression of *CAB* is rhythmically repressed during the night,

allowing a greater rate of induction during the day (Hicks *et al.*, 2001). In *elf3* mutants, *CAB* is constitutively activated, therefore suggesting a role for *ELF3* in gating light input. Moreover, *elf3* mutants become arrhythmic if pulsed with light during the night (McWatters *et al.*, 2000). These claims are re-enforced by the abundance of ELF3 peaking at dusk when it would be required for repression of light. The exactly role of ELF3 is still poorly understood. As with ELF4, the ELF3 protein shows no similarity to any other protein based on primary sequence predictions. In addition, the phenotypes of the *elf3* mutants seem to be more pronounced in *PHY B* signalling. Although an interaction between PHY B and ELF3 has been shown, the *elf3 phy B* double mutant has an additive effect, indicating that ELF3 may also function independently of PHY B in another role (Reed *et al.*, 2000).

1.4.4. LUX ARRHYHTMO (LUX)

LUX (also known as (PHYTOCLOCK, PHY)) is a Myb-domain transcription factor that has been proposed to function as part of the central oscillator (Hazen *et al.*, 2005; Onai *et al.*, 2005). MYB domain transcription factors form a large family of DNA binding proteins, with varying numbers of MYB repeats. Each MYB repeat has the general architecture of H, H-L-H and it is the H-L-H that is responsible for DNA contact (Reviewed in Chapter 2). *LUX* has an expression profile similar to that of *TOC1*; with peak mRNA levels during the subjective evening. The *LUX* promoter also contains an EE which LHY and CCA1 can bind, presumably repressing its transcription as with *TOC1*. The *lux* mutant shows dampened levels of *LHY* and *CCA1* mRNA, hinting that it is a positive regulator of both (Hazen *et al.*, 2005). Interestingly, the *lux* mutant has high levels of *TOC1* transcript which suggests that LUX may be important in up-regulating the levels TOC1 (Hazen *et al.*, 2005). It also shows arrhythmia of most clock outputs under constant light and darkness, highlighting its importance to the central oscillator.

1.4.5. CCA1 HIKING EXPEDITION (CHE)

A novel high-throughput 'promoter hiking' technique was used to identify transcription factors that could bind to promoter fragments of *LHY* and *CCA1* (Pruneda-Paz *et al.*, 2008). Functional redundancies amongst gene families make identification of clock-associated genes difficult using standard forward genetic screens. To determine direct *LHY / CCA1* regulators, a yeast-one hybrid system was used to screen transcription factors capable of binding the aforementioned promoters. This approach allowed several tiled fragments for each promoter to be used and screened against a comprehensive library of circadian regulated transcription factors. The authors isolated and characterised a class 1 TCP (TB1, CYC, PCF's) transcription factor that bound to a fragment of the *CCA1* promoter (GGTCCCAC) termed the class 1 TCP binding site (TBS) (Pruneda-Paz *et al.*, 2008). Overexpressing lines of CHE result in a reduced level of *CCA1* mRNA and *che1* and *che2* single mutants result in an increase in *CCA1* promoter activity (Pruneda-Paz *et al.*, 2009). Taken together, this suggests CHE is a negative regulator of CCA1. Furthermore, LHY and CCA1 can repress the transcription of *CHE* by binding to the CCA1 binding site (CBS) in the *CHE* promoter.

As previously discussed, the mechanism by which TOC1 performs its positive function in the oscillator is unclear. *CHE* has a similar expression profile to *TOC1* and is localised in the nucleus. Yeast two-hybrid screens and co-immunoprecipitation indicate that CHE binds to the N-terminal of TOC1, representing a method for TOC1 recruitment to the *CCA1* promoter, in the absence of a DNA binding motif (Pruneda-Paz *et al.*, 2008). Overexpressing *CHE* in cell lines with elevated TOC1 levels produced a significant period shortening that was not present in *CHE* mutants with wild-type TOC1 levels. This evidence would indicate that CHE may act as an antagonist to TOC1, as well as to unidentified positive *CCA1* regulators.

1.4.6. GI

The role of GI in the clock is difficult to interpret as it appears to have numerous functions, including inhibition of hypocotyl elongation by red light (Huq *et al.*, 2000, induction of flowering under long days (discussed later) and a separate role in the oscillator (Mizoguchi *et al.*, 2005). It is a large nuclear protein (~ 140 KDa) of unknown function that is expressed at a similar phase to TOC1 (Park *et al.*, 1999). The *GI* promoter contains several EE that LHY and CCA1 can bind (Harmer *et al.*, 2000), postulating at another feedback loop analogous to the LHY / CCA1 – TOC1 loop. Indeed, *plant*s overexpressing *LHY* show low levels of *GI* transcript that are arrhythmic (Fowler *et al.*, 1999). *GI* loss-of-function mutants resulted in low levels of *LHY* and *CCA1* mRNA (Park *et al.*, 1999; Fowler *et al.*, 1999), indicating a positive regulatory role. A mathematical model for the core oscillator was proposed (Locke *et al.*, 2005) and the authors confirmed that GI could perform the role of Y, as experimental data of GI was consistent with the model (Fig 1.4).

GI also interacts with the F-box proteins which are important in targeted degradation of central clock components (discussed later) and SPINDLY (SPY), an *O*-linked β-*N*-acetylglucosamine transferase (OGT) which catalyses the addition of an *O*-linked β-*N*-acetylglucosamine (*O*-GlcNAc) to the Ser/Thr of cytosolic and nuclear proteins (Wells *et al.*, 2001). This modification is important for localisation and half-life of the protein targets. It appears that GI negatively regulates SPY and this may inhibit or down-regulate subsequent *O*-GlcNAc of protein targets (Tseng *et al.*, 2004). To date, no targets have been determined.

1.5. Entrainment of the clock

The environment provides many different signals over the duration of the day / night cycle, including variation in temperature and both quantity and quality of light. The clock has therefore developed a sophisticated method for resetting itself allowing entrainment to the environment. If the circadian clock was advanced with respect to the environment, rhythmic processes would occur too early and the entrainment would have to delay its phase. Conversely, if the clock was delayed relative to the environment, entrainment would have to advance the clock. This can be observed

experimentally when a single stimulus, for example a light pulse, is given in constant darkness. If the pulse is given at noon, there will be no effect on the clock. However, a pulse of light administered before dawn will advance the clock and a pulse of light given after dusk will delay the clock. This sensitivity can be plotted as a phase response curve (PRC) in which the shift in phase is plotted against the time the stimulus is given across the circadian cycle (Fig 1.5). PRC's have characteristic shapes for each organism, so the effect of mutants on these curves can highlight possible mechanisms of entrainment (Johnson *et al.*, 1999).

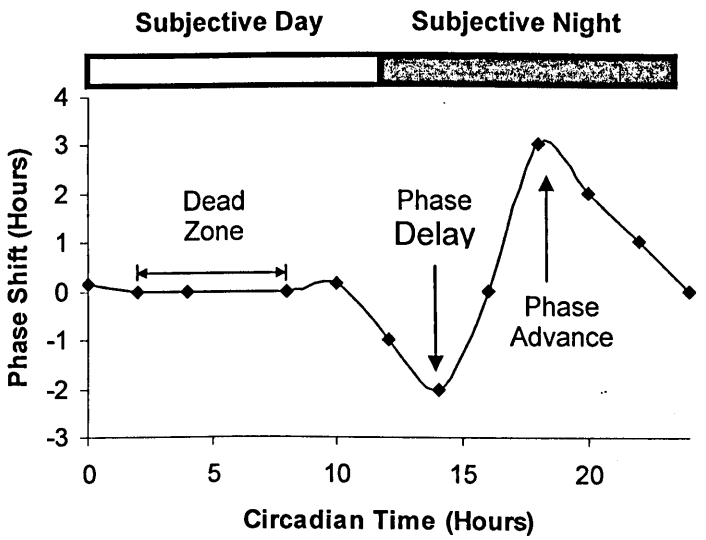


Fig 1.5. A typical Phase Response Curve for *Arabidopsis thaliana*. Following the curve through the 24 hour cycle, you find the 'dead zone' during the subjective day (CT 2-CT 8). Light pulses given at this time have little or no effect on the phase of the clock. As we move to the subjective night, (CT 10 – CT 14) a pulse of light delays the clock, in this case by a maximum of 2 hours at CT 14. The slope then reverses direction until it reaches a maximal phase advance in the middle of the subjective night. In this example, a light pulse given at CT 18 produces a phase advance of 3 hours. The curve falls back to the x intercept as the night progresses, eventually reaching the 'dead zone' the following day.

1.5.1. Light signalling targets

Elucidating the candidates required for light entrainment of the clock remains a challenge, although several proteins have been implicated. Light input is negatively regulated by ELF3 (as previously discussed); the *elf3* loss-of-function alleles are arrhythmic in light, but maintain rhythmicity in constant darkness (Hicks *et al.*, 1996). TIME FOR COFFEE (TIC) is another protein that appears to have an overlapping function with ELF3 in gating light input to the clock. Rhythmicity of clock outputs were abolished in *tic* mutants during the subjective morning (Hall *et al.*, 2003). The *elf3 tic* double mutant showed complete arrhythmia, suggesting gating of light by ELF3 in the night and TIC in the day (Hall *et al.*, 2003).

Light input is positively regulated by SENSITIVITY TO RED LIGHT REDUCED 1 (SRR1). *Srr1* mutants were defective in PHYB signalling, but also altered clock controlled expression of *TOC1* and *CCA1*, indicating a function independent of PHYB (Straiger *et al.*, 2003). The role of SRR1 with respect to the oscillator remains unclear.

A well characterised part of the input of light to the central oscillator involves the basic-helix-loop-helix protein transcriptional regulator PHYTOCHROME INTERACTING FACTOR3 (PIF3). PIF3 binds to the G-box located in the *LHY* and *CCA1* promoters, thereby inducing their expression (Martinez-Garcia *et al.*, 2000). The DNA-bound PIF binds to PHYB in its active form ((Martinez-Garcia *et al.*, 2000), providing a direct mechanism for light regulation of the negative arm of the central oscillator feedback loop. Interestingly, *pif3* knockouts do not result in alteration in period length of gene expression of *LHY* and *CCA1*. This would imply that other components are required for PIF3 function (Martinez-Garcia *et al.*, 2000).

1.5.2. Light Input

All organisms require methods of environmental perception in order to survive. *Plant*s rely on sunlight as their sole source of energy and being sessile, it is no surprise that they have acquired a sophisticated photoreceptor system. The photoreceptors perceive the external light cues and make adjustments that alter the developmental and physiological processes in a light-specific manner (Mathews, 2006). As chlorophyll has evolved to detect both red and blue absorption spectra, four kinds of photoreceptor are present in *plant*s; Phytochromes, Cryptochromes, Phototropins and a novel family of F-box proteins.

1.5.2.1 Red light photoreceptors

In *Arabidopsis* the red / far red light detecting phytochromes are encoded by five genes, *PHYA-E*. Phytochromes are dimers that consist of two identical apoproteins covalently linked to a linear tetrapyrolebilin compound that acts as a chromophore (Phytochromobilin). This can undergo a reversible polymerisation between two states depending upon the wavelength of light it receives. After initial assembly of the Phytochrome, it absorbs at 666nm (red light) and this is the biologically active form called Pr. Once red light is absorbed, the Phytochromobilin undergoes a conformational change that results in the Pfr form which interacts with *plant* proteins in the cytosol or nucleus (after translocation), therefore inducing a light response (Reviewed in Rockwell *et al.*, 2006). The conformational change is due to a *Z*-to-*E* photoisomerisation of the C15=C16 double bond in the tetrpyrole ring (current research is looking at the intermediate structural changes between Pr and Pfr). The ability of the Phytochromobilin to revert between the Pr and Pfr form allows it to act as a biological switch, turned on by absorption of red light and off by absorption of far-red light.

The Phytochromes have overlapping roles in *Arabidopsis.* However, they all show different molecular and spectral properties indicating many individual functions as well. For example, PHYA promotes seed germination and de-etiolation under far-red light (FR) and also promotes flowering. In stark contrast, PHYB delays flowering, but like PHYA promotes seed germination and de-etiolation, this time in response to red

light (R) (Reviewed in Thomas, 2006). The other three *Arabidopsis* Phytochromes also display distinct but overlapping functions.

1.5.2.2. Blue light Photoreceptors

The effect of blue light on *Plant* development has been understood since the beginning of the 19th century, but the photoreceptors responsible for blue light perception have only recently been deduced (Cashmore *et al.*, 1999; Briggs *et al.*, 2002). In *Arabidopsis*, the cryptochromes are encoded by two genes; CRYPTOCHROME 1 (CRY1) and 2 (CRY2) and were initially isolated by their role in the inhibition of stem elongation (Ahmad *et al.*, 1993; Lin *et al.*, 1998). Cryptochromes are flavoproteins that show a remarkable similarity to bacterial DNA lyases, and have also been found in mammals and flies, where they play an important role in blue-light photoreception and as crucial components of the circadian clock (van Gelder *et al.*, 2002). In *Arabidopsis*, it appears that CRY2 is the principle blue-light photoreceptor in the regulation of flowering time. The evidence for this stems from the late flowering phenotype of *cry2* mutants during long but not short days (Koornneef *et al.*, 1991; Guo *et al.*, 1998). Interestingly, the *cry2* mutant also has an effect on entrainment of the clock under white light (Más *et al.*, 2000) which would suggest some convergence between CRY2 and the phytochrome signalling pathway. Indeed, CRY2 has been shown to interact with PHYB where they co-localise to the nucleus (Más *et al.*, 2000). Additionally, the *cry* mutants show an altered entrainment under red light indicating that they are required in PHYA signalling to the clock (Devlin and Kay, 2001). So, under both red and blue light CRY2 regulates flowering time redundantly with CRY1 and PHYA.

The photoreceptor signalling pathways to the clock are not yet fully understood and the connections between the photoreceptors themselves require characterisation. Their largely overlapping roles make their function in the clock difficult to interpret. Furthermore, the *phyA phyB cry1 cry2* quadruple mutant still responds to the light signals that set the clock maintaining rhythmicity (Yanovsky *et al.*, 2000), suggesting that additional components are required for proper light signalling.

1.5.2.3. Phototropins and the F-box proteins

Phototropins 1 and 2 (Phot1 and Phot2) have an important function in blue-light phototrophic responses and are known to play a role in rapid processes including stomatal opening and chloroplast movement (Briggs and Christie, 2002). Both phototropins contain two light-oxygen-voltage (LOV) domains that non-covalently bind Flavin MonoNucleotide (FMN) in the dark. Blue light triggers the covalent attachment of the FMN molecule to each LOV domain, resulting in a conformational change that enhances the activity of a C-terminal Ser/Thr kinase domain (Christie, 2007; Tokutomi *et al.*, 2008). Although the phototropins autophosphorylate *in vitro*, the kinase targets *in planta* remain unidentified (Christie, 2007). As yet, the phototropins have not been implicated in controlling circadian processes. That said, the phototropin chromophore binding site (PAS/LOV domain) is highly similar to the PAS (period circadian protein, Ah receptor nulclear translocator protein and single-minded protein) signal sensor domain present in a novel family of F-box proteins that regulate flowering time and circadian rhythms in *Arabidopsis* (Nelson *et al.*, 2000; Klyosue *et al.*, 2000; Schultz *et al.*, 2001; Jarillo *et al.*, 2001).

The novel gene family consists of ZTL, FLAVIN-BINDING, KELCH REPEAT, F-BOX 1 (FKF1) and LOV, KELCH PROTEIN 2 (LKP2). These proteins all contain six kelch repeats (protein:protein interaction), a LOV domain (allows blue-light photoreception) and an F-box domain. The F-box motif is present in proteins that act as adapters, bringing substrates to ubiquitin protein ligase subunits for degradation via the proteasomal pathway (Craig and Tyers, 1999). Indeed, it has been shown that ZTL forms part of a Skp/Cullin/F-box (SCF) E3 ubiquitin ligase complex (Más *et al.*, 2003; Harman *et al.*, 2008). The unique combination of domains found in this family of proteins, implies involvement in light-dependant degradation of clock components. Clock proteins TOC1 and PRR5 are both substrates of ZTL and they are degraded via the SCF^{ZTL} pathway in a blue-light dependant manner (Más *et al.*, 2003; Harmon *et al.*, 2008). More recent studies have also shown a physical interaction between ZTL and GI upon photoexcitation (Kim *et al.*, 2007). The interaction stabilises both ZTL and GI and may prevent ZTL from interacting with its substrates TOC1 and PRR5 during the day, leading to decreased degradation of TOC1 and increased degradation during the night (Más *et al.*, 2003; David *et al.*, 2006; Kiba *et al.*, 2007;

Kim *et al.*, 2007; Fujiwara *et al.*, 2008). Recent research has added a further layer of complexity by showing that PRR3 directly binds to TOC1, leading to a decreased efficiency of ZTL-TOC1 binding (Para *et al.*, 2007). Thus, PRR3 is important in stabilising TOC1, preventing its degradation by inhibiting recruitment of TOC1 to the SCF^{ZTL} complex.

FKF1 indirectly promotes the expression of flowering regulator *CO* (Turck *et al.*, 2008). FKF1 has been shown to affect the stability of Cycling Dof Factor 1 (CDF1) which directly represses *CO* expression (Sawa *et al.*, 2007). Interestingly, FKF1 also associates with GI in a blue-light dependant manner. Only when GI-FKF1-CDF1 form a complex, is CDF1 degraded (Sawa *et al.*, 2007). This enables circadian regulation of *CO* so that expression is maximal during the afternoon in Long-day *plant*s (LDPs).

The third member of the family, LKP2 differs from ZTL and FKF1 in that it only contains a single LOV and PAS domain. *LKP2* overexpressing lines cause arrhythmia of the clock under constant conditions (Schultz *et al.*, 2001), but no substrates have yet been identified. Therefore, further studies will be required to deduce the function of LKP2.

It is apparent that we do not fully understand the roles of the photoreceptors and F-box proteins in the clock. It is particularly difficult to place the members of the ZTL family (that show such a diverse number of potential interactions) as input or central clock components. One important problem that remains to be addressed is role of F-box proteins in photo-entrainment. ZTL has been shown to interact with PHYB and CRY1, but the relevance of such interactions remains unexplored (Jarillo *et al.*, 2001).

1.5.3. Circadian regulation of light input

The circadian clock is able to control its own response to light by gating light input signals. The mRNA levels of all *PHYs* and *CRYs* are all under circadian control (Kozma-Bognar and Kaldi, 2008), although the protein levels themselves may not cycle. The exact control of photoreceptor protein levels are complex and remain unclear. CRY1, PHYB and PHYE levels are not rhythmically controlled (Kozma-Bognar and Kaldi, 2008), CRY2 is unstable in blue light (Sharrock and Clack, 2002), CRY2 and PHYA protein levels cycle in short but not long days (Yanovsky *et al.*, 2001) and there is evidence supporting a range of post-translational modifications including phosporylation, nucleo-cytoplasmic partitioning and nuclear translocation of the photoreceptors (Reviewed in Mockler *et al.*, 2003).

1.5.4. Temperature input

One of the key criteria for a circadian rhythm is that it has to be temperature compensated. Very little is known about how robust rhythms are maintained over a broad range of physiological temperatures in *plants*. Research in *Drosphila* has highlighted temperature-dependant mRNA splices and protein variants of the central clock protein PERIOD (PER) (Sawyer *et al.*, 1997). In *Neurospora*, FREQUENCY (FRQ) is a central clock component that appears in two isoforms. Mutants that lack either isoform are only rhythmic over a short range of temperatures (Liu *et al.*, 1997). Taken together, these studies suggest that temperature compensation is a fundamental property of core clock components. A study by Locke *et al*, looked for changes of gene expression over a wide range of temperatures for central clock components *TOC1*, *GI*, *LHY* and *CCA1*. Levels of *LHY* mRNA decreased with an increase in temperature and this was counter-balanced by an increase in the expression of *TOC1* and *GI* (Locke *et al.*, 2006). Conversely, *CCA1* mRNA was not affected at higher temperatures, but *CCA1* mRNA levels did increase under lower temperatures (Locke *et al.*, 2006). By combining mathematical modelling and the observation of many *gi* mutants over a wide range of temperatures, the authors' concluded; *LHY* and *GI* form an antagonistic counter-balance at higher temperatures to compensate the clock and at lower temperatures *LHY* is replaced by *CCA1*. This suggests a critical function of *GI*

in controlling temperature compensation of the clock, extending the range of temperature that can rhythmicity can be maintained (Locke *et al.*, 2006).

In addition to *GI* members of the PRR family have also been implicated in temperature compensation of the clock. The *prr7-3* and *prr9-1* mutants impaired the clocks response to thermocycles, with a *prr7-3 prr9-1* double mutant showing greater clock defect than either single mutant (Gould *et al.*, 2006). Importantly, the double mutant failed to reset the clock in response to temperature pulses and also failed to maintain rhythmicity in the dark; showing that these responses are light-independent (Gould *et al.*, 2006).

Despite these findings, very little is understood about temperature input and much work needs to be done. The implications of such research could one day help us modify or enhance the performance of *plant*s, allowing us to extend the geographical range conductive to growth.

1.6. Post-transcriptional regulation of the *Plant* clock

Phosphorylation is an important process in *Drosophila*, *Neurospora* and cyanobacteria clock functioning. At least one clock protein in each organism is phosphorylated before it is targeted for degradation. In *Arabidopsis*, the casein kinase CK2 has been shown to phosphorylate *CCA1 in vitro*. Overexpression of the CK2 subunit CKB3 results in short period circadian rhythms of several clock components (Sugano *et al.*, 1999). Furthermore, overexpression of a *CCA1* mutant that could not be phosphorylated resulted in a shortening of period (Daniel *et al.*, 2004), suggesting phosphorylation is required for CCA1 circadian function. Interestingly, CK2 can also phosphorylate LHY *in vitro*, suggesting a similar method of regulation to that of CCA1 (Sugano *et al.*, 1999).

PRR3 is also phosporylated by a circadian regulate kinase WITH NO LYSINE (WNK1), but the significance of this remains unclear (Murakami-Kojima *et al.*, 2002).

As previously discussed, the novel ZTL family of proteins are important in targeting a range of clock proteins for degradation. In addition to this family, two other proteins are required for light-input targeted degradation. DET1 and CONSTITUTIVELY PHOTOMORPHOGENIC 1 (COP1) negatively regulate light input to the clock. DET1 targets LHY for degradation (Song and Carré, 2005) and COP1 (an E3 ubiquitin ligase) is presumed to function in a similar manner. The *det1* and *cop1* single mutations result in a shortening of period in constant light (Millar *et al.*, 1995b; Song and Carré, 2005), although DET1 may also have a light-independent role close to or in the central oscillator (Song and Carré, 2005).

1.6.1. Unique role for LIGHT INSENSITIVE PERIOD 1 (LIP1)

LIP1 is the first GTPase that has been implicated in circadian clock in higher *plants*. GTPases are molecular switches that shuttle between a non-active GDP-bound state to an active GTP-bound state. GTPases are split into five sub-families based on structural and functional similarity. LIP1 represents a novel sub-family as it lacks certain fundamental properties of classic GTPases, including replacement of the catalytic glutamine$_{94}$ (Q_{94}) for a Histidine (H) (Kevei *et al.*, 2007). This base substitution results in a constitutively active state protein that still binds GTP. LIP1 is proposed to play a negative role in controlling circadian period that is suppressed by light in a fluence-rate dependant manner (Kevei *et al.*, 2007). The *lip1-1* mutant shows hypersensitivity to light pulses particularly in the first half of the subjective evening, resulting in large phase delays according to PRCs generated (Kevei *et al.*, 2007). It is not known how LIP1 negatively regulates the resetting of the clock. There are parallels with ELF3 and ZTL function; ELF3 attenuates light signalling in a similar way to LIP1 albeit at a later phase and ZTL regulates period length opposite to that of LIP1. However, the evidence suggests that LIP1 exerts its clock controlling properties by altering the distribution and or abundance of an unknown target (Kevei *et al.*, 2007).

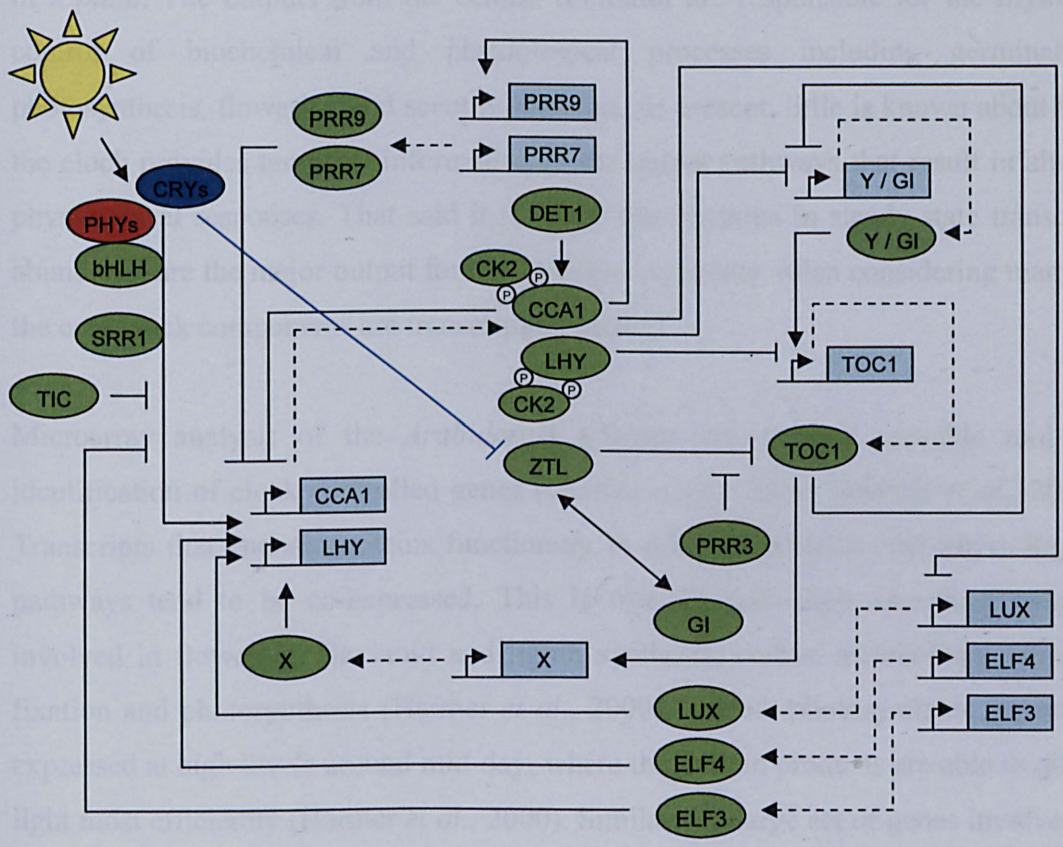


Fig. 1.6. Model of the *plant* circadian clock. This model is an overview of putative central oscillator components discussed. Blue rectangles indicate genes and green ovals represent proteins. Solid arrows show a positive regulatory effect with perpendicular lines indicating repression. Dashed lines between genes and proteins signal transcription / translation.
The well described LHY/CCA1 –TOC1 feedback loop is extended to include the mathematical model proteins 'X' and 'Y' (Locke *et al.*, 2005), as well as other loops including LUX, ELF4 and PRR7/PRR9. The red and blue light perceiving phytochromes and cryptochromes entrain the clock to the photoperiod by upregulating the expression of *LHY* and *CCA1*. Some proteins involved in light input are also included. Post-translational modification of clock components is important and the well defined interactions between degradation of TOC1 by ZTL and phosphorylation of CCA1 by CK2 are highlighted. Other proteins previously discussed are omitted for clarity.

1.7. Clock outputs

The clock regulates many developmental and cellular processes through the entire life of a *plant*. The outputs from the central oscillator are responsible for the rhythmic control of biochemical and physiological processes including germination, photosynthesis, flowering and scent production. At present, little is known about how the clock provides temporal information to the output pathways that result in altered physiological responses. That said it is likely that rhythms in steady-state transcript abundance are the major output for rhythmicity, especially when considering many of the core clock components are transcription factors.

Microarray analysis of the *Arabidopsis* genome has been a valuable tool for identification of clock controlled genes (Harmer *et al.*, 2000; Schaffer *et al.*, 2001). Transcripts that encode proteins functioning in related metabolic and physiological pathways tend to be co-expressed. This is true for transcripts encoding proteins involved in flowering, flavinoid and lignin synthesis, carbon metabolism, nitrogen fixation and photosynthesis (Harmer *et al.*, 2000). Indeed, photosynthetic genes are expressed at high levels around mid-day, where the protein products are able to utilise light most efficiently (Harmer *et al.*, 2000). Similarly, a large set of genes involved in the biosynthesis of photo-protective pigments peak co-ordinately just before dusk, when they will be required to protect the *plant* from damage by UV-B light (Schaffer *et al.*, 2001). The genes involved in auxin-transport and cell elongation also show co-regulation, as they are maximally expressed during the subjective afternoon where elongation rates are at their greatest (Harmer *et al.*, 2000). Collectively, analysis of gene expression suggests that many of the physiological processes controlled by the clock are a result of rhythmic transcript abundance. The challenge is to now elucidate the regulatory gene networks and how they are connected to physiology and development.

Although the exact mechanisms and gene networks for circadian control of development are not known, we do have information which implicates the clock from the earliest stage of development to the later stages of flowering and reproduction.

The circadian clock may play a role in the earliest life stage of a *plant* as germination is controlled by day length (Baskin and Baskin, 1976; Densmore, 1997). The fact that photoperiod is a requirement for germination in some species indicates the functioning of the clock in seedlings. Evidence that supports this notion includes, imbibition (absorbance of water) by *Arabidopsis* seedlings synchronising circadian-controlled gene expression (Zhong *et al.*, 1998) and gas exchange in *Allium cepa* seedlings maintain rhythmicity in darkness (Bryant, 1972).

During *plant* growth the circadian clock partially controls the elongation of hypocotyls (Dowson-day and Millar, 1999), cotyledon and leaf movements (Englemann and Johnson, 1998) and stem circumnutations (Niinuma *et al.*, 2005). In fact, these processes have been useful in the identification of putative clock components. In addition, the clock has been show important in shade-avoidance responses that result in stem and petiole growth when a *plant* is in competition for light (Niinuma *et al.*, 2005).

Parts of the reproductive *plant* cycle are also controlled by the clock. As previously discussed, several components of the core oscillator are important in controlling rhythmic expression of the floral inducer *CO* which results in the switch from vegetative growth to flowering. Interestingly, pollination is also likely to be partially controlled by the clock. In *Arabidopsis*, the flowers open in the morning when pollinators are most active and close in the evening maximising damage limitation (van Doorn and van Meeteren, 2003). *Plant*s also release their scent so that it concurs at the same time pollinators are likely to be present (Overland, 1960) thereby increasing the chances of successful pollination.

1.8. Gaps in our Knowledge

The current model of the circadian clock is primarily based on genetic screens which can only postulate at a particular function within the clock. As many of the inputs and outputs from the clock overlap with the central oscillator function it is difficult to determine where some of the putative oscillator proteins function in the regulation of the clock. Only some biochemical interactions have been elucidated at the protein level and this is an area which needs to be addressed. To build a complete picture, the interactions of the proteins with each other and with the genes of the central oscillator need to be deciphered as genetic screens are reaching saturation in the identification of clock components.

One major problem is that several of the key regulators of the central oscillator are *plant*-specific proteins that are novel in structure. Positive *LHY* and *CCA1* regulators ELF3, ELF4 and GI contain no known domains based on their primary amino acid sequence. In short, how do they perform their perceived functions? Do they form individual feedback loops with LHY and CCA1, or is there some convergence of these loops? As these proteins are expressed at a similar circadian phase, it would seem logical that they may form protein complexes with each or with as yet unidentified clock components. This is particularly likely for TOC1 which contains no known DNA binding domain and must therefore regulate *LHY* and *CCA1* indirectly.

Looking at these proteins from a structural stand-point may be beneficial in several ways. Firstly, by obtaining x-crystallographic data we can infer protein function based on active sites and domains discovered. This may help us determine specific functions and interactions of the clock proteins. More interestingly though, is the possibility of identifying novel folds and domains, which could have implications for understanding protein function beyond the limits of the clock network.

1.9. Aims

This aim of this thesis is to address the current gap in the knowledge of circadian protein function by a structural analysis. The initial aim is to attempt to express the circadian related proteins in *E. coli* and to determine which of these proteins express well enough for purification. The subsequent aim is to develop a purification protocol that will allow the production of mg quantities of protein for use in crystallography screening. Particular emphasis is placed on obtaining the positive regulators of *LHY* and *CCA1*; TOC1, ELF3, ELF4 and GI. These proteins are of unknown function and represent *plant*-specific novel structures, making them a highly desirable target for crystallisation and structural analysis.

Chapter 2. Bioinformatics

2.1. Introduction

The majority of our understanding of circadian clock associated proteins has come from genetic based evidence. This has led to a complicated picture of possible protein functions and interactions, but has yet to elucidate protein functionality at a biochemical level. Whilst the last several years have shed light into some interactions, the specific details of these remain largely unknown. To gain insight into the workings and importance of these proteins it is useful to employ a range of computational techniques. This approach not only provides insight into the structure of the proteins but also highlights the interest and potential importance of how a structurally based project could help our current understanding of the circadian clock.

In addition to the standard tools used in bioinformatics, one recent publication has provided a significant advance in the prediction of protein structure (Kelley *et al.*, 2009). The protein homology / analogy recognition engine (Phyre) is similar to other protein prediction tools in that it uses algorithms that match sequences based on three- dimensional (3D) structures that have been experimentally elucidated. This allows the researcher to build their own model based on existing knowledge that the number of folds found in nature is limited, and that remotely homologous sequences tend to adopt similar structures (Baker and Sali *et al.*, 2001). Although prediction of structure based on primary sequences remains the 'holy grail' of bioinformatics, the Phyre server provides a user-friendly interface that improves prediction and interpretation over pre-existing programmes.

The Phyre server uses the library of solved protein structures contained in the Structural Classification of Proteins (SCOP) (Murzin *et al.*, 1995) database, as well as the more recent structures deposited in the Protein Data Bank (PDB) (Berman *et al.*, 2000). The user submits a sequence (query) which is scanned against a non-redundant sequence database, creating a profile. This profile is then queried against three independent secondary structure programmes; Psi-Pred (McGuffin *et al.*,

2000), SSPro (Pollastri *et al.*, 2002) and JNet (Cole *et al.*, 2008), which show residues involved in the formation of α-helices, β-sheets and coils with confidence scores given for each residue. The secondary structure is then scanned against a fold-library using an algorithm that matches profile-profile alignments (Bennett-Lovesey *et al.*, 2008). The algorithm returns a score on which alignments are ranked, returning the top ten fitted alignments fitted to an extreme value distribution (E-value). The top ten hits are then used to construct models of the query sequence.

It is in the interpretation of the data that the Phyre server really aids when trying to postulate protein structure and function. For a full review on interpreting the outputs from the server, see Kelley *et al.*, 2009.

This chapter focuses on a combination of bioinformatic techniques to gain insight into the circadian clock proteins. There is no hard and fast method of summarising the findings and so effort has been made to condense the results and to highlight uncertainty. This chapter also breaks from convention, by combining results and discussions as appropriate. In addition, as the Phyre server returns matches of known protein families, a considerable amount of information regarding the background of such families is included. Whilst some of this could have been presented in the main introduction, it appears more sensible to keep the structural analysis separate but provide sufficient information to aid with explanation of the results.

The Biophysical properties of the circadian-associated proteins are highlighted in Table 1. These are used as references through-out the thesis.

Protein	Number of amino acids	Mw (Da)	Isoelectric point (pI)	Molecular Extinction coefficient	1g / L	Instability Index	GRAVY
TOC1	618	69,195	7.13	44,880	0.649	51.67	-0.818
TOC1-PRR	312	34,690	6.14	29,520	0.851	34.57	-0.297
LHY	645	70,425	5.77	50,670	0.719	48.03	-0.741
CCA1	608	66,976	5.70	48,700	0.727	60.51	-0.933
LUX	323	34,649	5.47	25,900	0.747	47.12	-0.790
ELF3	695	77,205	8.63	36,060	0.467	62.32	-0.918
ELF4	111	12,376	8.16	6,970	0.563	57.96	-1.046
GI	1173	127,876	6.60	158,050	1.236	48.96	-0.053
SPY	914	101,430	5.98	108,400	1.069	37.75	-0.214
ZTL	609	65,906	5.44	99,110	1.504	38.33	-0.079
LKP2	611	66,351	5.16	100,840	1.52	38.37	-0.091
FKF1	619	69,116	6.06	90,045	1.303	49.63	-0.236
SRR1	275	31,474	4.97	37,530	1.192	61.01	-0.436
LIP1	342	37,741	8.97	43,810	1.161	55.29	-0.517

Table 1. Biophysical properties of clock-associated proteins. Gene sequences were obtained from PubMed and then submitted for BLASTp (Altschul *et al.*, 1997) at the ExSPASy server. Outputs were linked to ProtParam (Gasteiger *et al.* 2005) and the information provided in the table is taken from this prediction.

2.2. *Plant* response regulators

In prokaryotes, two component signalling systems provide a quick and effective method of signal transduction through a large number of cells. This system is based a His-Asp phosphorelay system involving a histidine protein kinase (HK) and a response regulator (RR). The HK acts as a sensor for an input, which when stimulated phosphorylates its cognate RR. The RR is then responsible for modulating gene expression and / or any other cellular output. This process depends on the transfer of ATP onto a conserved histidine, serine, threonine or tyrosine, and then subsequent phosphorylation. Additionally, a histidine containing phosphor-transmitter (HPt) is frequently employed as an intermediate in the phosphorelay between the HK and RR domains (Fig 2.1).

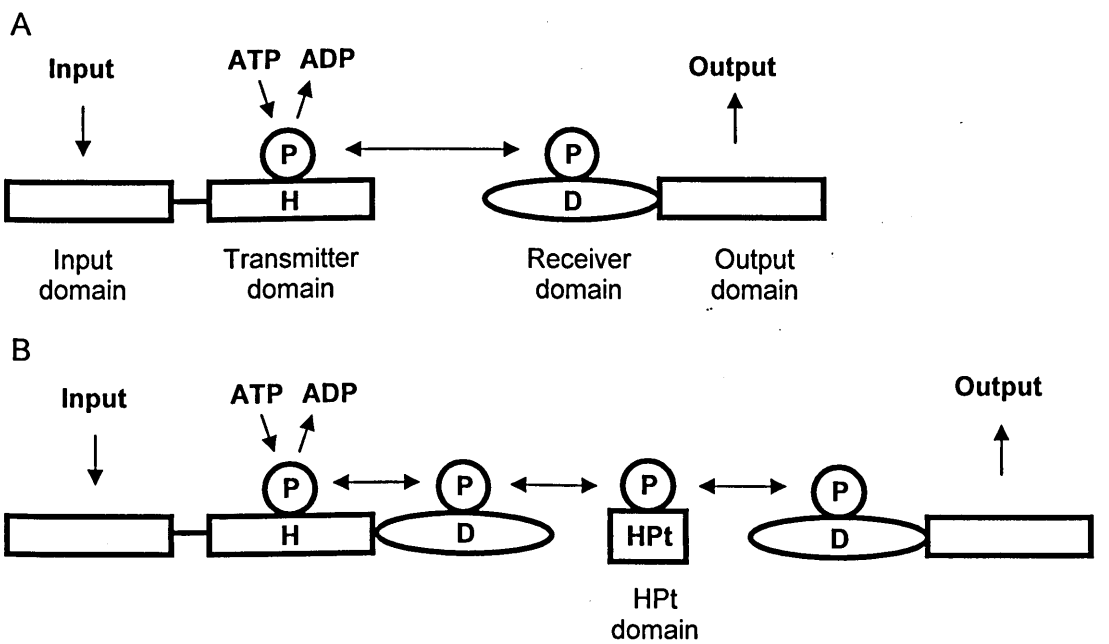


Fig 2.1. A schematic overview of classic two component systems. (A) The canonical phosphorelay pathway. The input domain senses the stimulus which induces autophosphorylation of a conserved histidine (H) in the presence of ATP. The phosphoryl group (P) is transferred to the receiver domain to a conserved Asp residue (D), which drives output pathways. (B) A two component signalling system, where the input domain, transmitter domain and response receiver domain are part of the same polypeptide. In this instance, HPt domain is phosphorylated and acts as an intermediate between the hybrid histidine kinase domain and another response regulator domain.

Initially, it was suspected that these TCS were specific to prokaryotes, as it differs so widely from known eukaryotic signal transduction pathways. However, many instances of phosphorelay systems have been found in eukaryotes with special focus in the higher *plant Arabidopsis thaliana*. In *Arabidopsis*, the response receiver domain (ARR) contains an invariant Asp phospho-accepting residue and these can be classified into three major sub-groups based on structural similarities; type-A ARRs, type-B ARRs and pseudo response regulators (APRRs). The APRRs are distinct from other RRs in that they do not contain the essential phosphor-accepting Asp site. Indeed, it was assumed that the APRRs do not undergo phosphorylation and therefore do not act in typical His-Asp phosphorelay pathways. Recent evidence suggests that the APRR 7/5/3 and TOC1 are phosphorylated *in vivo*, with the degree of phosphorylation dependant upon circadian period and therefore linked to function; enhanced binding or targeted degradation (Fujiwara *et al.*, 2008).

In addition to the RR domain, the *plant* ARRs are classified according to their C-terminal extensions (Fig 2.2). Whilst the Type-A ARRs only contain a short C-terminal extension, the Type-B ARRs consist of a functional GARP (maize Golden 2, ARR; *Arabidopsis* response regulators and Psr1; phosphorus stress response1 from *Chlamydomonas*) domain. The GARP domain is involved in DNA binding and is distantly related to the MYB-domain transcription factors, including LHY / CCA1. Finally, the PRRs have a conserved CCT (CONSTANS, CONSTANS-Like, TOC1) motif at their C-terminus, which is typical of the CONSTANS-like transcription factors; rich in basic residues and containing a nuclear localisation sequence (NLS). Downstream from the CCT, is a stretch of acidic residues that may be a transcriptional activation domain, suggesting that these proteins may be acting as nuclear transcription factors (Strayer *et al.*, 2000).

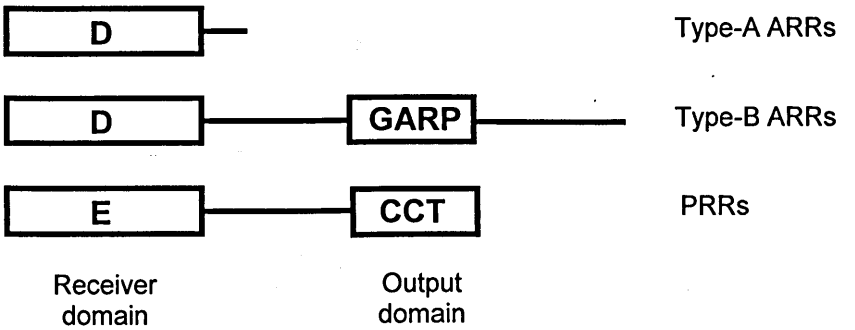


Fig. 2.2. Schematic representation and classification of *Plant* response regulators (ARRs). The Pseudo-response regulators (PRR) contain a phospho-accepting glutamate (E) instead of aspartate (D) in their response receiver domain. Type-A ARRs contain only a modest C-terminal extension, whilst Type-B ARRs contain a GARP domain and the PRRs contain the CCT motif.

2.2.1 TOC1 and the PRR family

As previously discussed, the role of the PRR proteins in the circadian clock is very difficult to interpret. As TOC1 was the first PRR protein to be described and its role in the circadian clock is arguably the best understood, we focus on this protein as a reference for elucidating how the other PRR's may function. The PRR's have a very high degree of sequence similarity as shown in a ClustalW alignment (Thompson *et al.*, 1994) overleaf (Fig 2.3). The pseudo response receiver domain and the CCT are indicated in boxes, with the conserved E (Glu68 with reference to TOC1) highlighted in red. The high sequence similarity in these regions suggests that the PRR proteins are acting in similar processes, even if they act on different targets.

At the start of this body of work, the focus was only on TOC1. As previously discussed, TOC1 has no known DNA binding motif therefore suggesting other proteins are required for its perceived function. Previous attempts to express full length TOC1 were unsuccessful, so work focused on the expression and purification of the truncated TOC1-PRR domain (residues 1-145). Full length TOC1 submission to the Phyre server results in a list of ten response receiver domains that are termed CheY-like, with E-values > $x10^{-24}$ and estimated precision of 100 %. Indeed, submission of the PRR domain retrieves the exact same list, albeit increased sequence identity.

```
PRR3    ------------------------------------------------------------
PRR7    ------------------------------------------------------------
PRR5    MWQTWPRQPILLDIFSNPNTLSTTVRSWSVRHPLSIITVKTFARFFLDIFFSPHYYRKNK 60
TOC1    ------------------------------------------------------------

PRR3    ----------------MCFNNIETGDEVETERQVFGSSEED--EFRVEDTARN--TNNV 39
PRR7    ----------------MNANEEGEGSRYPITDRKTGETKFDRVESRTEKHSEEEKTNGI 43
PRR5    VLFFALFSFISPLTNILICFVTVSLSLELSSSSSIIDLGFSKLSVCVVIMTSSEEVVEVT 120
TOC1    ------------------------------------------------------------

PRR3    QISQQQ-------------QQPLAHVVKWERYLPVRSLKVLLVENDDSTRHIVTALLKN 85
PRR7    TMDVRNGSSGGLQI----PLSQQTAATVCWERFLHVRTIRVLLVENDDCTRYIVTALLRN 99
PRR5    VVKAPEAGGGKLSRRKIRKKDAGVDGLVKWERFLPKIALRVLLVEADDSTRQIIAALLRK 180
TOC1    -------------------MDLNGECKGGDGFIDRSRVRILLCDNDSTSLGEVFTLLSE 40
                                  : ::    :::** : *. :   : :** :

PRR3    CSYEVTAVPDVLEAWRILEDEKSCIDLVLTEVDMPVHSGTGLLSKIMSHKTLKNIPVIMM 145
PRR7    CSYEVVEASNGIQAWKVLEDLNNHIDIVLTEVIMPYLSGIGLLCKILNHKSRRNIPVIMM 159
PRR5    CSYRVAAVPDGLKAWEMLKGKPESVDLILTEVDLPSISGYALLTLIMEHDICKNIPVIMM 240
TOC1    CSYQVTAVKSARQVIDALNAEGPDIDIILAEIDLPMAKGMKMLRYITRDKDLRRIPVIMM 100
        ***.*. . .  :.   *:     :*::*:*: :*  .*  :*  *  ..  :.******

PRR3    SSHDSMVLVFKCLSNGAVDFLVKPIRKNELKNLWQHVWRRCHSS-----SGSGSESGIHD 200
PRR7    SSHDSMGLVFKCLSKGAVDFLVKPIRKNELKILWQHVWRRCQSS-----SGSGSESGTHQ 214
PRR5    STQDSVNTVYKCMLKGAADYLVKPLRRNELRNLWQHVWRRQTSL-----APDSFPWNESV 295
TOC1    SRQDEVPVVVKCLKLGAADYLVKPLRTNELLNLWTHMWRRRRMLGLAEKNMLSYDFDLVG 160
        * :*.:  * **:  **.*:****:* ***  ** *:***            .   .

PRR3    -KKSVKPESTQGSENDASISDEHRNESGSSGGLSNQDGGSDNGSGTQSSWT--KRASDTK 257
PRR7    TQKSVKSKSIKKSDQDSGSSDE--NENGSIG-LNASDGSSD-GSGAQSSWT--KKAVDVD 268
PRR5    GQQKAEGASANNSNGKRDDHVVSGNGGDAQSSCTRPEMEGESADVEVSARD--AVQMECA 353
TOC1    SDQSDPNTNSTNLFSDDTDDRSLRSTNPQRGNLSHQENEWSVATAPVHARDGGLGADGTA 220
         .:.    .       .         . .   .  .  :  . .     :

PRR3    STS-----------------------PSNQFPDAPNKKGTYENG---------------- 278
PRR7    DSPRAVSLWDRVDSTCAQVVHSNPEFPSNQLVAPPAEKETQEHDDKFEDVTMGRDLEISI 328
PRR5    KSQFNETRLLANELQSKQAEAIDFMGASFRRTGRRNREESVAQYES-------------- 399
TOC1    TSSLAVTAIEPPLDHLAGSHHEPMKRNSNPAQFSSAPKKSRLKIGESSAFFTYVKSTVLR 280
         :                            *          : :  :

PRR3    CAHVNRLKEAEDQKEQIGTGSQTG-------------------MSMSKKAEEPGDLEKNA 319
PRR7    RRNCDLALEPKDEPLSKTTGIMRQDNSFEKSSSKWKMKVGKGPLDLSSESPSSKQMHEDG 388
PRR5    RIELDLSLRRPNASENQSSGDRPS-------------------LHPSSASAFTRYVHRPL 440
TOC1    TNGQDPPLVDGNGSLHLHRGLAEKFQVVASEG--------INNTKQARRATPKSTVLRTN 332
            :      :       *                            :  :     : .

PRR3    KYSVQALERNNDDTLNRSSGNSQVESKAPSSN----REDLQSLEQTLKKTR---EDRDYK 372
PRR7    GSSFKAMSSHLQDNREPEAPNTHLKTLDTNEASVKISEELMHVEHSSKRHRGTKDDGTLV 448
PRR5    QTQCSASPVVTDQRKNVAASQDDNIVLMNQYNTSEPPPNAPRRNDTSFYTGADSPGPPFS 500
TOC1    GQDPPLVNGNGSHHLHRGAAEKFQVVASEGINNTKQAHRSRGTEQYHSQGETLQNGASYP 392
          .        ..  .  : :                          :.          .

PRR3    VGD-RSVLRHSNL-SAFSKYNNGATSAKKAPEENVESCSPHDSPIAKLLG---------- 420
PRR7    RDD-RNVLRRSEG-SAFSRYN-PASNANKISGGNLGSTSLQDNNSQDLIKKTEAAYDCHS 505
PRR5    NQL-NSWPGQSSYPTPTPINNIQFRDPNTAYTSAMAPASLSPSPSSVSPHEYSSMFHPFN 559
TOC1    HSLERSRTLPTSMESHGRNYQEGNMNIPQVAMNRSKDSSQVDGSGFSAPNAYPYYMHGVM 452
            ..    :.  :     :    .            *   .

PRR3    --------------------SSSSSDNPLKQQ------SSGS----------------- 436
PRR7    NMNESLPHNHRSHVGSNNFDMSSTTENNAFTKPGAPKVSSAGSSSVKHSSFQPLPCDHHN 565
PRR5    SKPEGLQDRDCSMDVDERRYVSSATEHSAIGNHIDQLIEKKNEDGYSLSVG--------- 610
TOC1    NQVMMQSAAMMPQYGHQIPHCQPNHPNGMTGYPYYHHPMNTSLQHSQMSLQN-------- 504
                                 ..   ..           . .
```

```
PRR3      ------------------------------------------------------------
PRR7      NHASYNLVHVAERKKLPPQCGSSNVYNETIEGNNNTVNYSVNGSVSGSGHGSNGPYGSSN 625
PRR5      ------------------------------------------------------------
TOC1      ------------------------------------------------------------

PRR3      -------------------------------------DRWAQREAALMKFRLKRKERCF 458
PRR7      GMNAGGMNMGSDNGAGKNGNGDGSGSGSGSGSGNLADENKISQREAALTKFRQKRKERCF 685
PRR5      ----------------------------------KIQQSLQREAALTKFRMKRKDRCY 634
TOC1      --------------GQMSMVHHSWSPAGNPPSNEVRVNKLDRREEALLKFRRKRNQRCF 549
                                               ::  :** ** *** **::**:

PRR3      EKKVRYHSRKKLAEQRPHVKGQFIRK-----RDDHKSGSEDN------------------ 495
PRR7      RKKVRYQSRKKLAEQRPRVRGQFVRKTAAATDDNDIKNIEDS------------------ 727
PRR5      EKKVRYESRKKLAEQRPRIKGQFVRQ---------VQSTQAP------------------ 667
TOC1      DKKIRYVNRKRLAERRPRVKGQFVRKMNGVNVDLNGQPDSADYDDEEEEEEEEEEENRDS 609
           **:** .**:***:**:::***:*:                 .  .

PRR3      ---------
PRR7      ---------
PRR5      ---------
TOC1      SPQDDALGT 618
```

Fig 2.3. ClustalW alignment of the PRR proteins. Highlighted in boxes are the PRR domains (residues 1-145) and the CCT domains (residue 528-576) with reference to TOC1 sequence. Also shown in red is the conserved phospo-accepting E that characterises this family of ARRs. Stars and semi-colons represent conserved and semi-conserved amino acids respectively. PRR9 was omitted from the alignment due to differences in the sequences.

CheY from *E. coli* is the paradigm member of the bacterial single domain response regulators (lacking the output domain) and is involved in regulation of chemotaxis through regulation of information between chemoreceptors and the flagellar switch (motility). Chemo-effector concentrations are detected by specific methyl-accepting chemotaxis (MCP) proteins that assemble with the histidine phosphate kinase (HPK) protein, CheA and the coupling protein CheW. These accumulate as signal transduction 'clusters' at the poles of the bacterial cell, where CheA mediates phosphorylation of CheY, which increases CheY affinity for the FliM component of the flagellar switch complex. This results in a change in flagellar rotation from anti-clockwise to clockwise (Bren and Eisenbach, 1998) and therefore the bacterial cell moves in the desired direction. (Reviewed in Jenal and Galperin, 2009).

The RR domain of CheY contains a conserved (βα) 5 fold that catalyzes the transfer of a phosphoryl group from CheA to its own aspartic acid residues located in the acidic pocket. This results in a structural change from inactive to active where there is a small conformational change on the α4-β5-α5 RR domain surface (Kern *et al.*,

1999; Volkman *et al.*, 2001). This structural change mediates effector domains directly or as dimers (depending on the CheY-like protein), and have also been shown to engage in downstream effector protein binding (Dyer *et al.*, 2006).

Single RR domains are found in all domains of life, with many homologs belonging to the family. Although the above example illustrates control of bacterial motility, the many members of this abundant family are thought to act on numerous spatially separated downstream targets, either directly or in signal transduction cascades. Many may even work as molecular switches in protein:protein interaction networks, through phosphorylation dependant remodelling of the α4-β5-α5 RR domain surface. This may be to alter the sub-cellular localisation or activity of downstream targets. This is particularly interesting with respect to TOC1 function, which may be activating or localising other circadian proteins such as ELF4 / GI in order to up-regulate the transcription of the Myb-domain proteins LHY and CCA1.

The degree of sequence similarity of TOC1 and CheY from *E. coli* is shown as an alignment (Fig 2.4). There is remarkable sequence similarity between the two and it is likely that TOC1 is able to form the α4-β5-α5 RR domain surface that appears so important in CheY-like protein function (the residues involved in the formation of the α5 helix in CheY-like proteins, are not always conserved). The crystal structure of CheY indicates that there are at least three key residues to maintain the active confirmation. Firstly, Asp57 which is the phosphorylation site, bonds to BeF_3^- (which mimics phosphate) forming a hydrogen bond with Thr87 and a salt bridge with Lys109. The OH group of Thr87 also forms a hydrogen bond with the side chain of Tyr106 (Lee *et al.*, 2001). This allows the correct orientation of the α4-β5-α5 surface. TOC1 contains semi conservative substitutions with respect to CheY; Asp57 is replaced with Glu71 and Thr87 with Ser101. The Tyr106 (CheY) does appear to be conserved, suggesting that the active sites are similar enough to infer functionality. Questions that will not be answered by structural analysis are the exact targets of TOC1 and indeed, other PRR domain proteins highlighting the need for a combination of structural and biochemical characterisation of this interesting family of proteins.

```
TOC1      MDLNGECKGGDGFIDRSRVRILLCDNDSTSLGEVFTLLSECSY-QVTAVKSARQVIDALN 59
CheY      -------------MADKELKFLVVDDFSTMRRIVRNLLKELGFNNVEEAEDGVDALNKLQ 47
                       :  ..:::*: *: **    * .**.* .: :*  .:.. :.:: *:

TOC1      AEGPDIDIILAEIDLPMAKGMKMLRYITRDKDLRRIPVIMMSRQDEVPVVVKCLKLGAAD 119
CheY      AGG--YGFVISDWNMPNMDGLELLKTIRADGAMSALPVLMVTAEAKKENIIAAAQAGASG 105
          * *   .::::: ::*  .*:::*: *  *  :  :**:*:: : :   :: . : **:.

TOC1      YLVKPLRTNELLNLWTHMWRRRRMLGLAEKNMLSYDFDLVGSDQSDPNTNSTNLFSDDTD 179
CheY      YVVKPFTAATLEEKLNKIFEKLGM---------------------------------- 129
          *:***: :  * :  .::::.:  *
```

Fig 2.4. ClustalW alignment of the PRR domain from TOC1 with CheY. The archetypal bacterial phosph-accepting Asp57 of CheY from *Escherichia Coli* is highlighted in red, with the semi-conservative substitution of Glu68 in TOC1 also in red. The α4-β5-α5 RR domain surface of CheY is indicated in blue (β4), purple (α5) and blue (β5) respectively. Stars and semi-colons represent conserved and semi-conserved amino acids respectively.

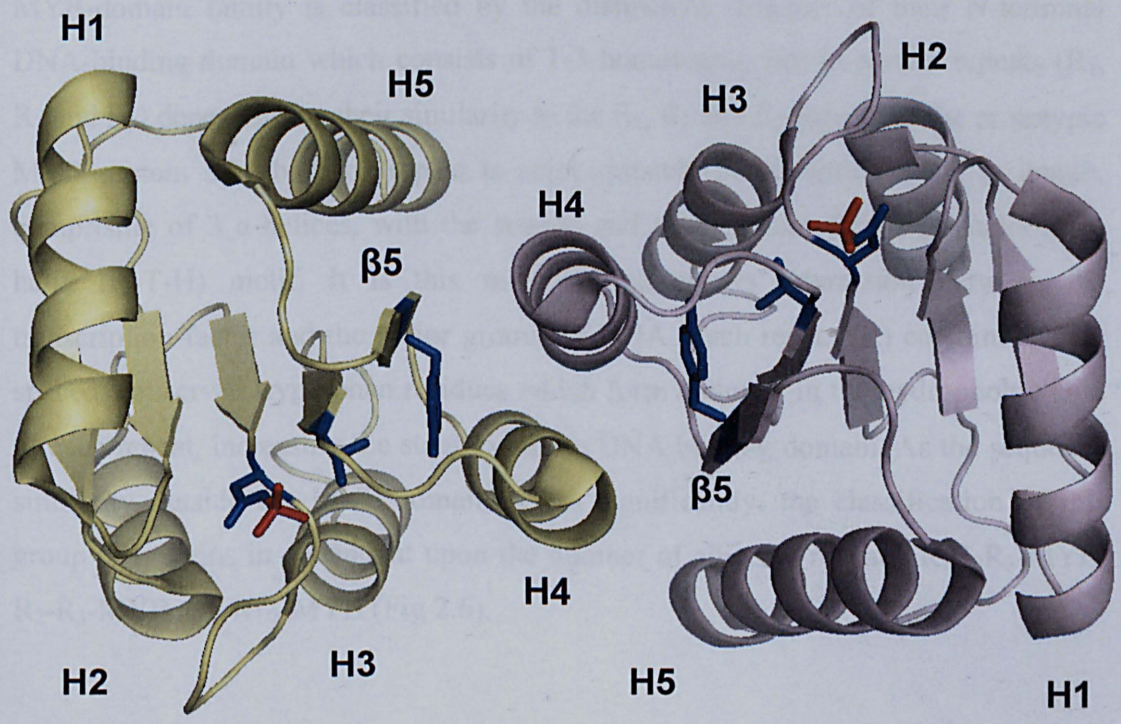


Fig 2.5. Ribbon diagram of two BeF3- activated CheY response receiver domains. The source of the Che Y protein was *E. coli* (PDB, 1FQW) and the crystal structure was determined at 2.4 Å resolution (Lee *et al.*, 2001). Shown is the dimmer crystallographic asymmetric unit, which is how the protein crystallised as CheY is not a dimmer under physiological pH. The active sites are directed toward the reader. The residues Asp57, Thr87 and Tyr106 involved in the active site are shown in blue and the BeF_3^- moiety which complexes with receiver domains, mimicking phosphorylation-activated states, is shown in red. This diagram was drawn with the Pymol software package.

2.3. MYB-domain transcription factors

MYB domain proteins are widely distributed in higher *plant*s and comprise one of the largest groups of transcription factors. A fundamental requirement of a transcription factor is the presence of a DNA binding domain, and the MYB-domain proteins are characterised by a highly conserved DNA-binding domain at their N-termini. The first *plant* MYB gene identified was the c-MYB-like transcription factor (gene *C1* isolated from *zea mays*) which is involved in the biosythesis of anthocyanin (Paz-Ares *et al.*, 1987). In the past twenty years, many more MYB-domain transcription factors have been identified and studied, indicating diverse roles in physiological and biochemical process from cell cycle control, seed and floral development and control of cell morphogenesis (Reviewed in Du *et al.*, 2009). The MYB-domain family is classified by the distinctive structure of their N-terminal DNA-binding domain which consists of 1-3 homologous but imperfect repeats (R_1, R_2 and R_3) dependant on their similarity to the R_1, R_2 and R_3 repeats of the prototypic MYB protein c-Myb. Each repeat is approximately 50-53 amino acids in length, comprising of 3 α-helices, with the second and third helices forming a helix-turn-helix (H-T-H) motif. It is this motif which allows interaction between the transcription factor and the major groove of DNA. Each repeat (R) contains evenly spaced, conserved tryptophan residues which form a cluster in the hydrophobic core of each repeat, increasing the stability of the DNA binding domain. As the sequence similarity outside the MYB-domain varies significantly, the classification of this group of proteins in dependant upon the number of adjacent repeats: $R_1R_2R_3$-MYB, R_2-R_3-MYB and $R_{1/2}$-MYB (Fig 2.6).

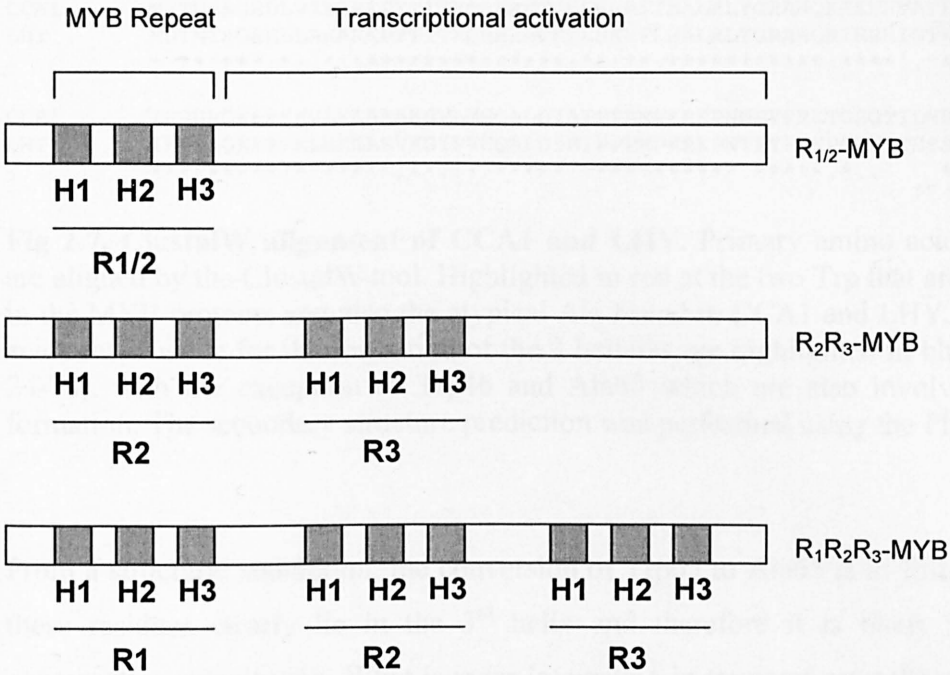


Fig 2.6. Schematic representation of the classification of MYB-domain transcription factor family. Shaded boxes represent α-helices labelled 1-3. A repeat of the DNA-bind domain containing 3 helices is labelled R. The MYB-domain proteins are grouped according to the number and type of repeats (shown to the right). Non-shaded areas represent protein specific domains associated with their transcriptional activity, and this is reflected in the diversity of the subfamilies members.

2.3.1 LHY / CCA1

LHY and CCA1 both belong to the $R_{1/2}$ subfamily of MYB-domain proteins. Unsurprisingly, they show high sequence similarity, especially towards the N-termini of the protein (Fig 2.7). They both contain a basic region Lys13 to Lys107 which contains the highly homologous MYB-domain sequence. In their MYB-repeat they contain the majority of the residues important for forming the hydrophobic core, which provides stability and maintenance of the H-T-H allowing binding of the domain to the major groove in DNA (Ogata *et al.*, 1992). Interestingly, only two of the three conserved tryptophans are conserved in LHY and CCA1 (Trp27 and Trp46), with the third replaced by alanine (Ala65). In terms of actual contact with the DNA, bases usually required are not present in either protein. Both proteins contain the SHAQKYF sequence which is *plant* specific and can be used as a means of classifying the single $R_{1/2}$ domains found in *plant* MYB proteins.

```
CCA1      METNSSGEDLVIKTRKPYTITKQRERWTEEEHNRFIEALRLYGRAWQKIEEHVATKTAVQ 60
LHY       MDTNTSGEELLAKARKPYTITKQRERWTEDEHERFLEALRLYGRAWQRIEEHIGTKTAVQ 60
          *:**:***:*: *:****************:**:**************:****:.******

CCA1      IRSHAQKFFSKVEKEAEAKGVAMGQALDIAIPPPRPKRKPNNPYPRKTGSGTILMSKTGV 120
LHY       IRSHAQKFFTKLEKEAEVKGIPVCQALDIEIPPPRPKRKPNTPYPRKPGNNGTSSSQVSS 120
          *********:*:*****.**:.: ***** ************.*****.*..    *:..
```

Fig 2.7. ClustalW alignment of CCA1 and LHY. Primary amino acid sequences are aligned by the ClustalW tool. Highlighted in red at the two Trp that are conserved in the MYB proteins and also the atypical Ala found in CCA1 and LHY. The amino acids responsible for the formation of the 3 helicies are highlighted in blue (residues 24-78), with the exception of Trp46 and Ala65 which are also involved in helix formation. The secondary structure prediction was performed using the Phyre server.

From a structural standpoint, the conversion of Trp65 to Ala65 is of little interest as these residues clearly lie in the 3rd helix and therefore it is likely that it is a conservative substitution. What is more interesting in terms of crystallography is the C-terminal (not shown). Secondary structure predictions using CCA1 and LHY sequences minus residues 1-80, show no similarity to any other protein. We are already aware that the N-terminal accounts for their perceived DNA binding activities, to EE in the promoters of clock outputs and components of the central loop itself, for example TOC1. However, the C-terminal may be involved in other interactions through as yet unidentified mechanisms.

When we perform a BLASTp on LHY, the results are a list of proteins termed LHY homologues, putative proteins, hypothetical proteins and CCA1 derivatives. If we attempt a BLASTp against solved structures contained in the Protein Data Base (PDB), proteins containing the MYB-domain are identified, although these are mainly human examples. Interestingly, BLASTp of the PDB against *Arabidopsis thaliana* produces no results indicating the lack of solved structures from this organism.

A c-Myb domain bound to DNA was one of the top scores with an E-value of 9.7 x 10^{-17} e and an estimated precision of 100 %. The chain identity was only given 19 % but this is due to the relatively small number of residues involved in DNA binding in the LHY protein. The solution structure of the Myb_DNA-binding domain complexed with DNA (PDB identifier 1MSF) is shown below in Fig 1.3 (Ogata *et al.*, 1994). All features previously described can be seen in Fig 2.8. However, there is

more detail regarding the specificity of interaction. In this case, residues Asn183 (R3), Lys-182 (R3) and Lys128 (R2) are responsible for sequence specificity for the AACTG of DNA (Ogata *et al.*, 1994). As LHY and CCA1 are distinct from the R_2R_3 family, other residues are likely to be responsible for specificity. Indeed, as LHY and CCA1 only contain one MYB domain, it is likely that they act together (hetero or homo-dimers) in order to bind downstream targets. Crystal analysis with the addition of fragments of targets promoters would provide insight into which residues are important for DNA binding, as well as an understanding for the remaining C-terminus of the LHY and CCA1 proteins.

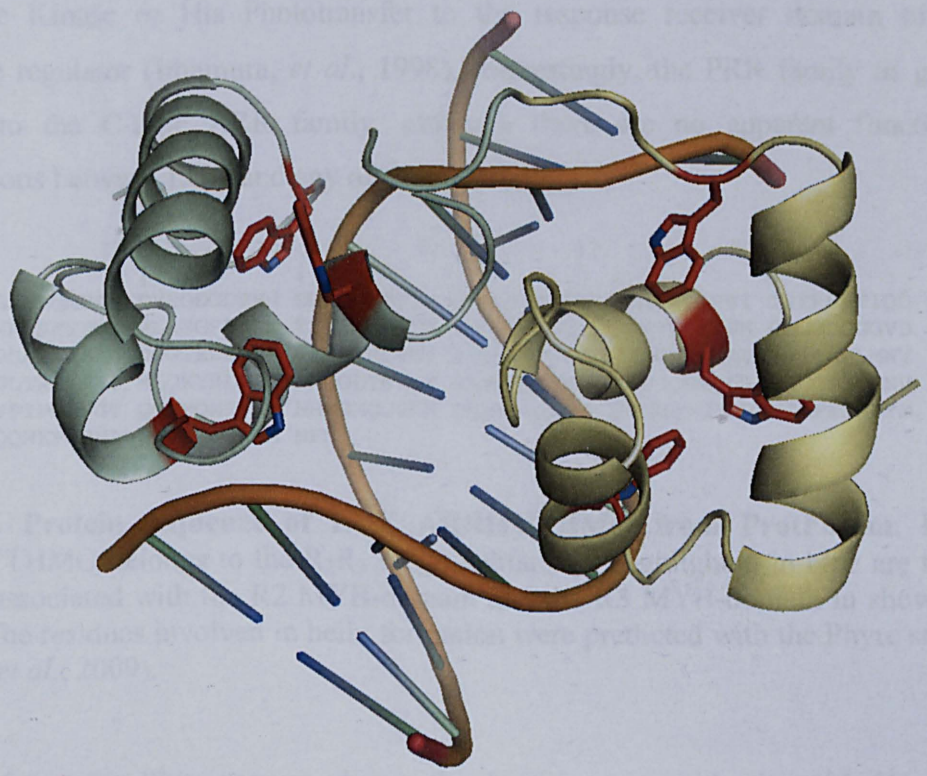


Fig 2.8. Ribbon diagram of c-MYB R2R3 complexed with DNA. The protein structure was deduced using NMR (Ogata *et al.*, 1994) and deposited in the PDB(1MSF). The R2 and R3 domains are shown in pale green and pale yellow respectively. The Trp residues important in maintenance of the hydrophobic core of each domain are shown in red. With respect to R3, the first helix (vertical) is shown on the right hand side. Helix 2 and 3 make the atypical H-T-H, with the third helix contacting the major groove of DNA (orange backbone). This binding is achieved through contacts in both the R2 and R2 third helices. This diagram was drawn with the Pymol software package.

2.3.2 LUX

The LUX protein differs significantly from LHY and CCA1 proteins. Although it contains the SHLQKYF (Leu192 replacing A) sequence and therefore can be classified as a *plant* specific Myb-domain protein, it is not part of the LHY / CCA1 gene family. Indeed, the conserved W residues seem to be absent in the LUX amino acid sequence. Furthermore, the LUX sequence contains two MYB-domains, classifying this protein as an R_2R_3 type MYB-domain transcription factor, rather than the $R_{1/2}$ type of LHY and CCA1 (Fig 2.9). The MYB domain in LUX shows greater similarity to MYB-domains found in the B-type response regulators (ARRs). The B-type ARR's are involved in signal transduction by transfer of phosphate by a Histidine Kinase or His Phototransfer to the response receiver domain of the response regulator (Imamura, *et al.*, 1998). Interestingly, the PRR family of genes belong to the C-type ARR family, although there are no apparent functional connections between LUX and any of the PRR family.

```
  1          11         21         31         41         51
     |          |          |          |          |          |
  1 MGEEVQMSDY DVSGDGDRVS EWEMGLPSDE DLASLSYSLI PPNLAMAFSI TPERSRTIQD
 61 VNRASETTLS SLRGGSSGPN TSSSNNNVEE EDRVGSSSPG SDSKKQKTSN GDGDDGGGVD
121 PDSAMAAEEG DSGTEDLSGK TLKRPRLVWT PQLHKRFVDV VAHLGIKNAV PKTIMQLMNV
181 EGLTRENVAS HLQKYRLYLK RMQGLTNEGP SASDKLFSST PVPPQSFQDI GGGGGSSGNV
241 GVPIPGAYGT QQMMQMPVYA HHMGMQGYHH QNHNHDPYHQ NHRHHHGAGG NGAFESNPYM
301 MQQNKFGSMA SYPSVGGGSA NEN
```

Fig 2.9. Protein sequence of LUX ARRHYTHMO from ProtParam. LUX ARRHYTHMO belongs to the R_2R_3 MYB subfamily. Highlighted in blue are the 3 helices associated with the R2 MYB-domain and the R3 MYB-domain in shown in violet. The residues involved in helix formation were predicted with the Phyre server (Kelley *et al.*, 2009).

Output from the Phyre server shows the best sequence identity (56 %) with DNA/RNA binding 3-helical bundle belonging to a GARP response regulator. In fact, the estimated precision is 100 % and with an E value of 4.7 x 10^{-08}e. Unfortunately, the protein (SCOP code *d1irza*) is not entered in the PDB and only has a SCOP code. Although this protein is classified as a GARP like protein, the SCOP entry also shows that the family it belongs to is Myb-DNA binding related. Indeed, the other close identities found on the Phyre server are all Myb domain proteins, including the Myb_DNA-binding domain (PDB identifier 1MSF) shown in Fig 2.8.

Although we can attribute the 3 helical bundles to LUX function in binding to the EE in promoters, it shows significant sequence difference from the LHY and CCA1 proteins to be classified as distinct. These facts coupled with the caveat that LUX may work with other proteins including TOC1 and GI (to promote the expression of *LHY* and *CCA1*), make LUX a potentially more interesting structural target than either LHY or CCA1.

2.4. F-box proteins

The ZTL family are a novel class of blue light photoreceptors. As previously described, they contain three characteristic domains LOV, F-box and six kelch repeats. Each domain will be considered in turn to help decipher how these proteins function.

2.4.1 LOV

Proteins involved in the perception of blue light contain a photosensory domain comprising of either one or two LOV domains. The main blue light receptors *in planta* Phototropin 1 and 2 (Phot 1 and Phot 2) have a photosensory domain that contains two LOV domains (LOV1 and LOV2) and a C-terminal Ser/Thr kinase domain belonging to the AGC family (cAMP-dependant protein kinase, cGMP-dependant protein kinase, and phospholipid-dependant protein kinase C). The LOV domains each non-covalently bind an FMN in the dark. Upon excitation by blue light, the FMN covalently binds to a conserved cystein residue in the LOV domain, resulting in a conformational change that increases the kinase activity of the photo-sensory domain (Tokutomi *et al.*, 2008). This binding is dark-reversible allowing quick switching between the two states.

The LOV1 and LOV2 domains in phototropins appear to have distinct properties. LOV1 is thought to be a dimerisation domain that is responsible for slowing the dark recovery of LOV2 (Kagawa *et al.*, 2004) and may attenuate the kinase activity of the phototropins. Indeed, *plants* expressing only the LOV2 domain show a reduced sensitivity to light, that is accounted for by the absence of dimerisation or faster dark recovery of LOV2 (Sullivan *et al.*, 2008). That said, it has been shown that LOV1 is not required for the photochemistry of phot1 and is only responsible for modest light

activation of phot2 (Cho *et al.*, 2007). The LOV2 domain appears to be more important than LOV1 in light activation of the phototropins. In the dark, LOV2 binds to the kinase output domain of phot2 and inhibits kinase activity (Cho *et al.*, 2007). This process is inhibited by blue light (Cho *et al.*, 2007). In the dark, an α-helix located between LOV2 and the output kinase domain (Jα-helix) is situated on the surface of the LOV2 domain. The presence of light causes the Jα-helix to unfold, alleviating the kinase domain from inhibition (reviewed in Demarsy and Fankhauser, 2009). Downstream targets for the kinase activity are poorly understood and the function of LOV1 remains unclear. Interestingly, the ZTL family only contain one LOV domain, LOV1. It differs from the phototropin LOV1 due to an absence of dark recovery, therefore suggesting a role in non-reversible light response (Imaizuni *et al.*, 2003). It has been shown that GI interacts with LOV of FKF1 and ZTL specifically on blue light perception (Sawa *et al.*, 2007; Kim *et al.*, 2007 respectively) and that LOV alone, is sufficient for binding of TOC1 to ZTL (Mas *et al.*, 2003).

2.4.2 F-box proteins

There are over 600 known F-box proteins in *Arabidopsis* with diverse functions, but they are often involved in the ubiquitin-proteasome degredation pathway. A general schematic of F-box function is shown in Fig 2.10a. In this example, E1 enzyme activates the Ubiquitin moiety (Ub) and E2 enzyme is responsible for the conjugation. The F-box protein (FBP) is part of an E3 ligase Skp_Cul1_F-box (SCF) complex, S phase kinase-assosiated protein 1 (Skp1), Cullin1 (Cul1), Ring-box 1 (Rbx1), which acts as a means of transferring Ub from the E2 enzyme to the FBP specific substrate. The FBP binds directly to Skp1, which interacts with Cul1, a scaffold protein which is associated with a RING domain protein, in this case Rbx1. The RING domain protein is responsible for the binding to E2, which is loaded with Ub moieties. In this way, the F-box proteins are able to recruit their specific substrates to the SCF complex (reviewed in Ho *et al.*, 2008). The substrates are often phosphorylated in order to provide target recognition to the FBP. The phosphorylation state of the substrate is important for F-box binding, although the degree and sites of phosphorylation are substrate specific.

There are deduced crystal structures of F-box proteins complexed with SCF (SCF^{FBP}) and these provide insight into the mechanism of Ub transfer (Zheng *et al.*, 2002; Orlicky *et al.*, 2003). Briefly, components of the SCF complex are organised into a C shaped arrangement that is approximately 59 Å from one end to the other. At one end is the N-terminal of the FBP responsible for substrate recognition and at the other is the E2 Ub carrying enzyme. For a transfer of Ub to the F-box, this rigid distance is required, so that the components of the SCF complex are in the correct orientation. A linker region between the F-box and the substrate recognition domain is also thought to be important in maintaining this distance, possibly by controlling the coupling between the two domains and therefore the enzyme activity of the surface. The F-box domain is a tri-helical structure which forms a hydrophobic surface for Skp1 binding (Cardozo and Pagano, 2004). Taken together, the precise arrangement of these protein complexes are important in formation of an active SCF^{FBP} complex.

ZTL was the first described FBP in the circadian system. A schematic representation of ZTL is shown in Fig 2.10(B). ZTL has been shown to associate with ASK1, AtCUL1 and AtRBX1 to form an active SCF^{ZTL} complex (Han *et al.*, 2004) which degrades TOC1 (Más *et al.*, 2003) and also PRR5 (Kiba *et al.*, 2007) in a light dependant manner. The F-box in FKF1 has been shown to be important in regulating *CO* expression, by degrading CYCLING DOF FACTOR 1 (CDF1) (Imaizumi *et al.*, 2005), a repressor of *CO* in a light dependant manner. A schematic of these roles are summarised in Figure 2.10(C and D).

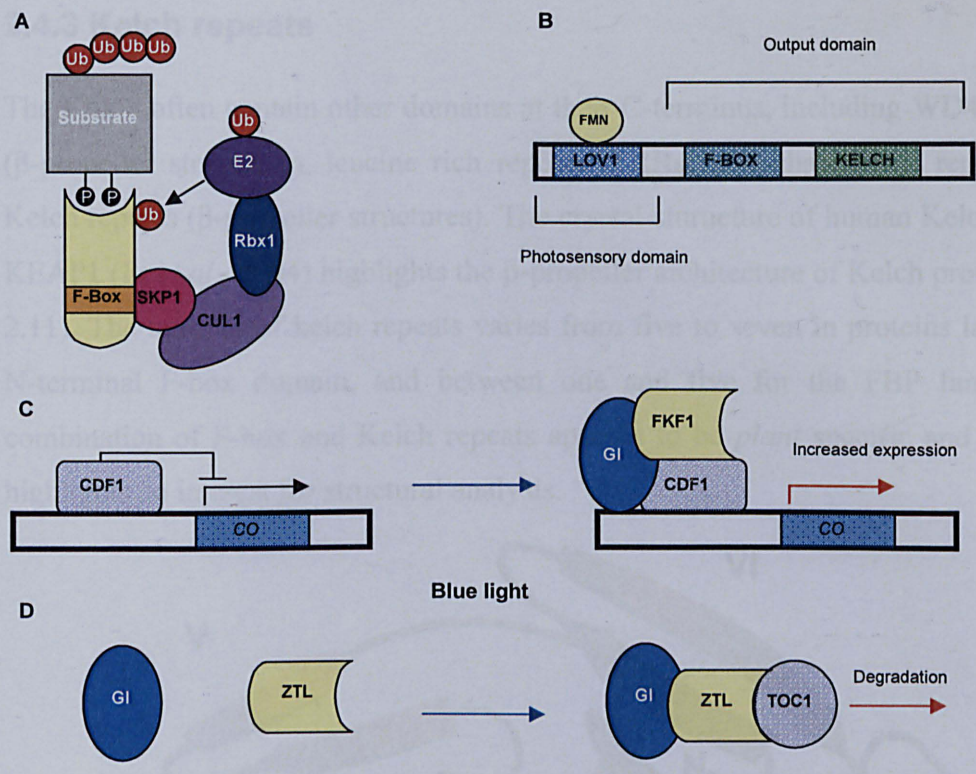


Fig 2.10. A schematic of F-box protein (A) involvement in targeted degradation; (B) ZTL architecture and (C, D) function in the circadian clock. (A) The FBP forms part of the SCF complex and mediates the transfer of Ubiquitin (Ub) from E2 (Ub conjugation enzyme) to the phosphorylated substrate. The ubiqutinated substrate is then degradated via the proteasome pathway. (B) Schematic representation of ZTL, highlighting the LOV1 domain that binds FMN, and the F-box / Kelch repeat output domains, responsible for SCF formation and protein:protein interaction respectively. (C&D) The F-box proteins FKF1 and ZTL interact with circadian protein GI in a blue light dependant manner. (C) FKF1 associates with GI through its LOV domain on exposure to blue light. The FKF1-GI complex binds to the CDF1 (*CO* repressor) targeting it for depredation. (D) ZTL interacts with GI through the LOV domain under blue light. The complex formed stabilises both ZTL and GI allowing more TOC1 to be targeted for degradation at the appropriate circadian time. This figure was largely based on the schematic shown in Fig 1 (A) and Fig 5 (A) from Ho *et al.*, 2008.

2.4.3 Kelch repeats

The FBP's often contain other domains at their C-terminus, including WD40 repeats (β-propeller structures), leucine rich repeats (LRRs; arch-shaped α-β repeats) and Kelch repeats (β-propeller structures). The crystal sturucture of human Kelch protein KEAP1 (Li *et al.*, 2004) highlights the β-propeller architecture of Kelch proteins (Fig 2.11). The number of kelch repeats varies from five to seven in proteins lacking an N-terminal F-box domain, and between one and five for the FBP family. The combination of F-box and Kelch repeats appears to be *plant* specific and therefore highlights an interest for structural analysis.

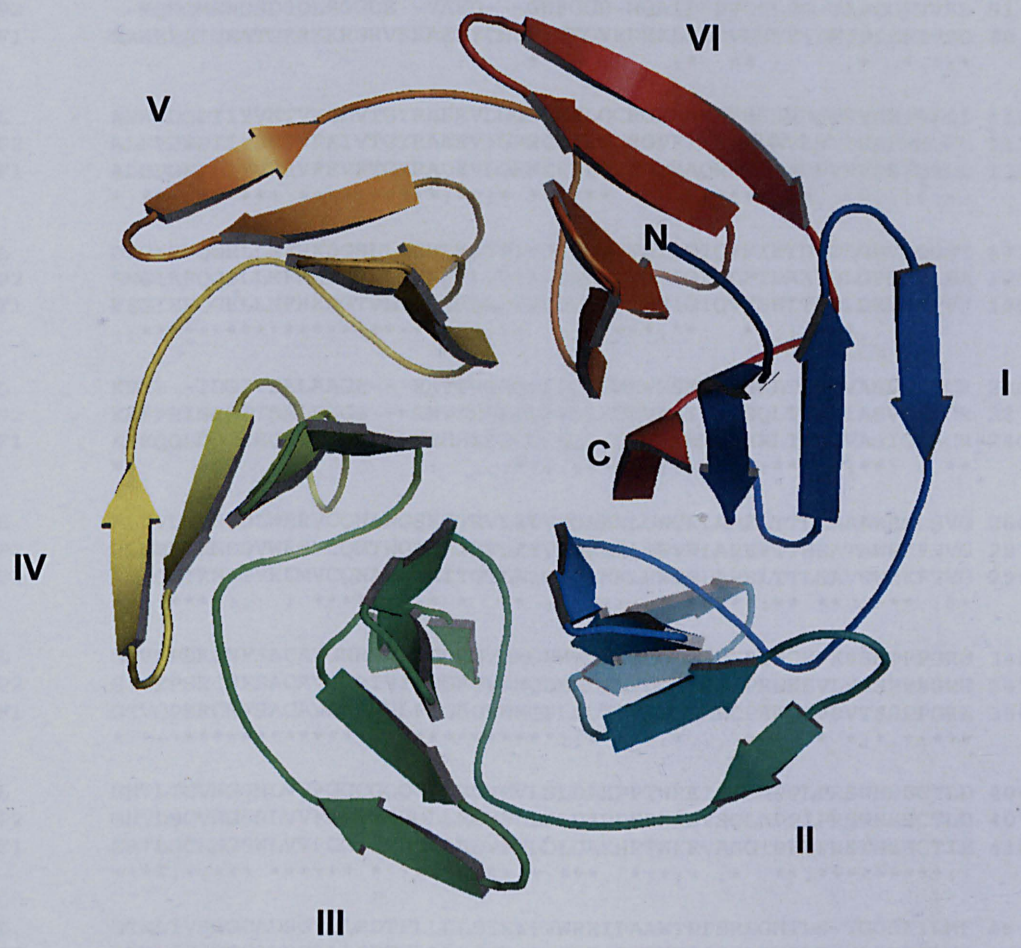


Fig 2.11. Ribbon diagram of the Kelch domain of human Keap1. The structure was solved to 1.85 Å (Li *et al.*, 2004) and deposited in the PDB (1U6D). The diagram shows the top-down view of the Kelch domain (residues 322-609). Each Kelch repeat is numbered I-VI, and contains four β strands which run anti-parallel to each other. Blade I contains both the N and C termini of the domain and this are labelled accordingly. This diagram was drawn with the Pymol software package.

An interesting point to consider with this set of proteins is that they all appear to have very similar sequences (shown in Fig 2.12), yet we already understand that they act on different targets. Deducing the crystal structures of each of these proteins will help elucidate how these proteins differentiate between their targets and may provide insight into unidentified targets / protein partners. The large degree of sequence similarity is shown in the ClustalW alignment below (Fig 2.12).

```
ZTL       ------MEWDSGSDLSADDASSLADDEEGGLFPG--GGPIPYPVGNLLH-TAPCGFVVTD 51
LKP2      --MQNQMEWDSDSDLSGGDE--VAED---GWFGGD-NGAIPFPVGSLPG-TAPCGFVVSD 51
FKF1      MAREHAIGEATGKRKKRGRVEEAEEYCNDGIEEQVEDEKLPLEVGMFYYPMTPPSFIVSD 60
                :   :..  . .      :    *      .  :*  ** :    :* .*:*:*

ZTL       AVEPDQPIIYVNTVFEMVTGYRAEEVLGGNCRFLQCRGPFAKRRHPLVDSMVVSEIRKCI 111
LKP2      ALEPDNPIIYVNTVFEIVTGYRAEEVIGRNCRFLQCRGPFTKRRHPMVDSTIVAKMRQCL 111
FKF1      ALEPDFPLIYVNRVFEVFTGYRADEVLGRNCRFLQYRDPRAQRRHPLVDPVVVSEIRRCL 120
          *:*** *:**** ***:.*****:**:* ****** *.* ::****:**. :*:::*:*:

ZTL       DEGIEFQGELLNFRKDGSPLMNRLRLTPIYGDDDTITHIIGIQFFIETDIDLGPVLGSST 171
LKP2      ENGIEFQGELLNFRKDGSPLMNKLRLVPIR-EEDEITHFIGVLLFTDAKIDLGPSPDLSA 170
FKF1      EEGIEFQGELLNFRKDGTPLVNRLRLAPIRDDDGTITHVIGIQVFSETTIDLDRVSYPVF 180
          ::****************:**:*:***.**  ::. ***.**: .* :: ***.

ZTL       KEKS--IDGIYSALAAGE---RNVSRGMCGLFQLSDEVVSMKILSRLTPRDVASVSSVCR 226
LKP2      KEIPRISRSFTSALPIGE---RNVSRGLCGIFELSDEVIAIKILSQLTPGDIASVGCVCR 227
FKF1      KHKQQLDQTSECLFPSGSPRFKEHHEDFCGILQLSDEVLAHNILSRLTPRDVASIGSACR 240
          *.          . :. *.   ::  ..:**:::*****:: :***:*** *:**:...**

ZTL       RLYVLTKNEDLWRRVCQNAWGSETTRVLETVPGAKRLGWGRLARELTTLEAAAWRKLSVG 286
LKP2      RLNELTKNDDVWRMVCQNTWGTEATRVLESVPGAKRIGWVRLAREFTTHEATAWRKFSVG 287
FKF1      RLRQLTKNESVRKMVCQNAWGKEITGTLEIMT--KKLRWGRLARELTTLEAVCWRKFTVG 298
          **  ****:.: : ****:**.* * .** :.  *:: * *****:** **..***::**

ZTL       GSVEPSRCNFSACAVGNRVVLFGGEGVNMQPMNDTFVLDLNSDYPEWQHVKVSSPPPGRW 346
LKP2      GTVEPSRCNFSACAVGNRIVIFGGEGVNMQPMNDTFVLDLGSSSPEWKSVLVSSPPPGRW 347
FKF1      GIVQPSRCNFSACAVGNRLVLFGGEGVNMQPLDDTFVLNLDAECPEWQRVRVTSSPPGRW 358
          * *:***************:*:**********::*****:*.:. ***: * *:*.*****

ZTL       GHTLTCVNGSNLVVFGGCGQQGLLNDVFVLNLDAKPPTWREISGLAPPLPRSWHSSCTLD 406
LKP2      GHTLSCVNGSRLVVFGGYGSHGLLNDVFLLDLDADPPSWREVSGLAPPIPRSWHSSCTLD 407
FKF1      GHTLSCLNGSWLVVFGGCGRQGLLNDVFVLDLDAKHPTWKEVAGGTPPLPRSWHSSCTIE 418
          ****:*:*** ****** * :*******:*:***. *:*:*::* :**:*********::

ZTL       GTKLIVSGGCADSGVLLSDTFLLDLSIEKPVWREIPAAWTPPSRLGHTLSVYGGRKILMF 466
LKP2      GTKLIVSGGCADSGALLSDTFLLDLSMDIPAWREIPVPWTPPSRLGHTLTVYGDRKILMF 467
FKF1      GSKLVVSGGCTDAGVLLSDTFLLDLTTDKPTWKEIPTSWAPPSRLGHSLSVFGRTKILMF 478
          *:**:*****:*:*.**********: : *.*:***..*:*******:*:*:*~ *****

ZTL       GGLAKSGPLKFRSSDVFTMDLSEEEPCWRCVTGSGMPGAGNPGGVAPPPRLDHVAVNLPG 526
LKP2      GGLAKNGTLRFRSNDVYTMDLSEDEPSWRPVIGYGSSLPG--GMAAPPPRLDHVAISLPG 525
FKF1      GGLANIGHLKLRSGEAYTIDLEDEEPRWRELECS--AFPG---VVVPPPRLDHVAVSMPC 533
          ****: * *::**.:.:*:**.::** ** :      . .*    ..*********:.:*

ZTL       GRILIFGGSVAGLHSASQLYLLDPTEDKPTWRILNIPGRPPRFAWGHGTCVVGGTRAIVL 586
LKP2      GRILIFGGSVAGLDSASQLYLLDPNEEKPAWRILNVQGGPPRFAWGHTTCVVGGTRLVVL 585
FKF1      GRVIIFGGSIAGLHSPSQLFLIDPAEEKPSWRILNVPGKPPKLAWGHNTCVVGGTRVLVL 593
          **::*****:***.*.***:*:** *:**:*****: * **::**** ******** :**
```

```
ZTL     GGQTGEEWMLSELHELSLASYLT--- 609
LKP2    GGQTGEEWMLNEAHELLLATSTTAST 611
FKF1    GGHTGEEWILNELHELCLASRQDSDL 619
        **:*****:*.* *** **:
```

Fig. 2.12. ClustalW alignement of the F-box proteins. An alignment of ZTL, FKF1 and LKP2 is presented. Conserved and semi-conserved residues are shown as stars and semi-colons respectively.

2.5. Proteins that are not classified according to sequence

The proteins discussed so far belong to discrete families. Whilst they may differ from other members of the family, they have enough structural (at least predicted) similarity to be grouped likewise. The remaining proteins introduced earlier are not members of previously reported families. They represent novel, *plant*-specific proteins. These are introduced over the next few pages.

2.5.1 ELF4

As previously described, ELF4 is a novel protein of unknown structure and function. BLASTp of ELF4 produces a list of hypothetical proteins as well as ELF4-like protein, presumably named due to their relatedness to ELF4. The only clue to a functional domain is the DUF1313 super-family on the pfam server (Fig 2.13).

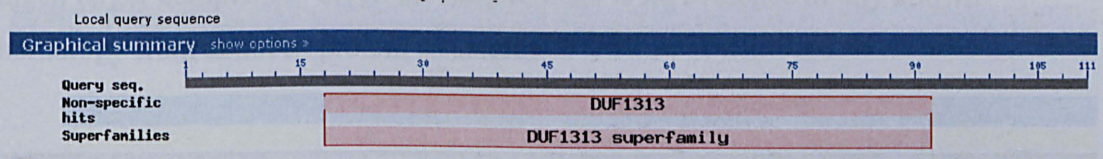


Fig 2.13. On-screen graphical summary of ELF4 protein sequence after BLASTp. The graphical domain of the DUF1313 is indicated below the ELF4 query (111aa) and spans from residue 16 to 92.

The DUF1313 super-family on pfam is described as a family of several hypothetical *plant* proteins approximately 100 amino acids in length. Pfam confirm that the function of this family is unknown.

Submission of the ELF4 sequence to the Phyre server confirms this prediction. In terms of output, the most similar protein in the PDB is chain A on the Syntaxin A protein. With the lowest E value above 0 and the highest estimated precision of prediction at 20 %, the conclusion holds that ELF4 structure resembles nothing so far published. That said, the secondary structure prediction (Psi-Pred, SSPro and JNet) highlights regions of disorder (likely unstructured) but also regions capable of forming α-helix (residues 23-56 and 62-86), in approximately a 50:50 ratio. This suggests ELF4 may be a good target for crystallisation as not only is it small, but it may form stable structures.

2.5.2 ELF3

As with ELF4, there is little in the literature regarding structure. Indeed, submission to the Phyre server confirms this as the top result in the fold recognition section is (PBD entry 1TWG) RNA polymerase II complexed with CTP. The RNA polymerase II structure consists of 10 peptides and the fold recognition between this entry and ELF3 is within chain A, a peptide consisting of 1733 residues. ELF3 only shows a 3 % identity with chain A and any similarity can be attributed to the large size of the peptide. Other predictions have similarly low percentage identities and relatively large E-value's (10^{-2} / 10^{-3}), showing similarity with small peptides that are part of much larger complexes. Given the output, there is no evidence of any known homology from deduced protein structures

2.5.3 GI

GI is a large nuclear protein of unknown structure and this was confirmed by the Phyre prediction programme. The top results show very small percentage identity and low E values with RNA polymerase. As GI has been shown to interact with other circadian proteins including the F-box proteins ZTL and FKF1, it is likely that there are domains along the length of the protein that have yet to be identified. Although the Phyre prediction did not highlight similarity of GI to structures contained in the PDB, the server did predict regions of secondary structure. By taking these regions and submitting them individually to the Phyre server we minimise the chance of poor

percentage identity against such smaller template sequences. Summarised below in Table 2, are the predictions based on truncations of the GI protein.

Again, there appears to be no confidant prediction of regions of similarity with the PBD database. As we are unable to predict domain boundaries for GI, it is difficult to determine which truncations should be submitted to the Phyre server. Indeed, it should be possible to enter much smaller regions in order to find a closer match. However, by submitting the regions of predicted secondary structure, we have gone some way to addressing this question. Table 2 shows the residues submitted to the Phyre server and the similarity to structures currently in the PBD database. It is clear from the output from such analysis that there appears to be very little confident matches with regard to the GI protein. Furthermore, the entries presented in Table 1 represent the most significant similarities from each submission. The conclusion is that it is not possible to predict functional domains of GI using these parameters and the Phyre server.

Residue	PDB entry	Functional summary	%Identity	E-value	Precision (%)
1-141	2VOX	The dimerisation domain of lap2alpha	19	11	20
261-397	2EBG	Hypothetical protein	26	5.8	35
435-601	2P72	Glycosyltransferase	17	8.4	25
693-807	1QVX	FAT domain of focal adhesion kinase	20	10	20
893-1170	2N05	Transcriptional regulator rha1	12	4.5	40

Table 2. Summary of Phyre predictions of GI based on regions of presumed secondary structure. The Table shows which residues from GI were submitted for each Phyre prediction. The best 'hit' is shown as a PDB entry with a description of each entry given under functional summary. Also highlighted are the % identity, E-value and precision (confidance) of the prediction.

2.5.4 SRR1

Analysis of the *SRR1* gene shows that it encodes a novel protein with homologs in several Eukaroytes including mouse, human *Drosophila* and yeast (Staiger *et al.*, 2002). The function of these proteins is yet to be determined. Indeed, more recent BLASTp analysis, identified a list of SRR1-Like protein. Again, these are

categorised based on sequence similarity to SRR1 and have yet to be assigned function.

Submission of SRR1 to the Phyre server confirms these findings. The top ten 'hits' have E-values greater than 0, % identities of less than 5 % with an estimated precision of less than 60 %. This further confirms SRR1 as a novel protein and indicates interest from a structural perspective.

2.6. Discussion

This chapter serves as an introduction to structural and biochemical analysis of clock-associated proteins. As previously mentioned, there is no defined way to present such analysis, so this chapter is a summary of what we know about particular families of proteins and how clock-associated proteins may function based on similarities to previously solved structures. There is little new information presented here. As the genetic evidence for the function of putative clock proteins is elucidated, similar bioinformatic approaches have been applied in order to assess the function of the newly discovered gene and its cognate protein. Many circadian publications only mention family identification and a broad description of the background into their modes of function. This chapter has focused on introducing the background of the protein families, including specific archetypal examples and tried to place context with respect to the proteins studied in this thesis. Differences between the clock-associated proteins and other family members have been discussed and this highlights the level of interest in structural studies of clock proteins. What is striking about the analysis conducted is that there are often fundamental differences between the queries and *bon fide* family members, such as the Asp to Glu alteration of the PRR family (phosphotransfer) or the lack of conserved Trp residues in the MYB-domain proteins (LUX). This occurs in the functional regions, so their relevance is interesting. Perhaps even more exciting are the remaining, undefined parts of the proteins. For ELF3, ELF4, SRR1 and GI, we are aware that they contain no known domains or motifs, despite containing regions of secondary structure (predicted and shown). However, even in reasonably well defined proteins such as the PRR family, there are huge sequence differences between database entries and also closely related

family members. Are these regions of undefined structure responsible for protein function, possibly protein:protein interactions?

The interest of these proteins from a structural standpoint goes beyond the circadian clock. Some of these proteins are likely to contain novel domains. Indeed, what is striking about the PDB entries are the lack of crystallised *plant* proteins. It is plausible that the *plant* circadian proteins represent completely unique structures. This is backed up by the caveat that circadian associated proteins between different organisms appear to share little homology.

Whilst much of what is reported is already understood, this chapter is the first example in the literature of a complete discussion of the proteins from a structural perspective, summarised in the same review. It certainly provides far more background than the current literature and therefore provides a link between the genetically based circadian information and the protein biochemistry of putative clock components.

Chapter 3. Materials & Methods

3.1. Initial expression screens for circadian proteins

3.1.1. Circadian gene cloning and vector construction

This part of the experimental procedure was carried out by our collaborator Dean Rea in the laboratory of Laszlo Polgár, Budapest. Briefly, the genes listed in Table 3 were amplified by polymerase chain reaction (PCR), digested with appropriate restriction enzymes and ligated into the expression vectors.

3.1.2. Small scale expression

The following vectors were used for small scale expression trials; pET26b-TOC1, pET32a-TOC1-PRR, pET28b-ELF3 and pET32a-ELF4. GI, SPY, ZTL, FKF1, LKP2, CCA1, LHY and SRR1, LIP1 and LIP1Δ234 were expressed in pMAL-c2x-cHis.

3.1.2.1. Transformation of *E. coli* strain B834 (λDE3) 'Rosetta'

1 µl of prepared expression construct was added to 100 µl of competent *E. coli* strain B834 (λDE3) cells (F^- *ompT hsdS*$_B$ (r_B^- m_B) *gal dcm* (DE3) pRARE (Cam^R), which contain the chloramphenicol resistance-carrying 'Rosetta' plasmid for supplementation of rare tRNA codons, and incubated on ice for 30 minutes (min). The mixture was then heat shocked at 42 °C for 45s and returned to ice for 2 min. The *E. coli* was plated onto pre-warmed (37 °C) Luria-Bertani (LB) agar containing 100 µg/ml ampicilin, 35 µg/ml chloramphenicol, and incubated at 37 °C overnight.

3.1.2.2. Expression of recombinant proteins

Vector containing colonies were used to inoculate 5 ml of LB medium containing 100 µg/ml ampicilin, 35 µg/ml chloramphenicol, and incubated at 37 °C with shaking at 180 rpm overnight. 50 µl of culture was added to 20 ml fresh LB containing 100 µg/ml ampicilin, 35 µg/ml chloramphenicol, and grown at 37 °C until an optical density (OD) of 0.6 at 600 nm was reached (OD_{600}). Protein expression was induced by addition of isopropyl-beta-thiogalactopyranoside (IPTG) to 0.5 mM, and growth was continued at 37 °C for 4 h.

3.1.2.3. Preparation of crude cell extract

Cells were harvested by centrifugation at 6000g for 15 min at 4 °C and re-suspended in 20 ml containing 50 mM HEPES pH 8.0, 1 M NaCl, 25 mM Imidazole, 10 % glycerol (buffer A). The cell suspension was sonicated (70 % power) on ice for 4 x 30 s and the soluble fraction was separated by centrifugation at 50,000g for 45 min at 4 °C. The insoluble pellet was re-suepended in 20 ml buffer A.

3.1.2.4. Polyacrylamide gel electrophoresis

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was carried out according to the method of Laemmli (1970) using 12 % gels cast and run using the Mini Protean II gel electrophoresis system (Biorad). 20 µl of expressed protein was added to 5 µl of SDS sample buffer (containing 10 % (w/v) SDS, 10 mM beta-mercapto-ethanol, 20 % (v/v) glycerol, 0.2 M Tris-HCl, pH 6.8 and 0.05 % w/v bromophenolblue) and heated to 95 °C for 4 min, before loading in to the gel. 5 µl of SeeBlue Plus2 pre-stained standard markers (Invitrogen) were also added to the gel for easy visualisation of protein molecular weights (Mw).
Samples were run on the gel for 2 h in 1 litre (l) of buffer containing 25 mM Tris-HCl, 200 mM glycine, 0.1 % (w/v) SDS (SDS running buffer) before staining with Coomassie blue (1 g Coomassie blue R250, 10 % acetic acid, 40 % methanol).

3.2. Sub-cloning circadian genes into pBAD-M41+

pP_{RO}-EX constructs containing the genes listed in Table 3 (page 77) were provided by Dean Rea. The majority of these were sub-cloned into a modified pBAD-M41+ vector (Invitrogen), except for GI, SRR1 and LIP1 which were sub-cloned into a modified pBAD-M41+ vector containing the multiple cloning site (MCS) of pP_{RO}-EX-HTa (Invitrogen), and LHY, CCA1 and LKP2 containing the MCS of pP_{RO}-EX-HTb (Invitrogen).

3.2.1. Production of competent TOP10 *E. coli*

100 µl of *E. coli* strain TOP10 (were cultured in 50 ml LB at 37 °C with shaking at 180 rpm until the OD_{600} reached 0.4-0.6. Cells were harvested by centrifugation at 4,000 rpm for 10 min at 4 °C and the pellet resuspended in 10 ml of ice cold 0.1 M $MgCl_2$ followed by centrifugation as before. The pellet was resuspended in 4 ml of ice cold 0.1 M $CaCl_2$, centrifuged as before and then repeated. Glycerol was added to a final concentration of 50 % (v/v) to the cells on ice and the cells were immediately frozen in liquid nitrogen in 200 µl aliquots and stored at -80 °C. All competent *E. coli* used in this thesis were prepared in this manner regardless of the strain.

The genotype of TOP10F' is provided below:

F'{*lac*Iq *Tn10* (TetR)} *mcr*A, Δ(*mrr-hsd*RMS-*mcr*BC), Φ80*lac*Z□M15, Δ*lac*X74, *end*A1, *rec*A1, *ara*D139, Δ(ara,leu)7697, *gal*U, *gal*K, *nup*G, *rps*L (StrR).

3.2.2. Digestion and ligation of constructs

1 µl of p pP_{RO}-EX constructs were added to 100 µl of TOP10 competent cells and these were heat shocked, plated and sub-cultured according to section 3.1.2.1 and 3.1.2.2 respectively. Plasmid DNA was extracted from the overnight culture using a miniprep kit (Qiagen) and 10 µl was double digested with 0.5 µl (Nco I and Xho I for pBADM-41+ constructs, Eco RI and Xho I for pBADM-Hta constructs and Bam HI and Xho I for pBADM-Htb constructs) each restriction enzyme (Fermentas) in a reaction mixture containing 8 µl H_2O and 4 µl restriction digest buffer for 2 h. Reaction mixtures were added to 5 µl of 30 % glycerol, 0.05 % bromophenol blue and 70 % TAE (DNA loading buffer) before running on 1 % agarose gel containing 10 mg/ml ethidium bromide for visualisation of DNA under UV light. The corresponding genes were cut from the gel and gel purified using a commercial kit (Qiagen). Vectors were digested and purified in the same way. 2 µl of vector was added to 10 µl of insert, 11 µl H_20, 2 µl ligase buffer and 1 µl T4 DNA ligase (Invitrogen) and then left overnight at 18 °C. 20 µl of ligation was added to 100 µl TOP10 cells and these were transformed and plated as before (section 3.1.2.1). To validate the success of the ligation, single colonies were sub-cultured overnight, DNA isolated, digested as before and the vector and insert were

visualised on a 0.75 % agarose gel. Samples positive for insert and vector were sequenced using the in-house sequencing service.

3.2.3. Expression in pBAD-M41+

Initial expression was performed as described in section 3.1.2.2. The only deviation was the addition of 0.2 % L-arabinose as a synthetic inducer in place of IPTG. Cells were harvested and visualised as highlighted in sections 3.1.2.3 and 3.1.2.4 respectively.

3.2.4. Western blotting

Polyacrylamide gels were blotted onto a nitrocellulose Hybond-C membrane (Amersham Biosciences) using a mini-trans blot cell (Biorad) at 20 volts (V) overnight. The nitrocellulose membrane was washed in 25 ml TBE for 10 min and then dried on blotting paper for 20 min. The membrane was immersed in 7 % milk (Marvell) in Tris-buffer saline (TBS) for 1 h to block non-specific proteins, before addition of 40 µl anti-His_6 anitbody conjugated to horse radish peroxidise (HRP) (Invitrogen) in 20 ml 7 % milk in TBS for 2 h. After 4 x 10 min washes in TBS-tween (TBS-T), excess liquid was removed from the blot and ECL detection solutions were added according to the manufacturer's instructions (Amersham). The blots were exposed on x-ray film (Kodak) for detection of His-tagged proteins.

3.3. Expression of TOC1 and ELF3 in Yeast

Constructs of pYES2 (Invitrogen) containing TOC1 and ELF3 were provided by our collaborator Dean Rea and these were used for expression trials in Yeast. Two strains DB2061 (Provided by Dr Carré) and INVSc1 (Invitrogen) were inoculated from glycerol stock onto a 2 cm^2 patch on non-selective YPAD agar (containing 12 g Bacto-peptone, 6 g yeast extract, 12 g glucose, 12 g agar supplemented with 100 mg adenine hemisulphate) before incubation overnight at 30 °C. The resulting yeast was used to transform pYES2-ELF3 according to the Quick and Easy TRAFO protocol (available at http://home.cc.umanitoba.ca/~gietz/Quick.html). Transformations by this method failed to yield any pYES2-TOC1 containing strains, so a more stringent method from the Yeast

molecular Biology protocol book was used instead (Gietz and Woods, 2002). Expression of the recombinant proteins was performed using the pYES2 version J protocol (Invitrogen), specifically the time course of protein induction by Galactose and detection of recombinant protein.

The genotype and phenotype of the INVSc1 host strain are provided below:

Genotype: MATa *his3Δ1 leu2 trp1-289 ura3-52*/MATα *his3Δ1 leu2 trp1-289 ura3-52*
Phenotype: His–, Leu–, Trp–, Ura–

3.4. Expression of TOC1-PRR

The N-terminal 312 amino acid (aa) (including the PRR domain) was cloned into pET32a and pP_{RO}EX-HTa by our collaborator Dean Rea. The vectors were transformed into *E. coli* strain C41 (λDE3) containing pRARE2 (rare tRNAs), expressed and harvested as indicated in section 3.1.2. The protocol deviated in that 2 l of culture were induced and left overnight at 20 °C instead of 37 °C for 3 h.

3.4.1. Nickel Affinity Chromatography

A 5 ml nickel sepharose Hi-Trap column (Amersham Biosciences) was equilibrated in buffer A using the ÄKTA fast protein liquid chromatography (FPLC) purifier. The clarified cell extract was applied and the column was washed with buffer A until the absorbance at 280 nm had returned to 0. Further washing in buffer B (50 mM HEPES pH 8.0, 1 M NaCl, 100 mM Imidazole, 10 % glyerol) to remove loosely bound proteins was followed by elution of TOC1-PRR with buffer C (50 mM HEPES pH 8.0, 1 M NaCl, 500 mM Imidazole, 15 % glyerol).

3.4.2. Protein Concentration and Determination of Protein Concentration

Eluted TOC1-PRR protein samples were pooled and then concentrated using a 10,000 molecular weight cut-off centrifugal concentrator (Vivaspin) as instructed by the manufacturer.

Protein concentration was determined from the absorbance at 280 nm using the calculated extinction coefficient, $A^{0.1\%} = 0.851$. This was calculated using the ProtParam tool available at http://us.exspasy.org/tools/protparam (Gasteiger *et al.*, 2005)

3.4.3. Buffer Exchange and Protease Digestion

The concentrated TOC1-PRR was divided into 2 centrifugal concentrators (Vivaspin) and 20 ml Thrombin digest buffer (20 mM Tris pH 7.4, 150 mM NaCl, 25 mM $CaCl_2$) was added to one sample and 20 ml Enterokinase digest buffer (20 mM Tris, pH 7.5, 150 mM NaCl, 2 mM $CaCl_2$) to the other. Samples were centrifuged at 4,000g until concentrated to 1 ml and then the procedure was repeated. Concentration of each sample was determined as previously described (section 3.4.2) Thrombin (Amersham) and Enterokinase (Novagen) were added as according to the manufacturer instructions and left at 4 °C overnight for digestion. Success of digestion was determined by a SDS-PAGE (section 3.1.2.4).

3.5. Coexpression of GI and SPY

GI has been co-expressed with SPY at the University of Minnesota (Tseng et al., 2004). pETcoco-SPY was provided upon request by Neil Olszewski and used in an expression trial with pMAL-GI as published (Tseng *et al.*, 2004). In addition, pETcoco-SPY was also co-expressed with pBADM41+-GI in *E. coli* strains C41, C43, M15 (containing the kanamycin resistant $pRED_4$) and JM101.

Cells were transformed with pETcoco-SPY (section 3.1.2.1), sub-cultured (section 3.1.2.2) with appropriate antibiotics and made competent (section 3.2.1) as previously described. 100 µl of competent cells were super transformed with either 2 µl pMAL-GI or pBAD-GI. Small scale expression was performed as before (section 3.1.2.2).

3.6. *In vitro* selection of aptamers

The random oligonucleotide library was synthesised and HPLC purified to concentration of 1 μmol (Operon) and this was diluted in 10 mM Tris pH 8.0 to produce a stock of 0.1 nmol / μl, which was stored at -20 °C. The diluted library (LICSelexApt) comprises of 40 random nucleotides flanked by the oligonucleotides; 5' –ggtattgagggtcgcatc -40N- gatggctctaactctcctct- 3', allowing easy amplification of the degenerate nucleotide pool. Primers LICSelexF and LICSelexR; 5'-ggtattgagggtcgcatc-3' and 5'-agaggagagttagagccatc-3' respectively, were used to amplify the aptamer pool by PCR during the selection process. The primers were obtained in both non-biotinylated and biotinylated forms (Invitrogen).

The His-MBP-SSR1 protein was purified (according to 3.1.2.2) and was selected due to its relative purity and the presence of an N-terminal His tag. Ni-NTA magnetic beads (Qiagen) were prepared by taking 150 μl of a 5 % slurry and equilibrated in 1 ml PBS-T (50 mM K_2HPO_4 pH 7.5, 150 mM NaCl, 0.05 % Tween -20). The Ni-NTA beads were allowed to settle, the supernatant removed, before being resuspended in 1.25 ml PBS-T. Purifed SRR1 fusion protein was diluted in PBS-T to give a final concentration of 2 mg / ml and 25 μl was added to the Ni-NTA beads in a 1 ml eppendorf, succeeded by 60 min incubation with mild rotation at 4 °C. The eppendorf was placed in a magnetic stand (Amersham Biosciences), to retain the beads and the supernatant removed. This was followed by 3 washes in 1 ml PBS-T and the resulting Ni-NTA-SRR1 bound beads were stored at -20 °C.

In the initial round of selection, 1 nanomole of LICSelexApt was diluted in100 μl of PBS-T in a PCR tube, heated to 95 °C for 2 min and instantly cooled on ice. The resulting ssDNA was added to 10 ml PBS-T containing 1 μg / ml bovine serum albumin (BSA) and 0.1 μg / ml dI-dC (Polydeoxyinosinic-Deoxycytidylic Acid). One hundred pmol of bead-bound SRR1 was then added to this mixture and incubated with rotation for 30 min at 20 °C. These tubes were applied to a magnet, the supernatant removed and 1 ml of PBS-T was added. The wash step was achieved by inversion of the eppendorf, reapplying the magnet and removing the supernatant. Each wash was repeated 2 further times, to remove any unbound ssDNA. SRR1 fusion and aptamers were eluted from the Ni-NTA beads with 20 μl of 20 mM Tris

pH 7.5, 500 mM imidazole and stored at 4 °C. A PCR tube containing 10 µl of the SRR1 / aptamer eluate was added to 10 µM LICSelex F and LICSelexR (biotinylated), 0.1 mM dNTPs, 0.2 mM $MgSO_4$, 1.25 U pfx polymerase (Invitrogen), 10 µl pfx buffer, and made up to 100 µl total volume in milli Q water. Amplification conditions were 95 °C for 2 min; 20 cycles of 95 °C for 30 s; 50 °C for 30 s; 68 °C for 30 s; 2 min at 68 °C. The resulting 80 bp product was visualised by loading 10 µl onto a 3 % agarose gel stained with ethidium bromide. The remaining 90 µl of PCR product and 23 µl of 5 M NaCl were mixed with 1.5 mg of M-280 strepavidin magnetic beads (Qiagen) for 15 min at 20 °C, and then washed 3 times in 1 ml PBS-T, as previously described. The ssDNA was separated from the complementary immobilised strand (biotinylated) by the addition of 50 µl of 100 mM NaOH. The tube was applied to a magnet; the supernatant containing the ssDNA was carefully removed and diluted in 1 ml of PBS-T containing 100 mM monobasic phosphate buffer to adjust the pH to 7.5. To prevent re-annealing, the material was heated to 95 °C for 2 min and then placed on ice immediately thereafter.

Subsequent rounds of selection were performed as described overleaf with only a few exceptions. The numbers of cycles of the elongation stage during PCR were altered during the rounds of selection. For each round, 10 cycles of elongation were performed and 10 µl of the PCR product was visualised on a 3 % agarose gel. In the event of no product present, the aptamer pool was subjected to a further 5 cycles of elongation until a product was detected. Rounds 1-3 required 20 cycles, (4-6) 15 cycles, (7-8) 10 cycles and (9-15) 15 cycles. The amount of protein used was decreased from 50 µg to 25 µg for rounds 10-15, and the incubation time was reduced to 15 min. A negative selection of non-specific aptamers was performed after rounds 4, 8 and 12, where the aptamer pool was incubated with Ni-NTA magnetic beads alone. 5 % slurry of Ni-NTA magnetic beads was added to 1 ml of ssDNA in PBS-T and incubated for 15 min with rotation. The eppendorf was applied to a magnet and the supernatant removed for use in subsequent rounds of SELEX.

After 15 rounds of SELEX two strategies were employed. The first was to test if the SELEX procedure had developed a pool of aptamers with specificity towards the target. The second involved cloning the aptamers into PSK II '-'vector, transforming

into competent *E. coli* and isolation for sequencing in order to determine the sequences responsible for the specificity.

3.6.1. Testing the polyclonal aptamers for protein specificity

Using the eluate after last round of SELEX it was possible to determine if the polyclonal aptamer pool had specificity towards SRR1. To determine specificity, 10 µl of the last round of elution was subjected to a PCR using the same conditions and reagents as described for the first round of SELEX selection, with the non-biotinylated LICSelexF primer being replaced with the biotinylated form. The double stranded PCR product was then heated to 95 °C for 2 min and instantly cooled on ice, yielding biotinylated ssDNA that could be used in a Dot-Blot to assess specificity.

3.6.2. Dot-Blot of polyclonal aptamers against purified SRR1

A nitrocellulose membrane (Amersham Biosciences) was divided into 6 equal squares each measuring 2 cm^3. To each piece, 2 µl of purified SRR1 (2 mg / ml) was added and dried using a low temperature setting on a hand held hair-drier. The polyclonal aptamer pool was diluted in H_2O, in 10-fold serial dilutions increments and added to each square in 5 µl aliquots. The membrane was blocked with 5 % BSA in TBS-T (20 mM Tris-HCl, 150 mM NaCl pH 7.5, 0.05 % Tween20) for 1 hr at room temp followed by 3 wash steps in TBS-T for 10 min each. The streptavidin-HRP conjugated antibody (Amersham Biosciences) was diluted 1:1,500 in TBS and 10 µl was applied to the membranes followed by incubation with shaking for 2 hr at room temp. The blots were then washed 3 times in TBS-T, for 15 min, and detection was performed using the ECL detection reagents (Amersham Biosciences) according to the manufacturer's instructions.

3.6.3. Cloning of the aptamer pool

2 µl of PSK II were mixed with 100 µl of competent *E. coli* (DH5α) cells and left on ice for 45 min. The mixture was heat shocked at 42 °C for 45 s and incubated back on ice for 2 min. All 100 µl of the transformation was plated on LB agar containing 50 µg / ml kanamycin, and incubated at 37 °C overnight. Individual colonies were

picked, used to inoculate 5 ml LB containing 50 μg / ml kanamycin and grown overnight at 37 °C with shaking of 180 rpm. The plasmid was extracted from the overnight culture using the standard miniprep (Quiagen) protocol. The purified vector was then blunt-end digested with Sma I to allow the incorporation of a T/A based cloning strategy.

The entire 50 μl of vector was used in a restriction digest containing, 2 μl Sma I (New England Bioloabs), 10 μl of 10 X NEB buffer 4 and 38 μl ultra pure water. The reaction was incubated at 25 °C for 2 hrs followed by heat inactivation of the enzyme at 65 °C for 20 min. After the blunt-end digest, 50 μl of the vector was incubated with 20 μM deoxyThymidine Triphosphate (dTTP), 1 U Taq polymerase and 1 X Taq buffer to add single 3' T overhangs at both termini of the blunt-end vector. This reaction proceeded for 1 hr at 70 °C. The resulting plasmid was purified by phenol extraction followed by ethanol precipitation.

3.6.4. Phenol extraction and ethanol precipitation of PSK II

An equal volume of 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA (TE) saturated phenol was added to the digestion mixture in a 1.5 ml microfuge tube and vortexed for 30 s before centrifugation at 13,000 rpm for 2 min in a bench top microcentrifuge. The upper aqueous layer was transferred to a clean tube, carefully avoiding the aqueous phenol interface. The aqueous phase was extracted once more by the addition of an equal volume of TE saturated phenol:chloroform (1:1), centrifuged as before and removed to a clean tube. An equal volume of water saturated ether was added, followed by vortex and centrifugation at 13,000 rpm for 3 min. The upper ether layer was discarded and the procedure was repeated once more.

To precipitate the DNA, 3 volumes of 95 % ethanol / 0.12 M sodium acetate was added to the tube and left on ice for 20 min. The DNA was precipitated by centrifugation at 13,000 rpm for 15 min at 4°C. The supernatant was once more discarded and the precipitate left to dry at room temp. Two volumes of 80 % ethanol were added to the precipitate followed by centrifugation as before. The supernatant was discarded and the microfuge tube placed in a speed-vac (Savant) to dry for

approximately 10 min. The DNA was dissolved in 50 µl 10:0.1 TE buffer and stored at 4 °C.

3.6.5. Preparation of the aptamer pool inserts

The aptamer pool was amplified by PCR under the same conditions as detailed in the first round of SELEX (note 25 cycles), using the non-biotinylated primers LICSelexF and LICSelexR. The aptamer template was diluted 1:100 in ultra pure water prior to PCR and 10 µl was used as the template. In order to make the aptamer pool complementary to the digested PSKII vector, 50 µl PCR product was incubated with 20 µM deoxyAdenosine Triphosphate (dATP), 1 U Taq polymerase and 1 X Taq buffer and heated to 72 °C for 8 min. This produced aptamers with 3' A overhangs as required in T/A cloning. The DNA was isolated by phenol:chloroform extraction followed by ethanol precipitation using the same protocol described. The DNA was resuspended in 50 µl of ultra pure water.

3.6.6. Cloning individual aptamers into PSKII

The ligation reaction contained 20 µl insert, 2.5 µl PSKII vector, 3 µl ligase buffer (Invitrogen), 1 µl T4 DNA ligase (Invitrogen) and 3.5 µl H_2O. To ensure efficient ligation, the reaction was left at 18 °C overnight. 10 µl of the ligation was transformed into DH5α competent *E. coli* (incubated and heat-shocked as before) and plated onto LB agar containing 50 µg / ml kanamysin (Kan). Plates were left at 37 °C overnight and 25 colonies were used to inoculate 25 x 5 ml LB cultures containing the antibiotic. The DNA was extracted by the standard miniprep procedure (QIagen) and these were sent for sequencing on the in-house AB sequencing facility according to their guidelines.

3.7. ELF4

3.7.1. pBAD-M41+-ELF4

pBADM-41+-ELF4 was expressed in *E. coli* C43 (λDE3) containing pRARE as detailed in section 3.4. The initial Ni-affinity chromatography was performed as described in section 3.4.1, and size exclusion was performed as described in section 3.6.2. Unsuccessful purification procedures were carried out according to the manufacturers instructions; ResourceQ and S (Amersham Biosciences) and amalose resin (New England Biolabs). Ammonium sulphate cuts from 10-90 % saturation were performed according to Deutscher, 1990.

3.7.2. pET32a-ELF4

pET32a-ELF4 was expressed in *E. coli* C41 (λDE3) containing pRIL (rare tRNA codon) as before (section 3.4). The first step in the purification was Nickel affinity chromatography which was performed in a similar manner (section 3.4.1) and concentrated using Vivaspin concentrators.

3.7.2.1. Size Exclusion Chromatography and Storage

A Superdex 200 gel filtration column (GE Healthcare life sciences) consisting of cross-linked agarose and dextran matrix was used in association with the ÄKTA purifier as a simultaneous polishing and buffer exchange step. The 50 ml Superdex column was equilibrated in 350 ml Enterokinase buffer containing 10 % glycerol. 5 ml of he Nickel column elution was applied to the Superdex 200 and ELF4 was eluted based on size. The resulting fraction was flash frozen in liquid nitrogen and stored in 1 ml aliquots (5 mg / ml) at -80 °C.

3.7.2.2. Digestion, Buffer Exchange and Anion Exchange Chromatography

100 mg of ELF4 fusion was digested with Enterokinase according to the manufacturer's instruction (Novagen). The digested ELF4 was buffer exchanged with 20 mM Tris-HCl pH 8.0 using the centrifugal concentrator (Vivaspin), concentrated

to 2 ml and applied to an 1 ml Resource Q (GE Healthcare life sciences) connected to an ÄKTA purifier. The sample was eluted from the column by application of a gradient from 0 – 1 M NaCl in Tris-HCl pH 8.0. Separation of the fusion from the ELF4 protein was determined using SDS-PAGE (section 3.1.2.4).

3.7.2.3. Desalting and Storage

5 ml cleaved and purified ELF4 samples were buffer-exchanged into 20 mM Tris-HCl pH 7.0 using 4 serially-linked 5 ml Hi-trap desalting columns (Amersham Biosciences) according to the manufacturer's procedure. Protein concentration was determined using the A_{280} as before (section 3.4.2) and samples were concentrated to 12 mg / ml. Freeze-thawing samples leads to aggregation so samples are either used immediately or were stored for up to 24 h at 4 °C.

3.7.2.4. Mass Spectrometry

100 μM of purified ELF4 was analysed by the Biological Mass Spectrometry and Proteomics service (Department of Biological Sciences, University of Warwick). Briefly, the bands of interest were excised then processed on the MassPrep protein handling system using the manufacturer's recommended methodology (available on request). The tryptic peptide extracts were transferred to suitable vials and analysed by means of nanoLC-ESI-MS/MS (NanoAcquity/Synapt HDMS) and the data used to interrogate appropriate databases for the sample source using ProteinLynx Global Server rel. 2.3.

3.7.2.5. Crystallisation screens

Crystal trials were performed in 96-well micro-titre plates. 50 μl of screening buffers were added to the plate reservoir and 1 μl was transferred to the adjacent well. 1 μl of 12 mg / ml ELF4 were added to each well to create a total of 2 μl under each condition. The screens used were as follows; MDL1 and MDL2 (Molecular Dimensions Ltd, (Jancarik and Kim, 1991), Clear Stratergy I and II (Molecular Dimesnsions Ltd), Hampton Index (Hampton Research; (Cudney *et al.*, 1994)), Wizard Sparse 1 and 2 (Emerald Biosystems Inc) and Newcastle screen (Prof. Richard Lewis, University of Newcastle, personal communication).

3.7.2.6. Circular Dichroism (CD)

CD scans were performed on a Jasco J-815 CD spectrophotometer. A data pitch of 0.5 nm, band width of 1 nm, and response time of 1 s were used for all scans. All spectra presented here are the average of 16 consecutive scans. Far-UV CD scans in the range 260-190 nm were performed using a 1 mm path length cuvette, with 0.25 mg / ml protein in 10 mM phosphate buffer pH 7.0. Near-UV CD scans in the range 320-240 nm were performed using 1 mg / ml protein in 10 mM phosphate buffer pH 7.0, using a cuvette of 10 mm path length. For Far-UV measurements, mean residue ellipticity ($[\theta]_{MRW}$) was calculated by this equation:

$$[\theta]_{MRW} = \frac{\theta}{Ncl \times 10}$$

where θ is the observed ellipticity (millidegrees), N is the number of peptide bonds in the protein, c is protein concentration (M) and l the pathlength of the cell used (cm).

3.8. LUX

3.8.1. *Plant* Growth

Ws Arabidopsis seeds were sterilised by decanting 40-60 into an eppendorf containing 1 ml 50 % bleech, 0.01 % Tween 20 with gentle agitation for 5 min. 0.5 ml sterile H_2O with 0.01 % Tween 20 was added, followed by several inversions allowing seeds to settle. The liquid was decanted and seeds were washed in 4 x 1 ml sterile H_2O. Seeds were sown on MSO agar plates 10-15 mm apart and left at 4 °C for 4 days before being transferred to continuous light in a grow-room at 21 °C.

3.8.2. Preparation of template

1 g of *plant* material was frozen in liquid nitrogen and ground into a fine powder in a pestle and mortar on dry ice. 15 ml of extraction buffer (100 mM Tris pH 8.0, 50 mM EDTA, 0.5 M NaCl, 10 mM beta-mercaptoethanol) is added to 2 ml SDS and mixed with the *plant* material and shaken thoroughly. The mixture was incubated at 65 °C for 10 min before addition of 5 ml 5 M potassium acetate followed by incubation on ice for 20 min. Then the tube was centrifuged at 25,000g for 20 min, the supernatant filtered through a micracloth into 10 ml isopropanol, and then frozen at -22 °C for 30 min. The tube was re-centrifuged at 20,000g for 15 min after which the supernatant was discarded and the pellet redissolved in 0.7 ml 50 mM Tris pH 8.0, 10 mM EDTA. The liquid was transferred to an eppendorf and spun in a bench top centrifuge for 10 min. 20 µl of RNase (2 mg / ml) was added and incubated at 37 °C for 30 min. DNA was precipitated from solution by adding 75 µl 3 M sodium acetate and 150 µl ethanol, mixing well and centrifuging at 13,000 rpm for 1 min in a bench top centrifuge. The supernatant was discarded and the DNA pellet washed in 80 % ethanol. The pellet was air-dried and redissolved in 10 mM Tris pH 8.0, 1 mM EDTA.

3.8.3. Polymerase Chain Reaction

Two primers were designed based on the Lux arrhythmo sequence found on Pubmed. 0.5 µl (stock 100 µM) of LUX_F25 forward (ATGGGAGAGAGGAAGTACAAAT-GAG) and LUX_R27 reverse (TTAATTCTCATTTGCGCTTCCACCTCC) primers were added to a reaction mixture containing 2 µl DNA template (section 2.7.2), 0.4 µl nucleoside triphosphatase (NTPs), 0.5 µl cloned Pfu turbo DNA polymerase (Invitrogen), 2 µl cloned Pfu buffer and this was made up to a total reaction volume of 20 µl with sterile H_2O. The PCR conditions were according to the manufacturer's instructions with the exception of annealing temperature which was increased to 55 °C and the number of elongation cycles changed to 25. The product was visualised on a 0.75 % agarose gel. The 1 kilobase (kb) product was excised from the gel and purified using a gel purification kit (Qiagen). The purified LUX DNA was used as a template in a similar PCR but instead using primers containing overhangs with restriction sites included. 0.5 µl LUX_For (TTTGCGGAATTCATGGGAGAGGAAGTACAA) and LUX_Rev (TTTGCGCTCGAGTTAGTCGACATTCTCATTTGCGCTTCC) with

EcoRI, XhoI and SalI restriction sites underlined respectively) were added to a reaction mixture containing 2 µl template (diluted 1:100), 0.4 µl dNTPs, 2 µl cloned Pfu buffer, 0.5 µl Pfu and 14.1 µl H_2O. PCR was performed as before and the DNA was isolated and gel purified from a 0.75 % gel as described above.

3.8.4. Cloning of Insert

30 µl of the LUX insert was digested with 1 µl EcoRI and 1 µl XhoI, and 50 µl of the pBADM41+-BN vector was digested with 1.5 µl of both EcoRI and XhoI according to the Fermentas protocol. The digests were run on a 0.75 % agarose gel and purified using the Qiagen gel purification kit. A reaction containing 1 µl vector, 20 µl LUX insert, 1 µl T4 DNA ligase, 3 µl T4 ligase buffer, 5 µl H_2O ligated the insert into the vector and this was verified using the in-house sequencing service.

3.8.5. LUX expression and purification

1 L pBADM-41+-LUX was expressed as described in section 3.2.3. The cells were harvested as described in section 3.1.2.3 and purified by Ni-affinity chromatography (Section 3.4.1). The MBP-LUX fusion was buffer exchanged and concentrated to 5 ml of 20 mM Tris-HCl pH 8.0 in a centrifugal concentrator (vivaspin). The MBP-LUX fusion was cleaned by anion exchange on a resource Q column (3.6.2.1) before being buffer exchanged into TEV protease buffer (containing 50 mM Tris pH 7.5, 0.5 mM EDTA, 2 mM β-mercaptoethanol) by size exclusion chromatogaphy as before (3.6.2.2). 100 µl TEV protease was added and left overnight at 4 °C. The digested product was subjected to size exclusion chromatography as described in section 3.6.2.1 and then further purified by anion exchange chromatography on MonoQ5/50 column (GE Healthcare life sciences) according to the manufacturers' instructions. The small quantities of LUX were detected by silver staining of SDS-PAGE. A combination of SDS-PAGE and Western blots were used to analyse purification procedures.

3.8.6. Silver staining of SDS-PAGE

The SDS-PAGE was fixed in 50 ml de-stain (50 % v/v methanol, 10 % v/v acetic acid and 40 % H_2O) for 30 min, washed in 50 ml 50 % v/v methanol, 5 % v/v acetic acid and 45 % H_2O for 10 min and then rinsed in distilled H_2O. The gel was then placed into 150

ml 10 % glyceraldehyde for 30 min before being transferred to H_2O and left overnight at 4 °C. The H_2O was decanted and replaced with 200 ml containing 1 g w/v DTT for 30 min, followed by 150 ml 0.1 % v/v silver nitrate for a further 30 min. The gel was rinsed in H_2O for 5 min and 150 ml 3 % sodium carbonate (Na_2CO_3), containing 75 µl formaldehyde was added until bands became visible. The reaction was stopped with 7.5 ml 2.3 M citric acid and the resulting gel was washed in 0.3 % Na_2CO_3 and then H_2O for 10 min each.

3.9. LIP1

The constructs used in the expression trials of LIP1 were pBADM-41+-LIP1, pMAL-LIP1 and pMAL-LIP1Δ.

3.9.1. Small scale expression and Ni purification of LIP1 and LIP1Δ

The vectors were transformed into *E. coli* strain C41 (λDE3) containing pRARE2 (rare tRNAs), expressed and harvested as indicated in section 3.1.2. The protocol deviated in that 0.5 l of culture were induced and left overnight at 20 °C instead of 37 °C for 3 h.

3.9.2. Large scale pMAL- LIP1Δ Expression

7 L of LB were inoculated with 5 ml pMAL-LIP1Δ overnight culture and grown at 37 °C until an OD_{600} of 0.4 was reached. Protein expression was induced at 25 °C with 1 mM IPTG and the culture was left overnight. Cells were harvested and purified by Nickel affinity chromatography as described in sections 3.1.2.3 and 3.4.1 respectively.

3.9.3. pMAL- LIP1Δ Digestion

The Ni purified fusion protein was buffer exchanged into Factor Xa protease buffer (50 mM Tris pH 7.5, 0.1 M NaCl, 1 mM $CaCl_2$) as indicated in section 3.6.2.2. The MBP-LIP1Δ fusion was concentrated to 1 mg / ml, 4 µl Factor Xa (Pierce) was added to 1 ml of concentrated fusion and left to digest overnight at 18 °C.

3.9.4. Assaying for GTPase activity

The GTPase activity of LIP1Δ was determined using a malachite green assay. The assay detects the release of inorganic phosphate using visible light spectroscopy, via the formation of an intermediate complex between malachite green molybdate and free orthophosphate at 630 nm.

To prepare the phosphate standard curve, 12 µl of 1 mM phosphate solution was added to 288 µl reaction buffer (20 mM Tris-HCl, pH 8.0, 200 µM GTP, 5 mM $MgCl_2$) to make a 40 µM phosphate solution. 500 µl of the phosphate solution was added to 1.5 ml reaction buffer to make a 10 µM phosphate standard. From this, dilutions were made to obtain a standard curve from 0 – 10 µM phosphate and the standards were made up to 100 µl total volume using the reaction buffer. 80 µl of the standards were pipetted in duplicate in clear-bottom 96-well plates (VWR International). Blanks containing reaction buffer and water only were added a negative controls.

LIP1 and LIP1Δ were diluted (10 µM and 30 µM final concentration) and 80 µl added in duplicate to the 96-well plate. The reaction was initiated by the addition of 20 µl Malachite Green Reagent (Invitrogen), the wells were mixed on a plate mixer and the plate was incubated at room temperature for 10 min. The absorbance of the solution at 630 nm was recorder over a period of 60 min on a DYN-EX MRX plate reader.

Chapter 4. Expression trials of circadian proteins.

4.1. Introduction

In order to investigate the structure of a protein by crystallisation, an expression and purification protocol that results in the production of milligrams of soluble protein is required. At the start of this thesis, no such protocol was available. Although there were several examples of over-expressed circadian proteins in the literature, the quantities obtained were in the μg range. A consortium based at the University of Wisconsin, Madison, known as the Center for Eukaryotic Structural Genomics (CESG) was founded as a collaborative effort to develop efficient methods of obtaining three dimensional crystal structures, with the initial focus on proteins found in the model *plant Arabidopsis thaliana*. The CESG designed an expression and purification protocol that utilises a pBAD vector containing an N-terminally His-tagged Maltose Binding Protein (MBP), which is responsible for increasing protein solubility and affords an effective one-step method of purification. In addition to this, a Tobacco Etch Virus (TEV) protease cut site is located between the MBP-fusion allowing easy separation of the native protein. The consortium reported much higher yields of soluble *Arabidopsis* proteins using this expression system (Jeon *et al.*, 2004).

In this chapter the expression of circadian proteins listed in Table 3 was investigated. Small scale expression trials in the constructs provided by Dean Rea were performed as described in section 3.1.2. Furthermore, pBAD constructs were made (section 3.2) and expression trials based on the CESG protocol were performed (section 3.2.3). The success the trials determined which proteins would be subject to further study. Particular emphasis was placed on obtaining the positive regulators of *LHY* and *CCA1*, including ELF3, ELF4, GI and TOC1.

Protein	pProEx	pET26b	pET28a	pET32a	pET39b	pGEX6p	pMAL-c2x	pMAL-c2x-cHIS	pYES2
TOC1	x	x	x	x	x	x	x		x
TOC1-PRR	x			x					
ELF3	x	x	x	x	x	x	x		x
ELF4	x	x	x	x	x	x (+4T)			
GI	x					x	x		
SPY	x								
ZTL	x	x	x	x	x	x			
FKF1	x					x			
LKP2	x					x		x	
CCA1	x					x			
LHY	x					x			
SRR1	x								
LIP1	x			x	x	x	x	x	
LIP1 Δ234	x			x		x	x	x	

Table 3. Constructs provided for expression of circadian proteins. This table highlights the constructs provided at the start of this body of work. Constructs were made in Hungary by our collaborator Dean Rea. The genes are listed on the far left. Of particular note are the TOC1-PRR and LIP1Δ234. TOC1-PRR contains residues 1-145 of the TOC1 sequence (the PRR domain only) and LIP1Δ234 is a truncated LIP1 (residues 1-234) that lacks the disordered 108 C-terminus. In this table x denotes provided construct.

4.2. Initial expression trials

Small scale expression trials of the constructs shown in Table 3 were initially performed in the laboratory of Laszlo Polgár, Budapest and re-performed at the start of this project. Briefly, constructs highlighted in Table 3, were transformed into B834 competent *E.coli* containing the pRARE plasmid (encodes rare tRNA) or BL21 Star Rosetta (Invitrogen) *E.coli*. Small scale expression trials were performed in a final culture volume of 50 ml, using varying amounts of inducer, various lengths and temperature of induction, before extraction of the protein in a similar manner as

described in Chapter 3 (3.1.2.3). The results are not reported here, but the most significant levels of expression (as judged by our collaborator and ourselves) were re-examined. Expression of GI, SPY, ZTL, FKF1, LKP2, LHY and CCA1 were greatest in the pMAL-c2x-cHis construct. ELF3 and TOC1 failed to express in any of the constructs provided, but the pseudo-response receiver domain (PRR) of TOC1 (TOC1-PRR) (Fig 4.1) was expressed in pET32a. Overexpression of ELF4 was also highest in pET32a and all these expression trials are summarised by SDS-PAGE (Fig 4.1).

Fig. 4.1 indicates an obvious overexpression of ELF4, CCA1, LHY, ZTL, FKF1, LKP2 and LIP1Δ234. It is encouraging to see overexpression of so many constructs and it appears that using MBP as a fusion partner to aid soluble expression is an effective way of expressing these clock proteins. However, the quantities required (mg) for crystallisation does mean that removal of the fusion protein will be extremely costly and therefore not viable for so many proteins. For these reasons, cloning the clock proteins into a modified pBAD vector which contains MBP and also a Tabacco Etch Virus (TEV) protease cleavage site would provide a financially viable alternative. There is an abundant supply of TEV protease in the Structural Laboratory which has been purified by Dr Roper (personal communication).

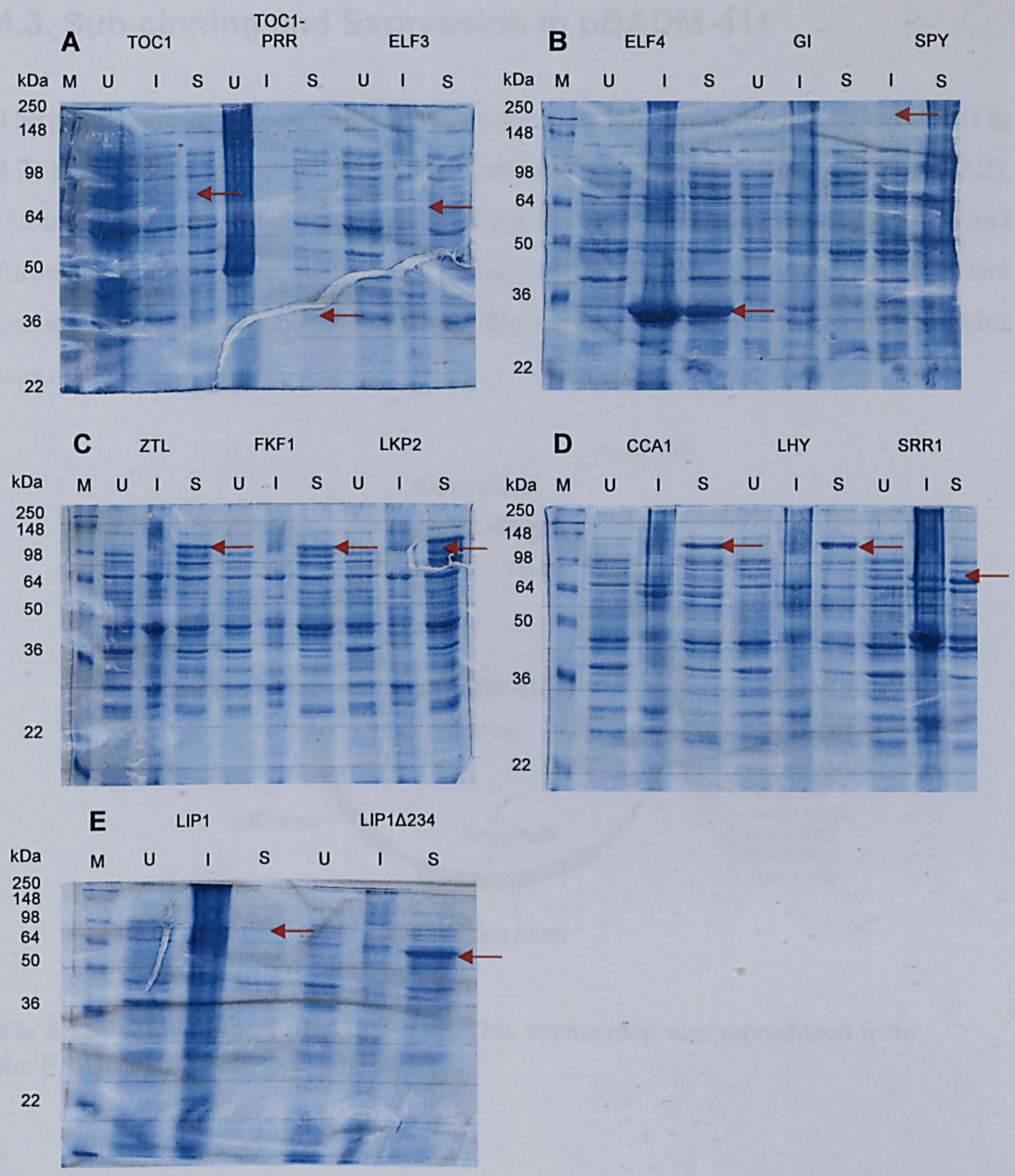


Fig 4.1. SDS-PAGE showing small scale expression trials. For each gel, 20 µg of protein was loaded, as determined using a BSA standard curve (appendix). In each case, M indicates protein markers in kDa, U are un-induced cell fractions, I are induced fractions and S are soluble fractions. Red arrows highlight the Mw of the proteins. (A) There is no obvious over-expression of TOC1, TOC1-PRR or ELF3. (B) There is no obvious overexpression of either GI or SPY. ELF4 is clearly overexpressed in pET32a with a visible band at 32 KDa. (C) The F-box proteins ZTL, FKF1 and LKP2 are all soluble, expressed as MBP-fusion proteins. (D) CCA1 and LHY are both overexpressed (~110 kDa) as MBP fusions. SRR1 is not obviously overexpressed. (E) Full length LIP1-MBP fusion (~70 KDa) is not overexpressed whereas LIP1Δ234 truncate is clearly overexpressed (~58 KDa).

4.3. Sub-cloning and Expression in pBADM-41+

The genes were sub-cloned from $pPro_{EX}$ constructs into a modified pBAD vector (Fig 4.2) as described in section 3.2. After ligation of the gene and vector (section 3.2.2), plasmid DNA was tested for presence of the insert by restriction digest (Fig 4.3) and further confirmed by sequencing. Cloning of TOC1-PRR, CCA1 and LKP2 were unsuccessful after several attempts and the decision to focus on the other proteins was made.

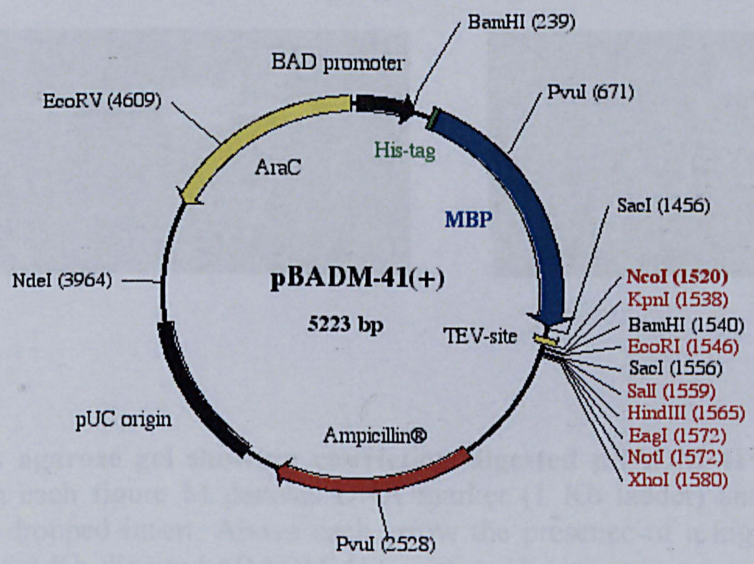


Fig 4.2. Vector map of pBADM-41+. This vector map was reproduced from the EMBL database.

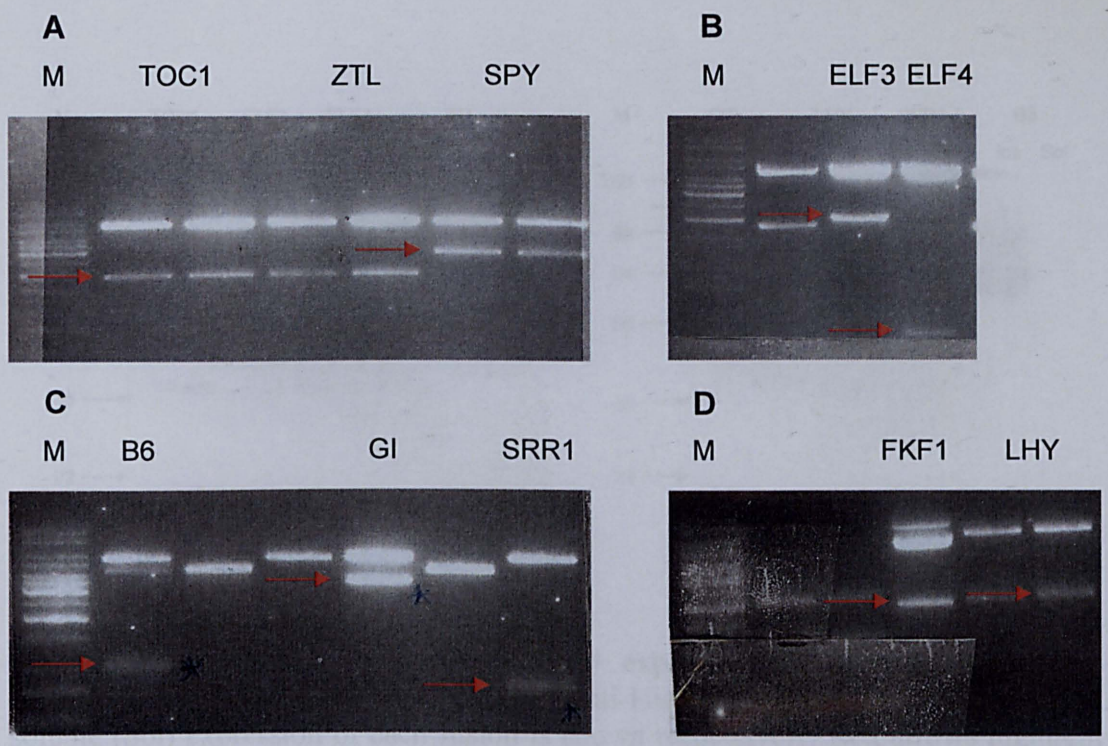


Fig 4.3. 1 % agarose gel showing restriction digested pBADM-41+ constructs for insert. In each figure M denotes DNA marker (1 Kb ladder) and red arrows represent the dropped insert. Above each arrow the presence of a higher Kb band indicates the 5.2 Kb digested pBADM-41+ vector. The inserts sizes are as follows; (A) TOC1 at 1.8 Kb, ZTL at 1.8 Kb and SPY at 2.7 Kb, (B) ELF3 at 2.1 Kb and ELF4 at 0.33 Kb, (C) B6 at 1 Kb, GI at 3.5 Kb and SRR1 at 0.8 Kb, (D) FKF1 at 1.8 Kb and LHY at 1.9 Kb.

The overexpression of pBADM-41+ constructs was evident for both ELF4 and ZTL, but less so for other constructs. The solubility and yield of overexpression was confirmed by Western analysis (Fig. 4.4) as described in section 3.2.4. TOC1 and ELF3 failed to express in pBADM-41+, whilst SRR1 and B6 were only expressed insolubly. Both LHY and SPY were overexpressed with approximately half the yield being insoluble.

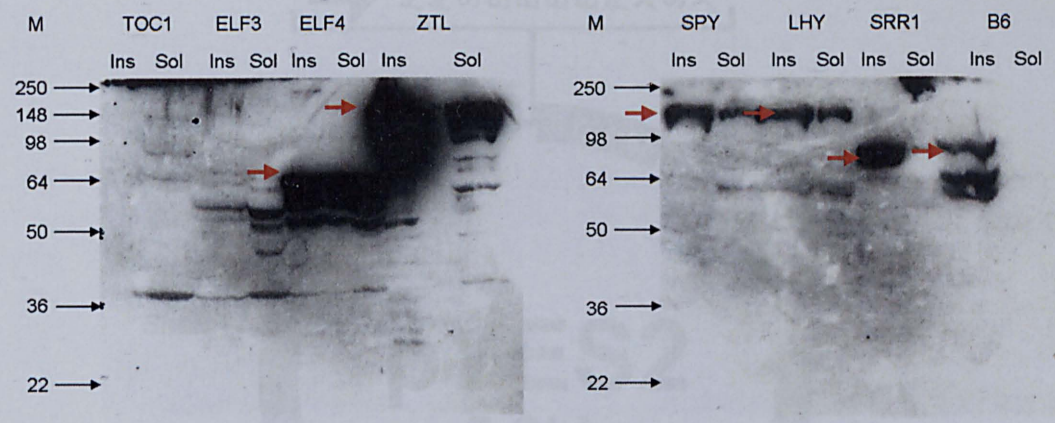


Fig 4.4. Western analysis of pBADM-41+ expression. The expression of His-tagged MBP-fusions was detected using anti-His antibodies. Insoluble (Ins) and soluble (Sol) expression of each fusion is shown respectively. Red arrows highlight the correct protein band. M denotes protein standard protein markers in kDa.

4.4. Expression of TOC1 and ELF3 in Yeast

As all expression trials of TOC1 and ELF3 had failed, an attempt to express both in yeast (section 3.3) was made. The genes were cloned into the expression vector pYES2 (Fig 4.5) and these were expressed in *Saccharomyces cerevisiae* strains DB2061 and INVSc1. The transformed Yeast cells failed to grow on the initial 2 attempts, but grew on the 3rd trial. No expression of TOC1 or ELF3 was detected by either SDS-PAGE or Western blots against the His-tagged proteins.

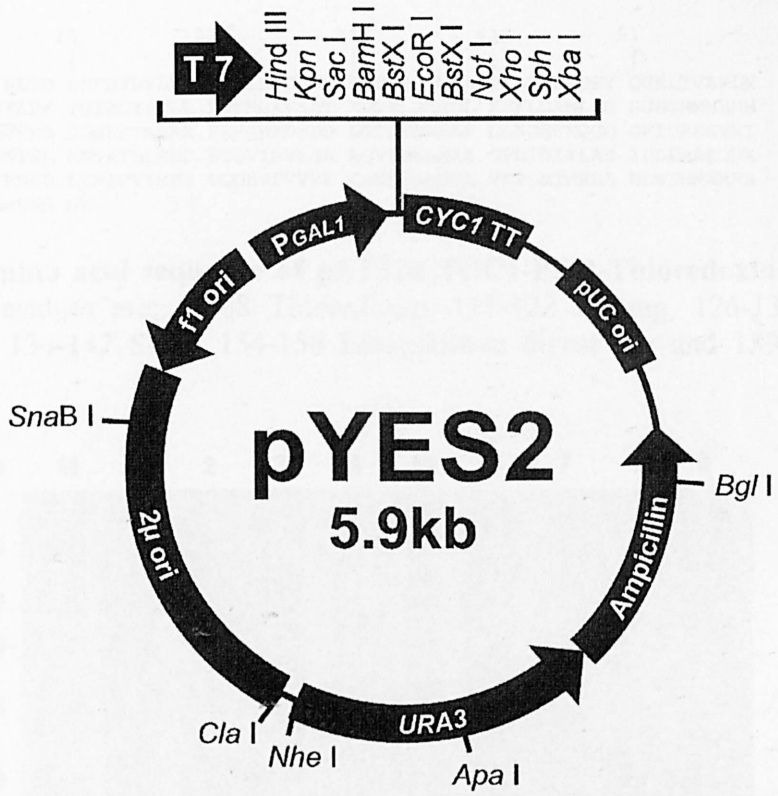


Fig 4.5. Vector map of pYES2. This map was reproduced from the pYES2 manual from Invitrogen.

4.5. Expression of TOC1-PRR and Ni-affinity purification

As expression of the full-length TOC1 was unsuccessful, focus turned to purification of the PRR domain. The N-terminal PRR domain of TOC1 was expressed in $pPro_{EX}$ and pET32a vectors (Fig 4.6) and initially purified by Ni-column chromatography as described in section 3.4 and 3.4.1. SDS-PAGE analysis of the expression and purification of the TOC1-PRR (Fig 4.7) shows large soluble overexpression of the 34 KDa TOC1-PRR-Thioredoxin fusion in the pET32a vector (Fig. 4.7, lane 7). The fusion protein was eluted from the Ni-column with 100 mM suggesting loose binding to the Ni-column as shown in Fig 4.7 lane 8 and as the predominant peak in Fig 4.8. The elution after addition of 0.5 M imidazole (Fig 4.7, lane 9) produced the purest fraction, which was used in subsequent digestion.

```
  1          11         21         31         41         51
  |          |          |          |          |          |
  1 MSDKIIHLTD DSFDTDVLKA DGAILVDFWA EWCGPCKMIA PILDEIADEY QGKLTVAKLN    60
 61 IDQNPGTAPK YGIRGIPTLL LFKNGEVAAT KVGALSKGQL KEFLDANLAG SGSGHMHHHH   120
121 HHSSGLVPRG SGMKETAAAK FERQHMDSPD LGTDDDDKAM DLNGECKGGD GFIDRSRVRI   180
181 LLCDNDSTSL GEVFTLLSEC SYQVTAVKSA RQVIDALNAE GPDIDIILAE IDLPMAKGMK   240
241 MLRYITRDKD LRRIPVIMMS RQDEVPVVVK CLKLGAADYL VKPLRTNELL NLWTHMWRRR   300
301 RMLGLEHHHH HH
```

Fig 4.6. Amino acid sequence of pET32a TOC1-PRR-Thioredoxin fusion. The following residues are; 1-108 Thioredoxin, 117-122 His-tag, 126-131 Thrombin digest site, 134-147 S-tag, 154-158 Enterokinase digest site and 159-305 TOC1-PRR.

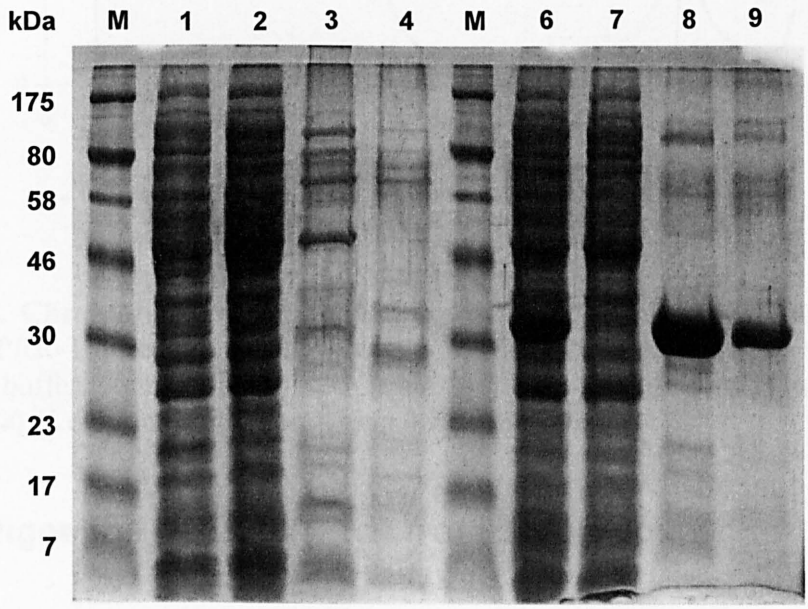


Fig 4.7. SDS-PAGE of TOC1-PRR expression and Ni-column purification. M denotes marker proteins with Mw in kDa. Lanes 1 and 6 show soluble overexpression of TOC1-PRR fusion in $pPro_{EX}$ and pET32a respectively. Lanes 2 and 7 represent flow-through fractions, lanes 3 and 8 are elution fractions after 100 mM imidazole wash and lanes 4 and 9 are elution fractions after addition of 0.5 M imidazole.

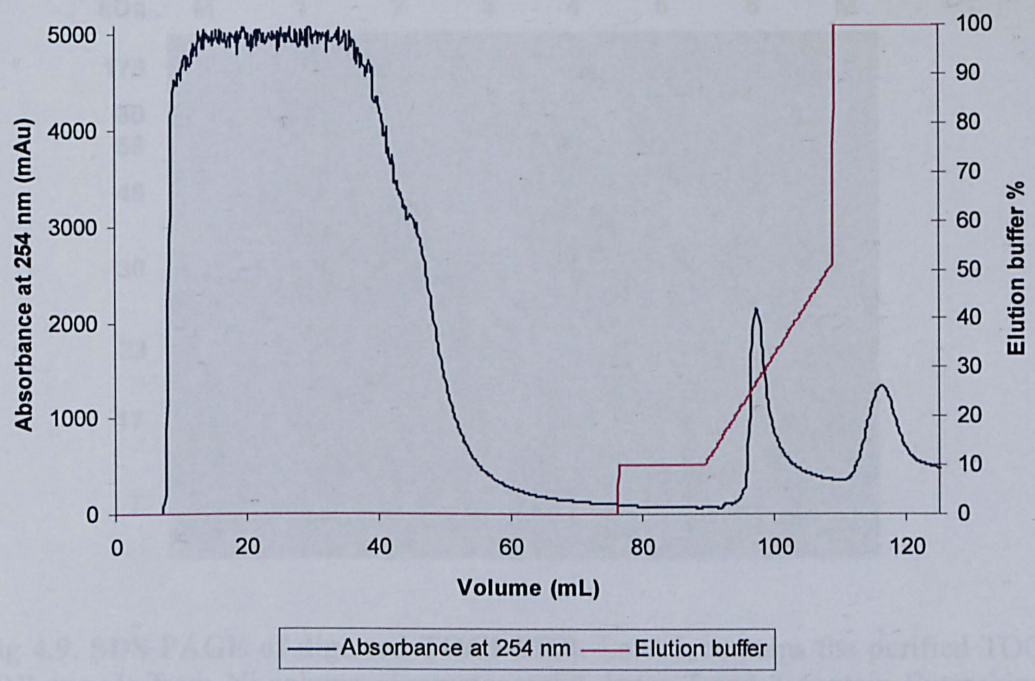


Fig 4.8. Chromatogram of Ni-affinity purification of pET32a-TOC1-PRR. The TOC1-PRR-Thioredoxin fusion shows a major peak elution around 90 ml at 20 % elution buffer (containing 100 mM Imidazole) and a smaller second peak around 115 ml at 100 % elution buffer (containing 0.5 M Imidazole).

4.6. Digestion of TOC1-PRR from Thioredoxin

The pET32a vector was digested by Enterokinase and Thrombin to remove the Thioredoxin and to determine which protease provided more successful cleavage. SDS-PAGE indicated that both proteases were efficient at removing the fusion protein (Fig 4.9). The Thrombin digest results in the removal of the 14 kDa Thioredoxin-polyHis (Fig 4.9, lane 5, lower band) from the 20 kDa S-tagged PRR (Fig 4.9, lane 5, upper band). However, according to the construct (Fig 4.6) digesting by Enterokinase will result in only the PRR domain and is therefore the preferred method of digestion. Unfortunately, digesting with Enterokinase resulted in precipitated TOC1-PRR. This precipitate was resuspended in 6 M urea for the purpose of SDS-PAGE, but this is not a practical solution when considering using the product for crystalisation trials.

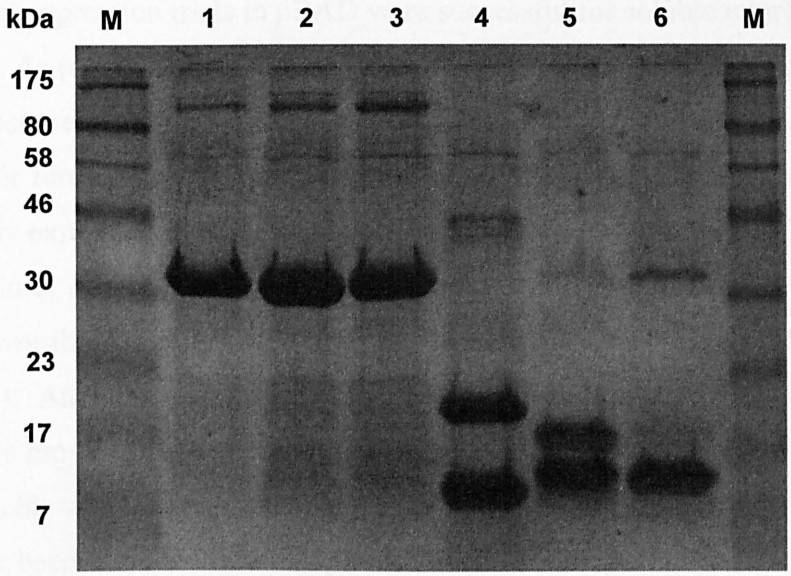


Fig 4.9. SDS-PAGE of digested TOC1-PRR. Lane 1 contains the purified TOC1-PRR sample from Ni-column chromatography, lanes 2 and 3 contain Enterokinase and Thrombin buffer exchanged TOC1-PRR respectively. Lane 4 shows the Thrombin digested sample, whilst lanes 5 and 6 indicate Enterokinase digested sample. Lane 6 represents the precipitated product of Enterokinase digestion re-suspended in 6 M urea.

4.7. Discussion

The aim of this thesis is to investigate the structures of circadian proteins by using x-ray crystallography. For this, mg quantities of protein are required and the evidence shown in Fig 4.1 would suggest that whilst the initial vectors appear successful at producing moderate levels of protein, a financially viable alternative was required. This set of vectors requires digestion of fusion protein using expensive proteases. The expense of the proteases required for digestion of mgs of fusion protein and the large scale-up required for significant levels of protein expression made existing constructs a practical impossibility. With this is mind, it was a logical progression to try and express the proteins in pBAD vectors in accordance with the CESG protocol. Not only does this system offer an increased solubility for ~80 % of proteins tested (Jeon *et al.*, 2005), but the MBP-fusion can be cleaved using TEV protease which is in plentiful supply in the laboratory.

The initial expression trials in pBAD were successful for soluble expression of ELF4 and ZTL. As previously mentioned, emphasis is placed on determining the structures of the positive regulators of *LHY* and *CCA1* and for this reason; pBAD-ELF4 was chosen for further experimentation (Chapter 6). Other positive regulators have been previously expressed in other laboratories. ELF3 was expressed in pET30a in *E. coli* BL21 (Liu *et al*, 2001) which is a similar expression system to pET32a. It has also been shown that GI can be co-expressed as MBP-GI with pETcoco-SPY (Tseng *et al*, 2004). Attempts to express ELF3 and GI according to the literature using constructs provided by the authors were unsuccessful. It is difficult to decipher why the protocols were not reproducible, but one suggestion is that important information may have been excluded from the publications.

As previously determined, no expression was detected for TOC1 and ELF3 in the initial trial (Fig 4.1), pBAD expression trials (Fig 4.4) or yeast expression trials (section 4.4). There were no protocols in the literature for expression of TOC1 which lead to attempting to express the TOC1-PRR domain. The PRR domain offers a great palace to start crystallisation trials; it is present in multiple clock proteins and is likely to play a key role in their functions, and it is quite small (16.5 kDa) allowing easier crystallisation and crystallographic analysis.

This domain showed high soluble expression (Fig 4.7) and was able to be purified in significant quantities as a Thioredoxin fusion. However, the digestion of the fusion is problematic. Digestion using Enterokinase removes all tags and the Thioredoxin fusion, but results in precipitation of the TOC1-PRR protein. Conversely, digestion using Thrombin maintains TOC1-PRR solubility, but also leaves the S-tag attached. It is possible to crystallise proteins with fusion partners and / or tags, but it can be difficult to determine the effect of such interactions on the native protein structure. The unstructured nature of the S-tag leads to increased solubility, but may make crystallisation more difficult.

The overexpression of the PRR domain was deemed successful, however due to time constraints no scale ups were attempted. It would be possible to continue this work by scaling up and attempt crystallisation. That said, the presence of the S-tag is likely to be required for obtaining reasonable amounts of protein.

Chapter 5. Development of DNA aptamers

5.1. Introduction

The development of affinity tags against proteins is of significant importance in molecular and biochemical research. Traditionally, the production of antibodies has been the most convenient way of providing specificity and allowing a wide range of experiments to be conducted. These experiments range from the very basic identification of a particular target, such as western analysis, to the very complex, including DNA/protein interactions, protein/protein interactions and even for the purification of proteins from whole cell extracts. Whilst antibodies are a staple feature of everyday research, there are problems concerning their use.

The first major step in their production involves the injection of the protein into live mammals, such as mice or rabbits. The animal is then bled and the resulting serum purified for the antibody in question. This is time consuming as the protein needs a significant incubation period to illicit the immune response that produces the antibody. It is also costly, in terms of keeping the animals and purifying the serum. In addition, the specificity of the antibody can vary from highly specific to non-specific. When considering the large number of proteins investigated in this thesis, it makes sense to find alternative affinity tags to replace the antibodies usually used.

In the early 1990's, several laboratory's reported an in vitro selection and amplification technique for the isolation of nucleic acid sequences that bind a target with high specificity and affinity (Robertson and Joyce, 1990; Tuerk and Gold, 1990; Ellington and Szostak, 1990). The technique was named SELEX, an acronym for Systematic Evolution of Ligands by EXonential enrichment, and these highly specific nucleic acids (Both RNA and later, single stranded (ss) DNA) were given the name aptamers. Aptamers bind non-nucleic acid targets including peptides, proteins, organic / inorganic molecules and drugs, based on their three dimensional structures, which include hairpins, loops, stems, pseudoknots, triplexes and quadruplexes. Aptamers show a specificity towards their targets that is at least comparable to the speciifictiy of antibodies towards their targets, with K_d values in some cases, reaching the pmol range (Jenison *et al.*, 1994). Furthermore, aptamers

specificity have been shown to discriminate between 10,000 and 12,000 fold (Jenison *et al.*, 1994; Geiger *et al.*, 1996) to that of their targets, even between cases where there is significant structural similarity of the target.

The high affinities and specificities of aptamers show the potential that they present in terms of diagnostic, bio-analytical and therapeutic applications. Furthermore, aptamers have significant advantages in their production over existing methods used for the production of monoclonal antibodies. Aptamers are made by the in vitro process SELEX, which removes the need for live animals and extends the list of potential targets to non-immunological and even toxic targets, which would be excluded during the production of monoclonal antibodies. In addition, aptamers can be produced against specific regions of a target, rather than relying upon the immune system of an animal selecting the epitopes of the target. As aptamers are selected from an oligonucleotide library containing large quantities of nucleaic acid (typically 10^{15}) and go through many rounds of selection, the affinity achieved can be greater than that of natural selection. The direct sequence knowledge of the aptamers allows for easy modification, including the introduction of an enzyme or fluorophores, or even immobilisation on a resin or micro-array that will not decrease or interfere with the aptamers ability to bind its target. Above all other considerations, arguably the most useful property of aptamers production is the conditions under which they are produced can be non-physiological to reflect the experiment or assay they are required for. They are much more likely to be stable and function under these conditions than an antibody.

5.1.1 General principle of aptamer selection by SELEX

The process starts with the synthesis of a degenerate nucleic acid library, consisting of 10^{14}-10^{15} DNA or RNA molecules that have a random region flanked by fixed regions that allow amplification of the sequence by PCR. The library is commonly incubated with an immobilised target (microtitre plates, affinity column, magnetic beads) and given time to form unique specific three-dimensional structures against the target. The bound oligonucleotides are separated from the unbound and non-specific nucleic acid sequences by a series of stringent wash steps. Target-bound oligonucleotides are eluted by the introduction of an affinity elution, and this

depends upon the kind of immobilisation performed. Once the aptamers have been separated from the target, they can be amplified by PCR. Resulting double-stranded DNA is separated into ssDNA, before being incubated with the target once more, ready for the next SELEX round (not necessary for RNA). Iterative cycles of selection and amplification decreases the size of the random nucleic acid library and increases the specificity of the aptamers towards the target (Fig 5.1). The number of rounds required and the specificity of the aptamer pool at the end of SELEX will vary from the methodology and the targets used. The enriched pool is then cloned and sequenced to determine the exact nucleotide sequence responsible for the affinity. Individual clones are usually screened for specificity by using fluorescent or radioactive binding assays, Surface plasmon resonance (SPR) or fluorescence correlation spectroscopy. A typical SELEX will yield approximately 50 specific aptamers. ClustalW analysis can help with identifying homologous sequences, but the specificity is usually determined by the aforementioned techniques.

5.1.2 Aptamer selection against circadian clock proteins

To determine the structural properties of the putative clock proteins, we require large amounts of protein, some of which can be used for the production of aptamers. If we are able to make aptamers towards the clock proteins, it will allow us and others to design experiments that go beyond structural analysis. As we are expressing many proteins, the development of antibodies is too expensive. We can therefore attempt to develop a SELEX protocol for the production of aptamers against any circadian clock associated protein that we purify.

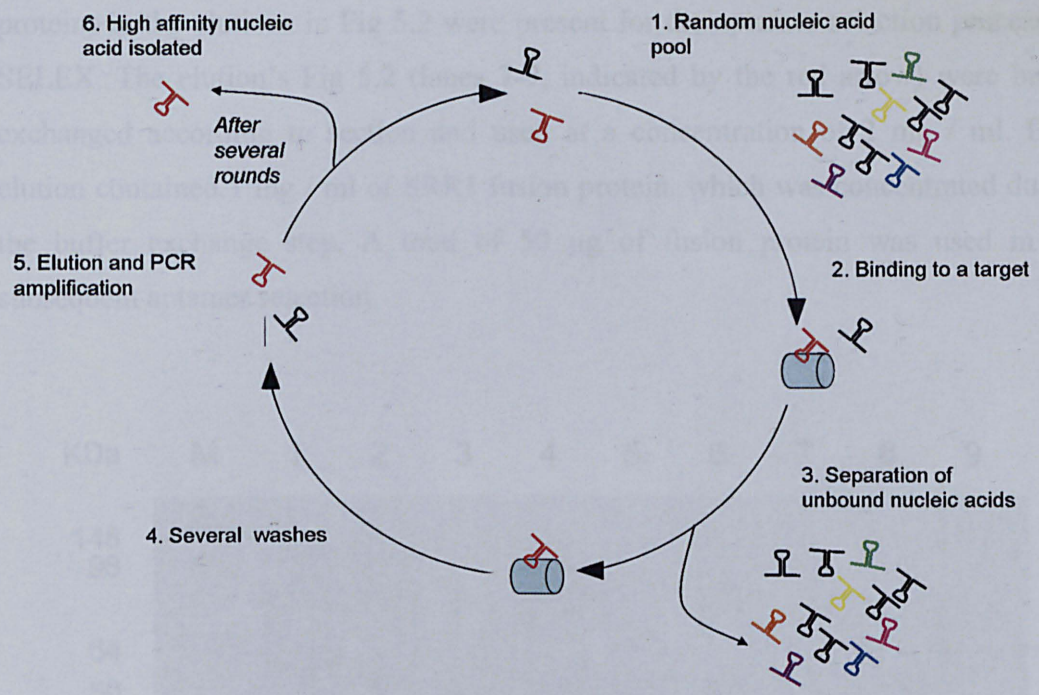


Fig 5.1. A schematic showing the production of DNA aptamers based on Systematic Evolution of Ligands by EXponential enrichment (SELEX). (1) A random single stranded DNA (ssDNA) library is incubated with a target, in this case a protein. (2) The ssDNA forms unique three dimensional structures around the target, with a minority showing specificity. (3) Unspecific ssDNA is removed from the target by gentle removal of the protein from the pool. (4) Several stringent washes remove any remaining unspecific ssDNA that may be associated with the protein due to steric interactions. (5) ssDNA is eluted from the protein by applying a high salt wash. The resulting eluate is amplified by PCR, enriching the pool for DNA that has specificity for the protein. The double stranded DNA is denatured to produce ssDNA that is responsible for the required specificity. The newly amplified pool is the subjected to further rounds of enrichment and amplification (steps 1-5), typically between 12-15 rounds, with each round improving the specificity of the aptamer pool for the protein. (6) The resulting pool contains highly specific sequences of ssDNA which can be sequenced, produced and used in subsequent experiments.

5.2. Results

The basis for aptamer selection was due to the attempt to purify numerous proteins. Although the main focus of the thesis did not involve the crystallisation of SRR1, it was chosen due to the relative ease at which it was purified. The MBP-SRR1 fusion is approximately 72 KDa and was purified using Ni-chromatography as described in the previous chapter. The resulting SDS-PAGE is shown in Fig. 5.2. Although the SRR1 appears to contain numerous contaminants, these were significantly less in concentration than other circadian proteins after one affinity purification. Other

proteins in the elutions in Fig 5.2 were present for the aptamer selection process by SELEX. The elution's Fig 5.2 (lanes 7-9, indicated by the red arrow) were buffer exchanged according to section and used at a concentration of 2 mg / ml. Each elution contained 1 mg / ml of SRR1 fusion protein, which was concentrated during the buffer exchange step. A total of 50 μg of fusion protein was used in the subsequent aptamer selection.

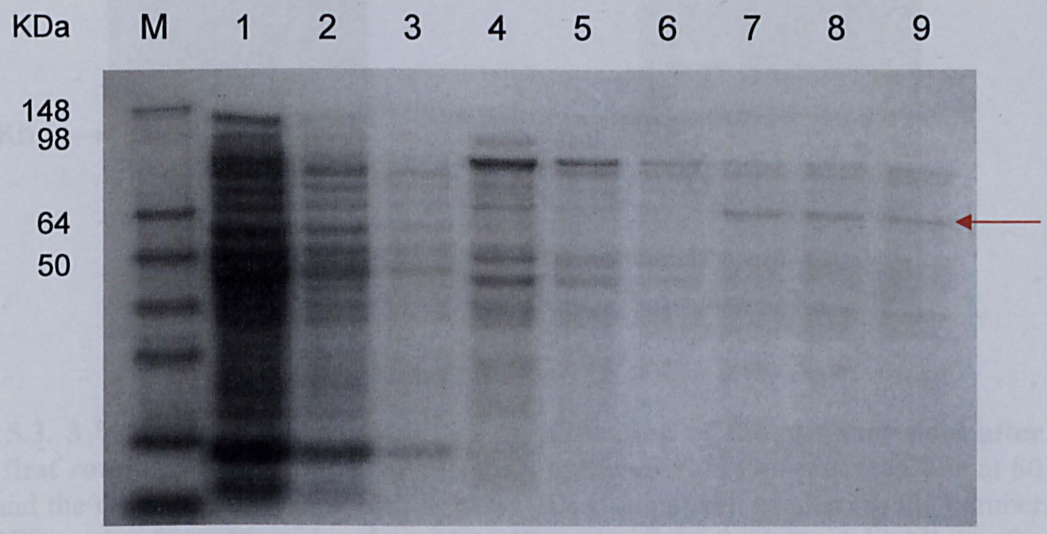


Fig 5.2. SDS-PAGE showing Ni-affinity chromatography purification of MBP-SRR1. Lane 1 shows the soluble lysate, lane 2 contains flow-through, lanes 3 and 4 contain the fractions after 100 mM imidazole washes, lanes 5 and 6 contain the fractions after 200 mM imidazole washes and lanes 7-9 represent elution of MBP-SRR1 with 300 mM imidazole, highlighted by the red arrow. M denotes protein standards in KDa.

After the first round of SELEX, amplification of the aptamer pool required optimisation. The initial PCR conditions differed from the procedure detailed in the methods section and produced two products, of which one was the correct size (Fig 5.3 (A)). The likely explanation is due to mis-annealing of the LICSelexF and LICSelexR primers which result in the production of conaptamers (joined aptamers). Optimisation of the protocol included reducing the annealing temperature from 56 °C to 50 °C and optimising the concentration of the co-factor $MgSO_4$. In addition, the number of PCR cycles had to be carefully monitored. Reducing the number of cycles from 25 to 20 or 15 still produced the double band. As few as 10 cycles were required to produce a product of the correct 80 bp size (Fig 5.3, (B, lane 2)). Figure

3.3 (B) highlights the difference in product between 15 and 10 PCR cycles (lanes 1 and 2, respectively).

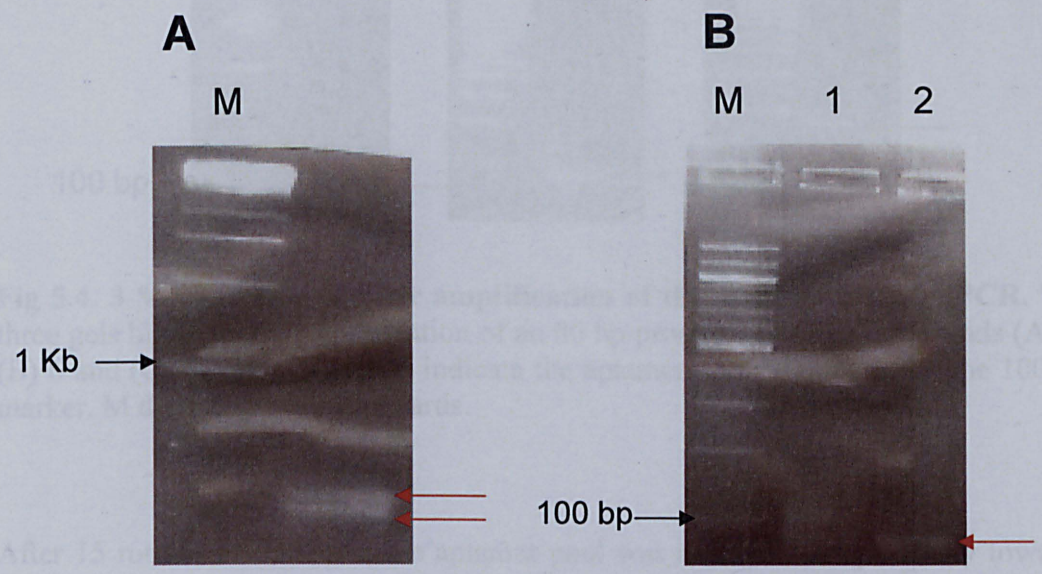


Fig 5.3. 3 % agarose gels showing the amplification of the aptamer pool after the first round of SELEX. (A) Shows the amplification of two products, one at 80 bp and the other slightly larger (red arrows). (B) Comparison of altering the number of PCR cycles from 15 cycles (lane 1) to 10 cycles (lane 2). Lane 1 highlights the presence of two bands and lane 2 contains the expected 80 bp oligonucleotide, indicated by the red arrow. M denotes protein standards.

Unfortunately, too few cycles result in no product at all and too many cycles produce bands of incorrect sizes. For these reasons, the aptamer pool was amplified using 10 cycles in the first instance and if no product was detected, the pool was subjected to a further 5 cycles. This eliminated the production of a double band and also proved that the pool contained ssDNA of the appropriate size. In total, 15 rounds of SELEX were performed and the aptamer pool was visualised after each round. Figure 5.4 shows the amplification of the aptamer pool after rounds 3 (A), 6 (B) and 9 (C). Any one of the rounds could have been selected for this figure.

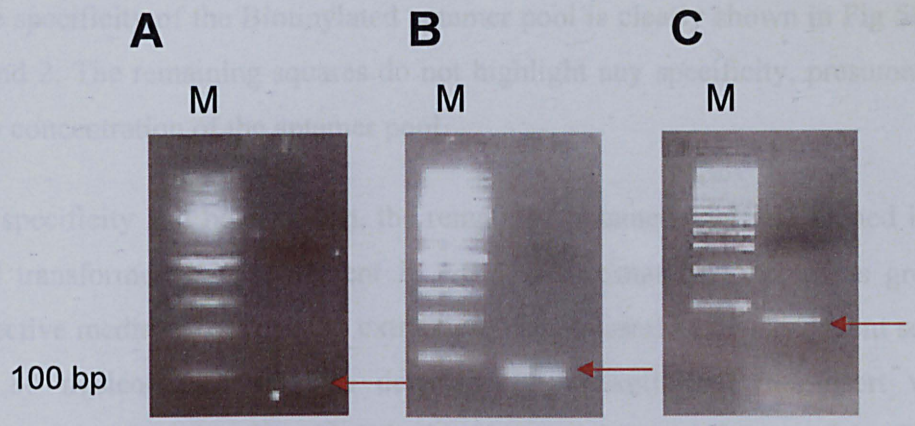


Fig 5.4. 3 % agarose gels after amplification of the aptamer pool by PCR. The three gels highlight the amplification of an 80 bp product after SELEX rounds (A) 3, (B) 6 and (C) 9. The red arrows indicate the aptamer pool slightly under the 100 bp marker. M denotes protein standards.

After 15 rounds of selection, the aptamer pool was assessed for specificity towards SRR1 by a Dot-blot analysis. In the final amplification step, the LICSelexF was replaced with the biotinylated primer which resulted in the production of Biotinylated aptamers. The SRR1 was blotted on to a nitrocellulose membrane and incubated with the biotinylated aptamer pool. Specificity towards the SRR1 fusion was detected by Dot-blot analysis using a streptavidin-HRP conjugated antibody (Fig. 5.5).

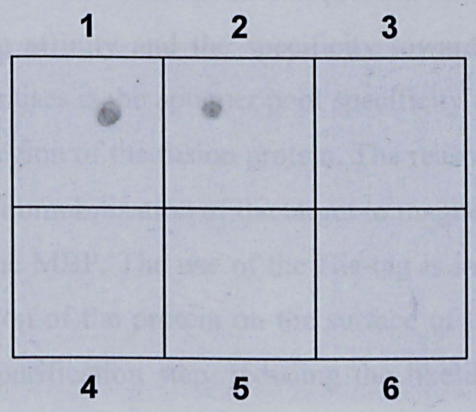


Fig 5.5. Dot-Blot of Biotinylated aptamer pool incubated with SRR1. The aptamer pool was diluted in ten-fold serial dilutions (1-6) from the initial Biotinylated pool (square 1). 5 µl of the aptamer pool was added to each square. Squares 1 and 2 show the detection of Biotinylated aptamers that have bound to SRR1.

The specificity of the Biotinylated aptamer pool is clearly shown in Fig 5.5, squares 1 and 2. The remaining squares do not highlight any specificity, presumably due to low concentration of the aptamer pool.

As specificity had been shown, the remaining aptamer pool was cloned into PSKII and transformed into competent *E. coli*. Approximately 35 colonies grew on the selective media, however after extraction of the plasmid and subsequent sequencing; no 80 nucleotide insert was detected. It is likely that the insert was never incorporated into the plasmid. Whether this was due to a failure of the T / A based cloning strategy or the ligation itself is unknown.

5.3. Discussion

The results clearly indicate that the selection protocol used in this thesis has produced aptamers with specificity towards SRR1 fusion protein (Fig 5.5). Unfortunately, the cloning of the aptamer pool did not provide any sequence information to elucidate nucleotides required for the observed specificity. Given more time, further attempts to clone the aptamer pool would have been made and these sequences screened for common nucleotide regions. That said, there would be plenty more testing of individual aptamer sequences in order to assess the specificity, in terms of binding affinity and the specificity towards the target itself. The first question this point raises is the aptamer pool specificity may be towards SRR1, MBP or even the linker region of the fusion protein. The reason for using the tagged fusion protein was for the immobilisation of the target to magnetic beads, via the N-terminal His-tag found on the MBP. The use of the His-tag is important because it promotes the proper orientation of the protein on the surface of the bead, and simultaneously provides an extra purification step, reducing the likelihood of producing aptamers towards a contaminant. Indeed, the purified SRR1 fusion shows several other proteins are present in the elution (5.2), so this was an important feature of this particular aptamer selection process. Subsequent aptamer specificity would have to be tested against purified SRR1 and purified MBP alone. It would be easy to synthesise the individual aptamers with tags or fluorescent molecules and perform western blots against lanes of either purified protein. This would certainly allow the

specificity to be determined, however, it would not indicate the specificity for these targets alone. If the aptamers were to be used in any situation where other proteins are present, we would need to infer that there is minimal non-specific binding. For example, if the interest lies in determining the relative abundance of a circadian associated protein over the daily light / dark cycle, how could we be sure that the aptamer binds only to our target? To this end, a range of experiments utilising whole *Plant* cell extracts would required. It is feasible to test individual aptamers against whole cell extracts run on an SDS-PAGE followed by western analysis. Highly specific aptamers should only produce one band on such blots. Combining the results of this kind of experiment with sequence alignment would provide nucleotide sequences that are a specific towards the target as possible. In any case, the very same problems exist with non-specific binding of antibodies and this does not prevent their use.

The SELEX protocol developed in this thesis highlights a straightforward and efficient method of developing affinity tags towards proteins. It has incorporated our knowledge of the strong and highly specific biotin-streptavidin interaction for; generating ssDNA from biotinylated PCR products using streptavidin magnetic beads, and also the detection of biotinylated aptamers specificity by western blots. If time had allowed, we could have further used this interaction to immobilise biotinylated aptamers to streptavidin beads for the purification of a protein target from cell extracts. This would have been particularly useful in isolating protein: protein interactions between circadian associated proteins. Moreover, using aptamer affinity chromatography would allow a means of purification without the need for affinity tags that may adversely affect protein structure, function or ability to form crystals for structural determination. Perhaps the most exciting use for developing aptamers in this project however, lies in the concept of producing DNA arrays. In theory, we could have produced highly specific aptamers towards all of the circadian associated proteins previously discussed. The aptamers could then be bound to a DNA chip and whole cell extracts washed over the chip at different circadian times. It would be possible to determine the level of protein accumulation over the circadian cycle and to therefore infer possible interactions. If the LHY / CCA1 interaction had not been characterised, then an experiment along these lines would show similar levels of protein accumulation at similar points of the circadian cycle. This may infer

an interaction or working in a similar pathway / manner. To simultaneously be able to monitor the protein levels in a given system would be very powerful in understanding unidentified interactions.

The benefit of the work presented here, is that this protocol can be used and adapted to suit the requirements of any research project. Any protein containing any type of tag could be subjected to this process of SELEX. As long as the affinity steps have been considered and appropriate tags (protein or aptamer) have been used, this protocol can be used for the development of aptamers towards any protein. As previously discussed, the aptamers can be used in place of antibodies in everyday research, highlighting a massive potential in their use. Their biological properties suggest an advantage over using antibodies and these have been addressed. Indeed, recent years have seen the development of aptamer technology dramatically increase, with ever more publications showing either the use of aptamers in medical diagnostics or an improvement in the methodology of selection, suggesting replacement of antibody technology. There is also increasing interest in commercial development of aptamer technology, as indicated in the number of industrial patents filed over the past few years. The protocol here describes a way to manufacture aptamers, in short time frames and at low cost which would suit the average academic researcher.

Chapter 6. ELF4

6.1. Introduction

ELF4 is an interesting circadian protein that appears to exert its effect as a light input into the clock and also as a putative central oscillator protein (as discussed in section 1.4.2). Evidence suggests that it is fundamental for proper clock functioning as all free-running periods are attenuated in the *elf4* loss-of-function mutant (Kikis *et al.*, 2005). Interestingly, *ELF4* loss-of-function appears to have a more pronounced effect on clock function that any of the other potential oscillator proteins, which result in an altered period or amplitude of rhythm. For example, *TOC1* loss-of-function results in a shorter period of clock outputs (Millar *et al.*, 1995b; Somers *et al.*, 1998; Strayer *et al.*, 2000), but the rhythmicity of outputs is maintained. There appears to be no redundancy in the function of ELF4.

There are several reasons for wanting to study the ELF4 protein in greater detail. From a structural standpoint it is interesting in that ELF4 contains no known domains or motifs based on its primary amino acid sequence. As ELF4 is expressed at similar phase to GI, LUX and TOC1 it is possible that it forms a part of a complex with these proteins. By obtaining a crystal structure it may be possible to predict such interactions which we can not decipher from the sequence. Furthermore, ELF4 may represent a species specific protein with a novel structure which may have implications for the study of non-circadian proteins. Indeed, a BLASTp of ELF4 at the NCBI results in a list of ELF4-Like 1-4 proteins, highlighting other proteins of unknown function may also have a similar structure and may therefore function in a similar manner.

In addition, results show that pBADM-41+-ELF4 and pET32a both express very well (Fig 4.2. and 4.5) and this is a decent starting point to attempt purification and crystallisation.

6.2. pBADM-41+-ELF4 expression and purification

As determined previously, there was high expression of ELF4 in pBADM-41+ (Fig 4.5). The expression trials were scaled up to 8 L and the resulting culture prepared for purification as laid out in section 3.1.2.2 and 3.1.2.3. ELF4 was subjected to Ni-affinity chromatography followed by size-exclusion chromatography into TEV protease buffer. At this stage, there was approximately 80 mg of fusion protein produced. The resulting fraction was cleaved by TEV protease before subsequent removal of the His tagged MBP. Although it appeared that ELF4 and MBP had been separated, the His-tagged MBP failed to bind to the Ni-column. To overcome this problem, the fraction was applied to an amylose resin which should bind the MBP itself. As before, no binding was detected.

As separation based on the His tag and MBP had failed, purification utilising ELF4s intrinsic properties were applied. Digested fractions were applied to anion and cation exchange chromatography, ammonium sulphate cuts and size exclusion chromatography, none of which showed separation. Hydrophobic interaction chromatography (HIC) was also attempted (phenyl-Sepharose, buthyl-Sepharose and octyl-Sepharose columns), but ELF4 and / or MBP failed to bind to the columns used leading to co-elution. An attempt to purify the fusion protein utilising the properties of MBP was also attempted. ELF4 fusion was applied to an amylose resin which should immobilise the MBP. TEV protease was included in the wash and elution buffers in order to cleave ELF4 from the fusion. The MBP did not bind to the resin.

The type and order of the purification procedure appeared to make no difference in the separation of ELF4 from MBP. With every column used, the yield of ELF4 decreased but the purity did not improve. Consequently, purification of ELF4 in pET32a was trialled in place of pBADM-41+.

6.3. pET32a-ELF4 expression and purification

Expression of ELF4 in pET32a is shown in Fig 4.2 (B). The vector contains a Thioredoxin (Trx) fusion protein to aid solubility, and His and S-tags to aid in purification (Fig 6.1). The fusion was expressed as described in section 3.6.2, and subjected to Ni-affinity chromatography which resulted in a yield of 200 mg of protein. The resulting fraction (Fig 6.2.a, lane 4) was applied to size exclusion chromatography in order to buffer exchange ready for digest. Approximately 170 mg of protein was eluted from the gel filtration (Fig. 6.2.a, lane 5).

```
  1          11         21         31         41         51
  |          |          |          |          |          |
  1 MSDKIIHLTD DSFDTDVLKA DGAILVDFWA EWCGPCKMIA PILDEIADEY QGKLTVAKLN    60
 61 IDQNPGTAPK YGIRGIPTLL LFKNGEVAAT KVGALSKGQL KEFLDANLAG SGSGHMHHHH   120
121 HHSSGLVPRG SGMKETAAAK FERQHMDSPD LGTDDDDKAM ETKRNGETKR RRNVAEEAEQ   180
181 GEDPAMETWE NLDRNFRQVQ SVLDRNRSLI QQVNDNHQSR METADNMETS KNVALIQELN   240
241 GNISKVVNME TYSDLNTSFS SGFHGGKNGH DGGGAAGTRA LE
```

Fig. 6.1. Amino acid sequence of pET32a ELF4-Thioredoxin fusion. The following residues are; 1-108 Thioredoxin, 117-122 His-tag, 126-131 Thrombin digest site, 134-147 S-tag, 154-158 Enterokinase cut site, 160-281 ELF4.

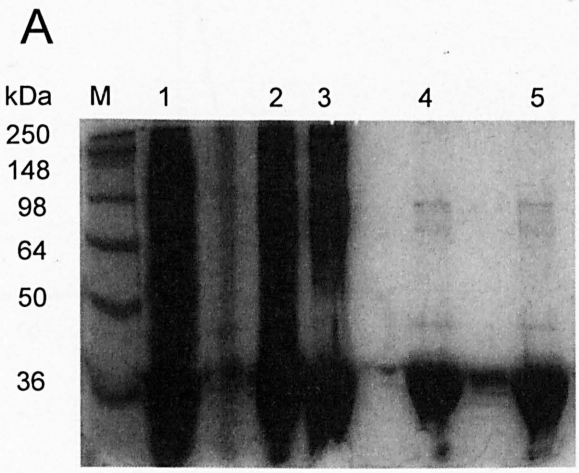
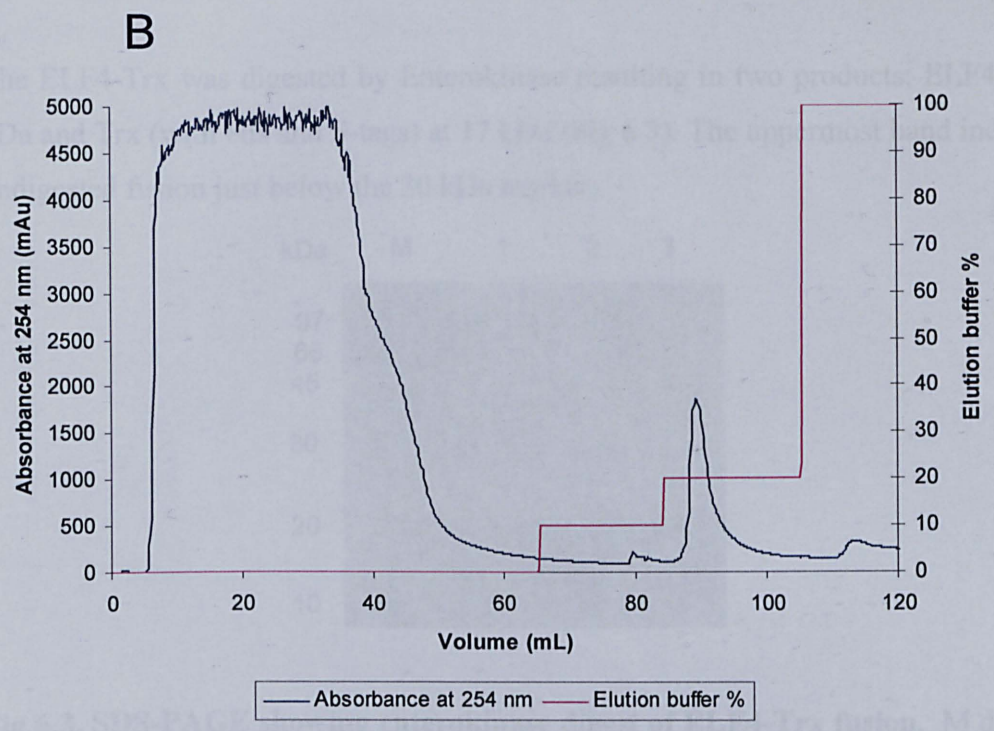


Fig. 6.2a. SDS-PAGE of ELF4-Trx fusion after Ni-affinity chromatography and size-exclusion chromatography. (A) M denotes protein markers. Lane 1 shows crude extract, lane 2 is induced fraction, lane 3 soluble fraction, lane 4 Ni-purified fraction and lane 5 ELF4-Trx after size-exclusion chromatography.



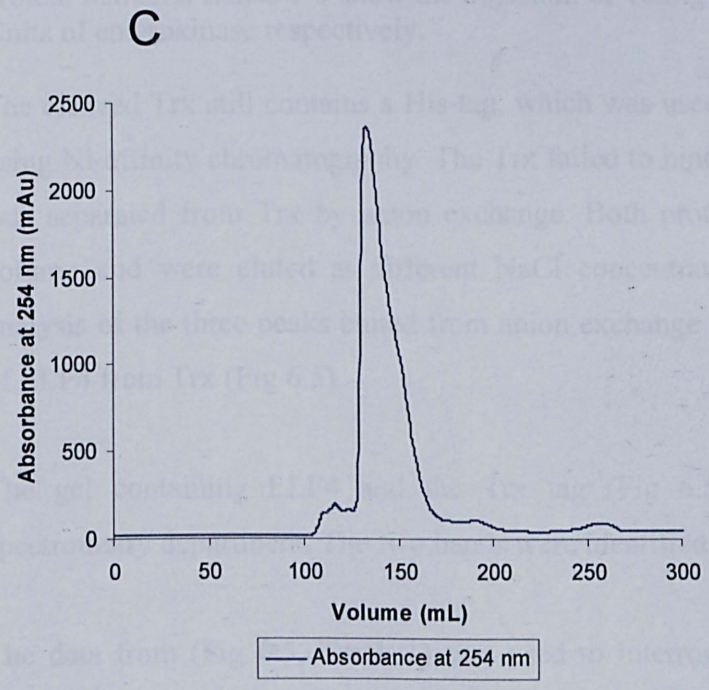


Fig 6.2b. Chromatographs showing Ni-affinity purification and size-exclusion of ELF4-Trx fusion. (B) ELF4-Trx fusion showed a peak at 20 % elution buffer containing 100 mM imidazole. (C) ELF4-Trx fusion after size-exclusion shows an absorbance of 2500 mAu at 254 nm.

6.4. ELF4-Trx digestion and anion-exchange.

The ELF4-Trx was digested by Enterokinase resulting in two products; ELF4 at 13 kDa and Trx (with His and S-tags) at 17 kDa (Fig 6.3). The uppermost band indicates undigested fusion just below the 30 kDa marker.

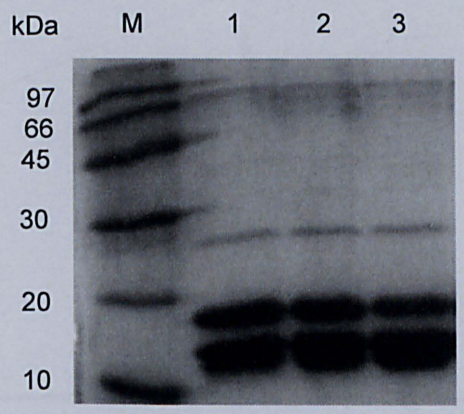


Fig 6.3. SDS-PAGE showing enterokinase digest of ELF4-Trx fusion. M denotes protein markers. Lanes 1-3 show the digestion of 10 mg of ELF4-Trx with 1,2 and 3 Units of enterokinase respectively.

The cleaved Trx still contains a His-tag, which was used as an anti-purification step using Ni-affinity chromatography. The Trx failed to bind to the Ni-column, so ELF4 was separated from Trx by anion exchange. Both proteins bound the Resource Q column and were eluted at different NaCl concentrations (Fig 6.4). SDS-PAGE analysis of the three peaks eluted from anion exchange clearly shows the separation of ELF4 from Trx (Fig 6.5).

The gel containing ELF4 and the Trx tag (Fig 6.5) was given to the Mass spectrometry department. The two bands were identified to be the correct proteins.

The data from (Fig 6.5, Track 1) was used to interrogate the *Arabidopsis* protein database (rel. 3.42 May 2008, http://www.ebi.ac.uk/IPI/IPIarabidopsis.html) and was identified as: Gene Symbol ELF4 (ELF4 EARLY FLOWERING 4, Accession IPI:IPI00528125.1|TREMBL:O04211|REFSEQ:NP_565922|TAIR:AT2G40080.1) with 2 peptides identified and 24% sequence coverage. The data from (Fig 4.5, Track 3) was used to interrogate the UniProt KB database rel. 13.3 May 2008 and the

protein was identified as: Thioredoxin – 1 Accession P0AA25 and homologues. The on-screen shots are shown in Figure 6.6.

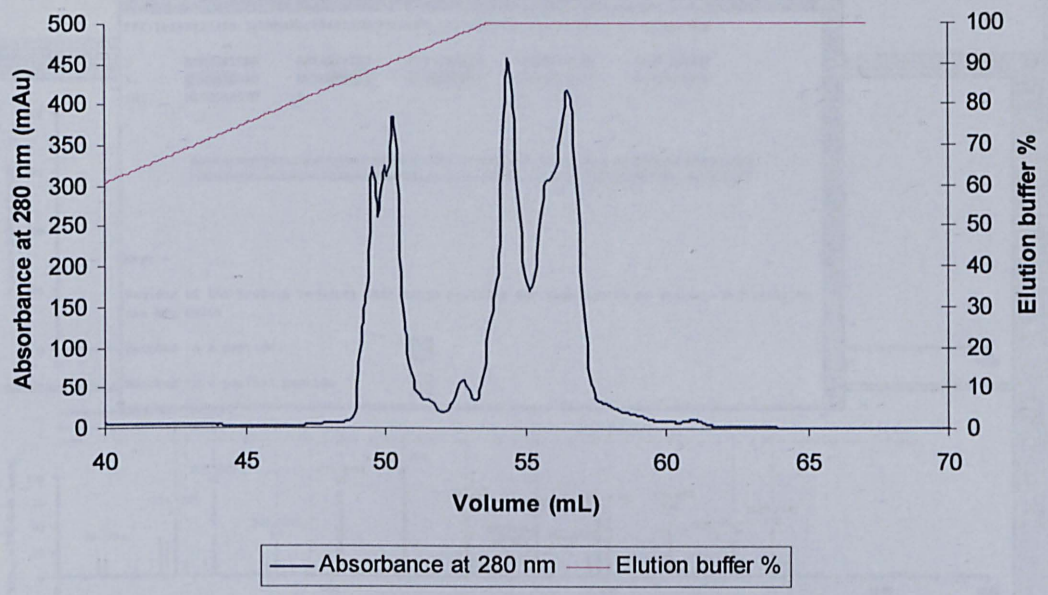


Fig 6.4. Chromatograph showing anion exchange of cleaved ELF4 and Trx. The peak around 50 ml with approximately 0.8 M NaCl represents ELF4, whilst the two converging peaks at 55 ml represent the elution of Trx.

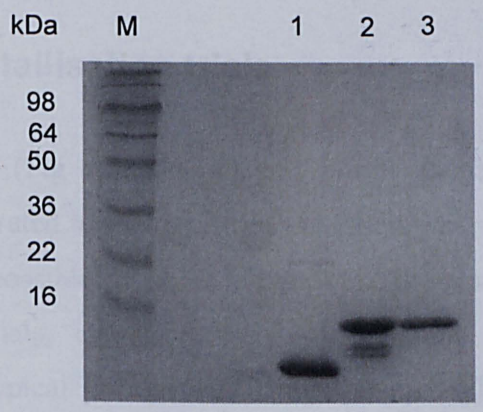


Fig. 6.5. SDS-PAGE of digested ELF4-Trx separated by anion exchange. Lane 1 is pure ELF4 (12 kDa), lane 2 contains the Trx tag and degradation products, lane 3 contains pure Trx (16.5 kDa). M denotes protein markers.

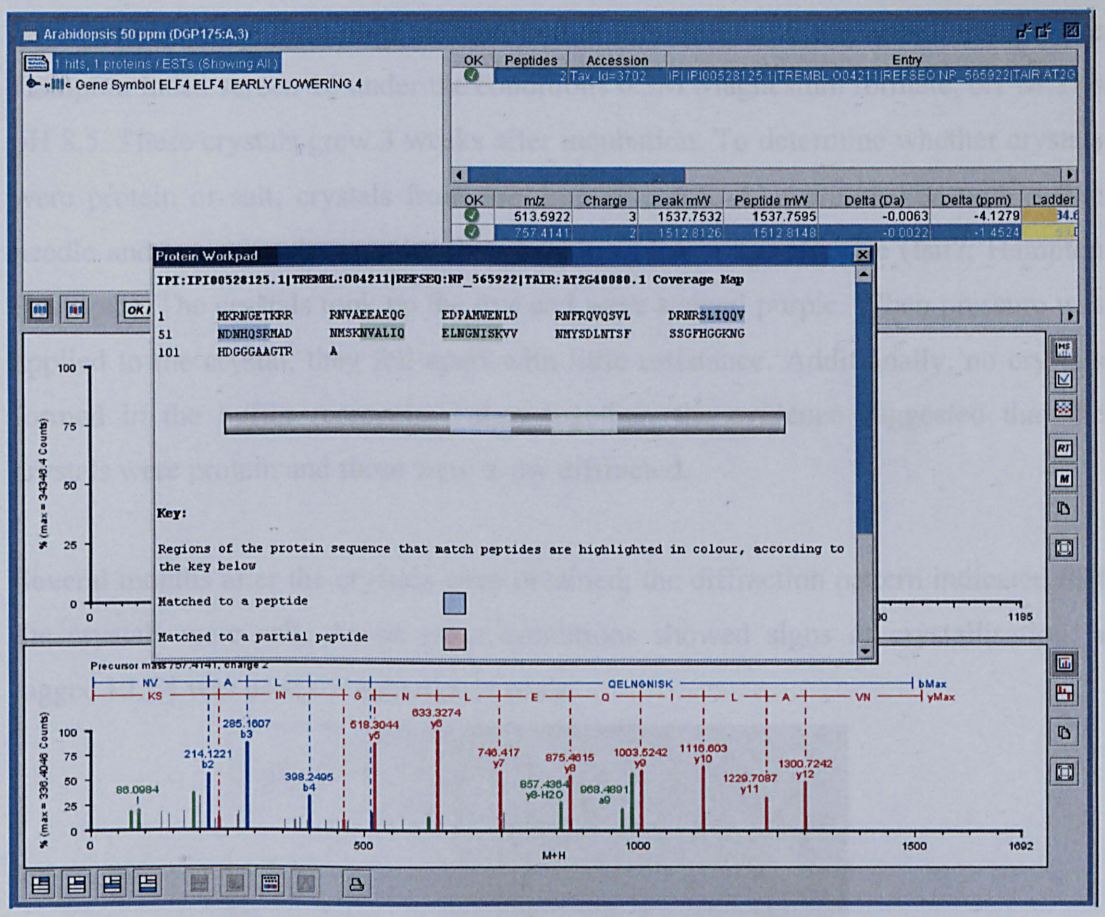


Fig 6.6. On-screen shot of ELF4 confirmation from mass spectrometry analysis. The output file shows the results from mass spectrometry of the band excised from Fig 4.5. The top left identifies the protein as ELF4 with the sequence coverage is shown in the middle of the screen shot. Two peptides are highlighted in purple and green respectively.

6.5. ELF4 Crystallisation trials

The purified ELF4 (Fig 6.5, lane 1) was buffer exchanged according to section 3.2.6.3 and concentrated to 10 mg / ml. Pure ELF4 was very unstable and needed to be kept as cool as possible. In order to maximise the amount of ELF4 that could be used in crystal trials, crystal screens were set-up immediately after sample concentration. A typical purification using the methods a stated in this chapter, yielded 1 mg / ml ELF4 from 4 L of initial culture. Each 96-well screen required 96 µl of ELF4 at 10 mg / ml. Therefore, every crystal screen performed required the growth of 4 L culture.

Initial screens were performed as described in section 3.6.2.5. Crystals formed in the Hampton Index screen 16 under the conditions 0.3M Magnesium formate, 0.1 M Tris pH 8.5. These crystals grew 3 weeks after incubation. To determine whether crystals were protein or salt, crystals from the drop (Fig 6.6, A) were broken with a fine needle and two crystals were incubated with a protein specific dye (Isit?; Hampton Research). The crystals took up the dye and were stained purple. When pressure was applied to the crystal, they fell apart with little resistance. Additionally, no crystals formed in the buffer reservoir. Taken together, the evidence suggested that the crystals were protein and these were x-ray diffracted.

Several months after the crystals were obtained; the diffraction pattern indicated that the crystals were salt. As no other conditions showed signs of crystallisation, a tagged ELF4 was used for subsequent trials.

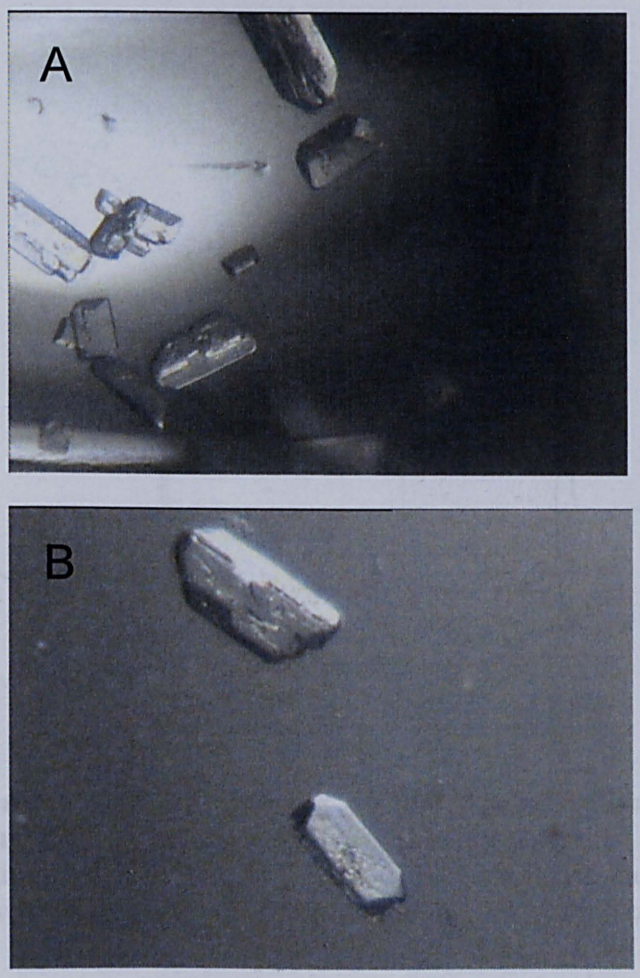


Fig 6.7. Picture of crystals formed in Hampton Index crystal screen 16. (A) Crystals of varying sizes found in the drop (B) Two crystals isolated for x-ray diffraction. The upper crystal is about 0.5 mm in the longest dimension. These crystals formed under the same conditions.

6.6. Purification and crystal trials of S-tagged ELF4

The purification applied was identical to the previous protocol (Section 4.3). After size-exclusion chromatography, ELF4 was successfully digested with thrombin in place of enterokinase. This removed the Trx tag, but left a 15 aa N-terminal S-tag as shown in Fig 6.5. The cleaved Trx was removed by Ni-affinity chromatography and the S-tagged ELF4 came through the flow through (Fig 6.8). The chromatograph indicates a much higher absorbance at 280 nm than Fig 6.8, showing a greater yield of ELF4. The solubility of S-tagged ELF4 was far greater than native ELF4 with almost no precipitation after subsequent buffer exchange into Tris HCl, pH 8.0. Consequently, the yield of S-tagged ELF4 was approximately 20 mg from the cleavage and purification of 50 mg of fusion protein.

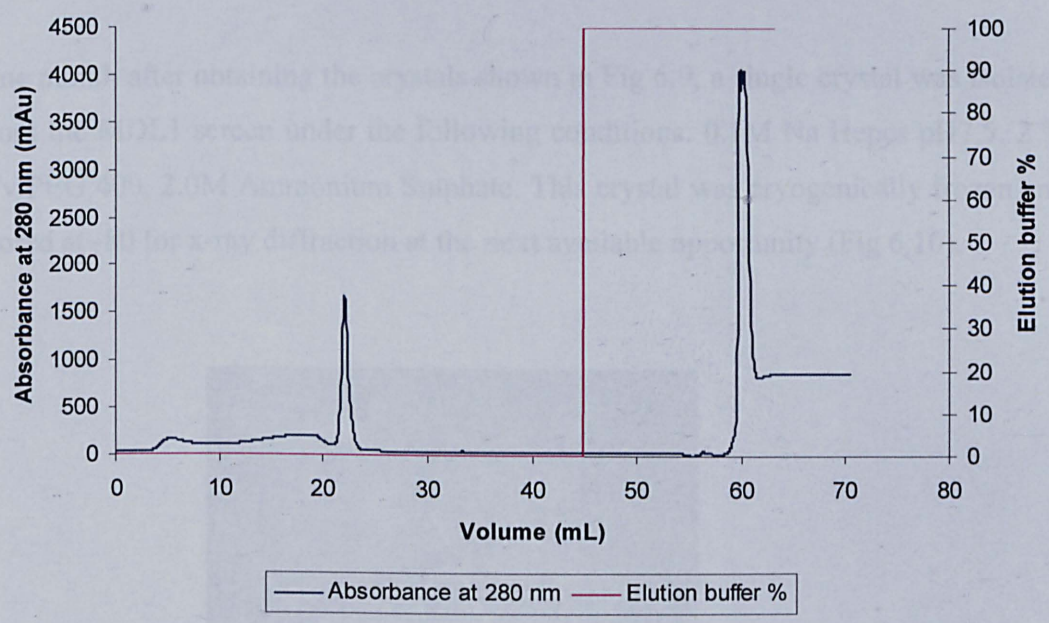


Fig 6.8. Chromatograph showing ELF4 purification by Ni-affinity chromatography after digestion with Thrombin. S-tagged ELF4 is present in the flow-through with a peak elution at 22 ml whilst Trx binds the Ni-cloumn and is eluted at 60 ml with 100 % elution buffer containing 0.5 M imidazole.

A total of 5 crystal screens covering 480 conditions were performed as described in section 3.6.2.5. Crystals formed in 0.2M Magnesium Chloride, 25 % PEG 2K MME after 2 days at 18°C (Fig 6.9). The crystals dissolved under the heat of a lamp that was used to take pictures. We were unable to form crystals by repeating the condition or screening around the condition by altering $MgCl_2$ concentration and pH.

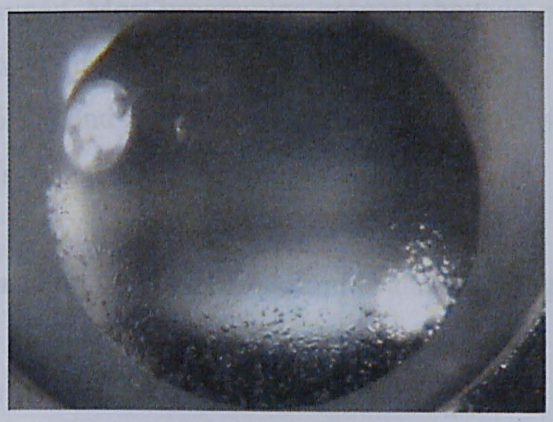


Fig 6.9. Picture of crystals formed in Clear Stratergy screen 3. The crystals appear very fine and are not suitable for x-ray diffraction.

One month after obtaining the crystals shown in Fig 6.9, a single crystal was isolated from the MDL1 screen under the following conditions; 0.1M Na Hepes pH7.5, 2 % v/v PEG 400, 2.0M Ammonium Sulphate. This crystal was cryogenically frozen and stored at -80 for x-ray diffraction at the next available opportunity (Fig 6.10).

Fig 6.10. Picture of a crystal formed in MDL1 screen 30. A single crystal of about 0.3 mm in the longest dimension.

6.7. Circular Dichroism of ELF4

Predication programmes based on mathematical algorithms can potentially highlight structural properties based on sequence similarity to proteins that have had their structure determined. Most programme outputs show no predication for the structure of ELF4. By applying CD to ELF4, we may be able to infer certain secondary structure (Fig 6.11) and may also be able to determine if the protein is suitable for further crystallisation trials.

CD was performed on the pure ELF4 sample and the S-tagged sample as described in section 3.6.2.6. In addition, Near-UV CD was applied to the S-tagged ELF4 due to the amount of protein available.

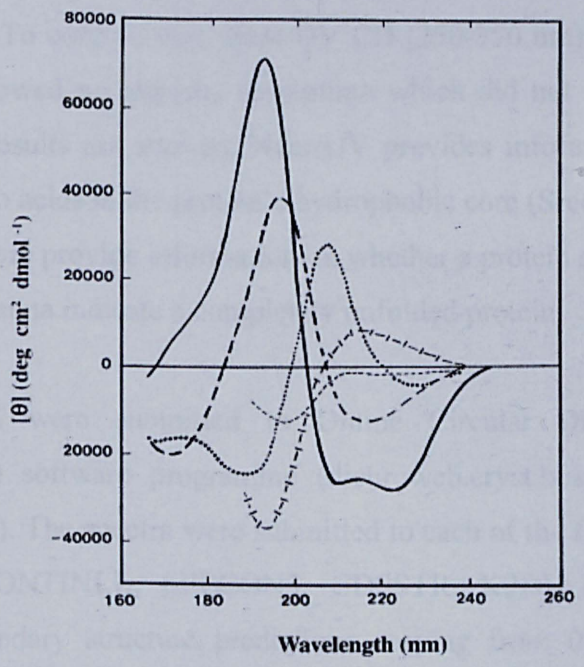


Fig. 6.11. Far-UV CD spectra observed from different types of protein secondary structure. α-helix (solid line), anti-parallel β-sheet (long dashed line), type I β-turn (dotted line), 31-helix or poly (pro) II helix irregular (cross dashed line) or unstructured (short dashed line). (Reproduced from Kelly *et al.*, 2005).

Circular dichroism (CD) spectroscopy is a biophysical technique that measures the difference in absorption of right-handed and left-handed circularly polarised light. Proteins are naturally chiral compounds, and when folded can give rise to distinctive CD spectra (Fig 6.11). By examining the CD spectra in the Far-UV (190-260) range we can infer information regarding secondary structure. In terms of this Chapter, it is useful to provide information on structure as this may have implications regarding crystallisation. A structural robust protein is more likely to crystallise than an unstructured protein.

The CD spectra on ELF4 clearly show the characteristics of a protein with β-sheet content (Fig 6.12 (A)). Each spectra up to 80 °C, exhibits a maxima around 198 nm and a minima around 220 nm. There is also a less pronounced minimum at 209 nm, which would indicate the presence of α-helix. At 95 °C, a protein should be denatured and the spectra clearly show all structure has been lost.

The CD spectra on S-tagged ELF4 indicate the protein is completely unstructured (Fig 6.12(B)). To confirm this, Near-UV CD (250-350 nm) was also performed and the spectra showed no maxima or minima which did not change at for any given temperature (results not shown). Near-UV provides information on the packing of aromatic amino acids in the protein's hydrophobic core (Sreerama and Woody, 2004) and can therefore provide information on whether a protein appears folded or not. No maxima or minima indicate a completely unfolded protein.

These spectra were submitted to Online Circular Dichroism Deconvolution (DICROWEB) software programme (dichroweb.cryst.bbk.ac.uk) (Whittmore and Wallace, 2008). The spectra were submitted to each of the four analysis programmes separately (CONTINLL, SELCON3, CDSSTR, K2D). Each output gave very different secondary structure predictions varying from 0 – 100 % α helix (not shown). More importantly, the normalised root-mean-square deviation (NRMSD) values were poor, in one instance being greater than 1.0. Low NRMSD values (<0.1) indicate that there is less discrepancy between the calculated structure and the crystallographic data in the database. NRMSD values are a measure of the reliability of the prediction and were therefore not used for further analysis.

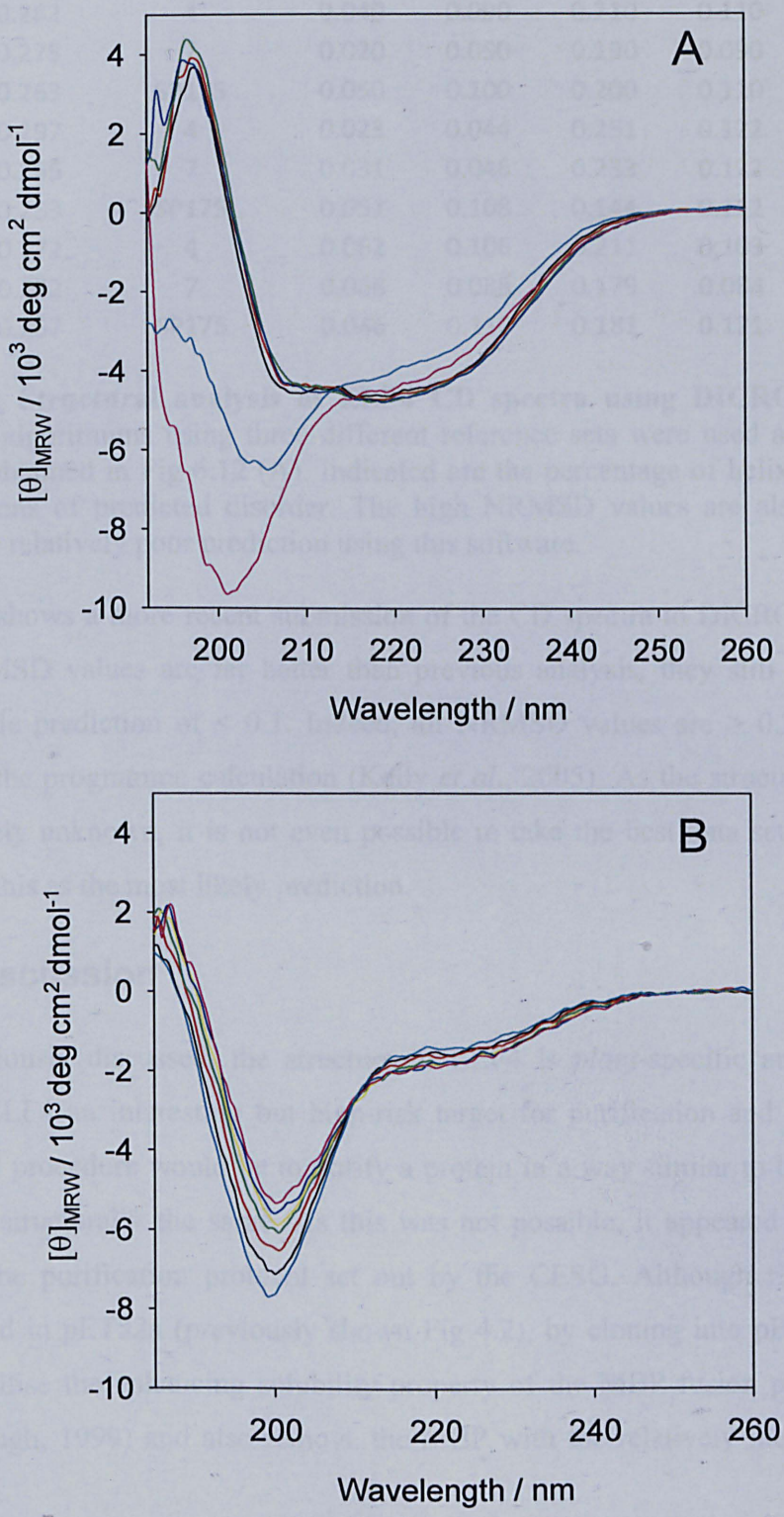


Fig 6.12. Circular dichroism spectra of native ELF4 and S-tagged ELF4. CD spectra of (A) ELF4 and (B) S-tagged-ELF4 at 20 °C (black line), 40 °C (red line), 60 °C (green line), 70 °C (yellow line), 80 °C (blue line) 95 °C (pink line) and 20 °C after re-cooling (turquoise line). $[\theta]_{MRW}$ represents mean molar ellipticity per residue (Section 3.7.2.6).

Algorithm	NRMSD	Reference set	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered
CDSSTR	0.282	4	0.040	0.090	0.210	0.110	0.250	0.300
	0.275	7	0.020	0.050	0.190	0.090	0.190	0.450
	0.263	SP175	0.050	0.100	0.200	0.110	0.130	0.400
SELCON3	0.297	4	0.023	0.044	0.251	0.122	0.233	0.321
	0.265	7	0.031	0.046	0.233	0.122	0.209	0.338
	0.253	SP175	0.051	0.108	0.144	0.112	0.135	0.443
CONTIN	0.272	4	0.062	0.106	0.211	0.103	0.206	0.311
	0.282	7	0.068	0.088	0.179	0.084	0.155	0.426
	0.257	SP175	0.046	0.110	0.181	0.121	0.128	0.415

Table 4. Structural analysis of ELF4 CD spectra using DICROWEB. Three separate algorithums using three different reference sets were used against the CD spectra obtained in Fig 6.12 (A). Indicated are the percentage of helix, strand, turns and regions of predicted disorder. The high NRMSD values are also included to show the relatively poor prediction using this software.

Table 4 shows a more recent submission of the CD spectra to DICROWEB. Whilst the NRMSD values are far better than previous analysis, they still lie outside an acceptable prediction of < 0.1. Indeed, all NRMSD values are > 0.25, suggesting error in the programme calculation (Kelly *et al.*, 2005). As the structure of ELF4 is completely unknown, it is not even possible to take the best data set from Table 4 and use this as the most likely prediction.

6.8. Discussion

As previously discussed, the structure of ELF4 is *plant*-specific and novel. This makes ELF4 an interesting but high-risk target for purification and crystallisation. Standard procedure would be to purify a protein in a way similar to that of proteins that are structurally the same. As this was not possible, it appeared appropriate to follow the purification protocol set out by the CESG. Although ELF4 was well expressed in pET32a (previously shown Fig 4.2), by cloning into pBADM-41+ we could utilise the enhancing solubility property of the MBP fusion partner (Kapust and Waugh, 1999) and also remove the MBP with the relatively inexpensive TEV protease.

Initial pBADM-41+-ELF4 expression was encouraging (Fig 4.5), so a significant amount of time was spent on this construct, trying to remove the MBP after cleavage by TEV protease. None of the purification methods previously discussed resulted in

separation of ELF4 from MBP despite complete cleavage of the fusion protein. It is strange that the cleaved His-MBP did not bind the Ni-column or amylose resin. The reason for this is unclear, but it may be due improper folding of the individual proteins in solution. It is also unclear why MBP and ELF4 did not bind to any of the anion exchange, cation exchange or HIC columns. This problem has occurred with other members of the laboratory working on MBP fusions. As yet, no satisfactory explanation or protocol to overcome this has been deduced. Although there are several examples of proteins co-crystallised with MBP (Smith *et al.*, 2003), it was not pursued in this case due to the difference in Mw. The MBP is almost 4 times as large as ELF4 and would therefore be likely to interfere with the structural properties of ELF4.

The expression of ELF4 in pET32a was also very high (Fig 6.2 B), but the purification procedure was problematic. The first problem was cleaving the Trx tag with enterokinase, which according to the manufactures instructions would cost £100 to cleave 1 mg of fusion protein. By altering the digestion conditions, we were able to cleave far more fusion (Fig 6.3) but it was still expensive. Furthermore, the stability of ELF4 decreased when removed from its fusion partner. Some passenger proteins are unable to form into their native conformations even after they have been made soluble by their fusion partner. These proteins are likely to resist aggregation whilst part of the fusion, but will become aggregated once removed. It is difficult to assess whether the protein is fully folded in its native state until cleavage is performed. In the case of ELF4, it would appear that it is insoluble on its own. Altering the buffers salt concentration, glycerol concentration, pH amongst others can help solubilisation of a target protein; however no screens performed suggested a significant improvement for ELF4. It is also useful to leave out solubilising agents for crystallisation trials, where you wish to increase the insolubility of the protein. Therefore, although yield was low, the best method for purification was to work at 4 °C and to perform fewer purification steps quicker, to minimise aggregation. This was not a consideration for the S-tagged protein which enhanced the solubility of ELF4 significantly, resulting in a much larger yield of protein.

Crystallisation trials failed to provide any protein crystals, with the exception of the crystal shown in Fig 6.10 which is yet to be diffracted. Approximately 480 initial

conditions were screened for the pure and tagged version of ELF4. A recent symposium suggested that if you do not obtain a 'hit' from 2 screens (192 conditions) there is little point in attempting to crystallise the protein under its current conditions (ICCMB 12; personal communication). Indeed, ELF4 is a small (12 kDa) protein that would be highly suitable for studies by Nuclear Magnetic Resonance (NMR). Unfortunately, this would be time consuming and therefore represents a project in its own right.

As nothing was known about the structure, it was difficult to determine whether crystallisation trails were likely to be successful. There are structural prediction programmes which assess the likelihood of crystallisation based on primary sequence. In the case of ELF4, these programmes suggested that the crystallisation trails would be difficult. That said, the process of crystallisation can be quite random and is worth attempting before discounting.

To determine why the trails performed did not provide any protein crystals, we used CD to investigate structural composition. The results indicate that ELF4 contains β-sheet and possibly some α-helix (Fig 4.12 (A)). A protein with a more pronounced band at 209 nm than at 220 nm is indicative of an α + β protein (Pelton and McLean, 2000). As the spectra shows stronger minima at 220 nm and a maxima closer to 200 nm, any α-helix content is likely to be outweighed by β-sheet content. Another interesting observation from the CD spectra is the stability of ELF4. The structure hardly changes at all, even when heated to 80 °C. It requires heating to 95°C to lose its structure and once denatured it does not appear to refold.

S-tagged ELF4 provides some interesting observations. The S-tag was originally identified as a fragment, following cleavage of RNase A by Subtilisin (Richards, 1992). The 15 residue peptide (S-tag) has been utilised as a 'carrier peptide' for purification of fusion proteins. It is small, excessively soluble with little structure and net charge at neutral pH. It therefore makes an ideal candidate for purification procedures as it is not likely to interfere with the structure of the fusion protein (McCormick and Mierendorf). When we look at the CD spectra for S-tagged ELF4, we clearly see no structure (Fig 4.12 (B)). This could be interpreted in one of two ways. Firstly, the S-tag may be interfering with the secondary structure of ELF4. As

previously highlighted, this is probably not the case. More likely, is that the S-tag masks the structural signal of the native protein. This would be particularly evident if ELF4 contains only a small amount of secondary structure combined with a reasonable amount of random coil or unstructured sequence. This conclusion would also highlight the difficulty in obtaining crystals for either native or S-tagged ELF4. A protein that is highly unstructured is more fluid in solution and therefore less likely to form an ordered crystal lattice.

Although we can only postulate at potential structure, the hypothesis of a small amount of very stable structure (mainly β-sheet) combined with a high proportion of random coil may help to explain the function of ELF4. It would not seem unreasonable that ELF4 may act as a hub protein which is promiscuous in binding. Indeed, ELF4 appears to exert an effect on the oscillator, and input and output pathways, suggesting a multi-functioning role. Interestingly, genome-wide studies of hub proteins indicate that deletion of a hub protein tends to be more lethal than deletion of a non-hub protein in regulatory protein networks (Jeong *et al.*, 2001). This may help explain the arrhythmicity observed in all outputs tested in the *ELF4* loss-of-function mutant.

Chapter 7. LUX

7.1. Introduction

Months before the start of this thesis, another CAB::LUC screen identified five mutant alleles of a clock gene called *LUX ARRHYTHMO* (LUX) (Hazen *et al.*, 2005). This gene encodes a Myb-domain transcription factor which is necessary for positive regulation of *LHY* and *CCA1* and is repressed in a similar manner to repression of TOC1 (Reviewed in section 1.4.4). LUX's role as a positive regulator and the fact that it is mandatory for proper clock functioning highlight its potential as a crystallographic target. In addition, as *TOC1* is upregulated in the *lux* mutant (Hazen *et al.*, 2005) it may be working as a protein complex with TOC1 in the control of *LHY* and *CCA1*. Structural analysis may provide insight into what or how LUX binds to protein partners.

In this chapter, the *Lux* gene was cloned into pBADM-41+-BN according to section 3.7.2 and 3.7.3. A similar initial purification to those shown in Fig 4.4 was used and the protocol was adjusted to maximise the amount of LUX protein obtained. The overall aim was to perform crystal screens should sufficient protein be produced.

7.2. Cloning of *LUX*

Initial attempts at cloning LUX were performed using LUX-FOR and LUX-REV (Table 4), and were unsuccessful. Alterations to the PCR conditions, Mg^{2+} concentrations, and amounts of primer / template added were also unsuccessful. The cDNA templates extracted from *Arabidopsis* backgrounds *Ler*, *Ws* and *Col* did not result in amplification of LUX using the aforementioned primers. As the LUX gene does not contain any Introns, genomic DNA was extracted (as described in section 2.7.1 and 2.7.2) from wt *col Arabidopsis*, and used as a template. The PCR produced non-specific products due to the poor annealing of LUX-FOR and LUX-REV to the template. LUX-F25 and LUX-R27 corresponding to the start and end of the *LUX* gene without overhangs were used to amplify the 992 bp *LUX* gene (Fig 7.1 (A). The LUX products shown in Fig 7.1 (A) were used as template with the primers LUX-FOR and LUX-REV in a PCR that resulted in the amplification of LUX with

overhangs beneficial for cloning (Fig 7.1 (B)). LUX was then successfully cloned into pBADM-41+-BN and verified by digestion of the construct with the appropriate restriction enzymes to remove the insert (LUX gene) from the plasmid (Fig 7.1 (C)). The construct was also sequenced in-house to make sure there were no point mutations and that the gene had been cloned in-frame.

Name of Primer	Primer Sequence
LUX-FOR	5' TTTGCG**GAATTC**ATGGGAGAGGAAGTACAA 3'
LUX-REV	5' TTTGCG**CTCGAG**TTA**GTCGAC**ATTCTCATTTGCGCTTCC 3'
LUX-F25	5' ATGGGAGAGGAAGTACAA 3'
LUX-R27	5' TTAATTCTCATTTGCGCTTCCACCTCC 3'

Table 5. Primers used for amplification of LUX. The nucleotides underlined in bold type indicate the restriction enzyme sites. Overhangs are added to the 5' end of the primers to allow directional cloning of the PCR product into pBADM-41+-BN.

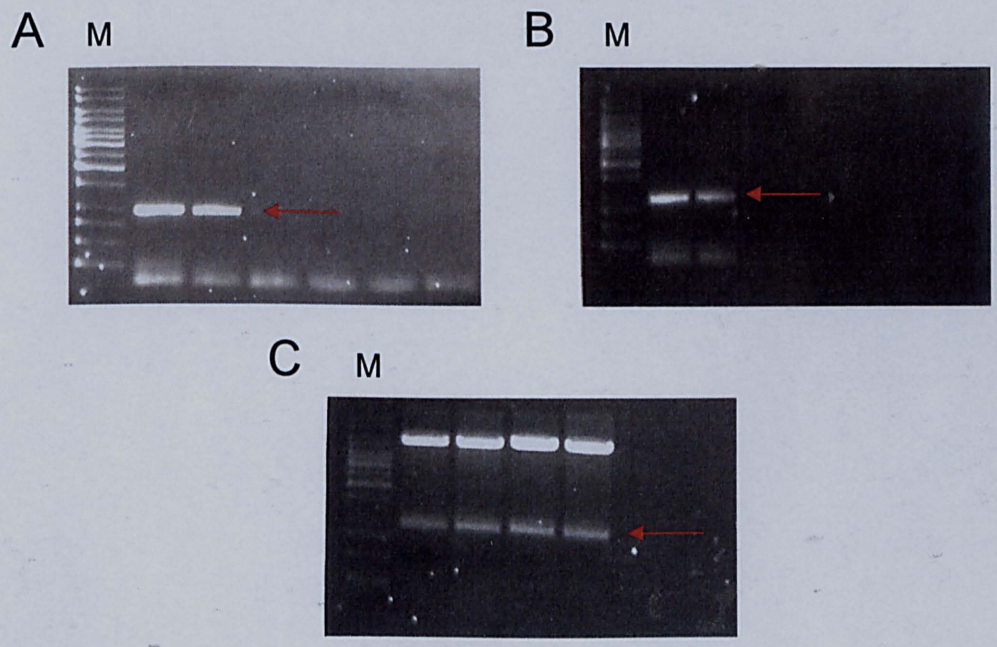


Fig 7.1. 1 % agarose gel showing the cloning of LUX. (A) shows the amplification of LUX using primers LUX-F25 and LUX-R27. (B) shows the amplification of LUX using the product of (A) as template and primers LUX-FOR and LUX-REV. (C) shows the digestion of LUX insert from pBADM-41+-BN (upper band). In each case, M denotes DNA markers and the red arrow highlights the *LUX* gene.

7.3. Expression of pBAD-LUX

Small scale expression was performed according to section 3.7.5. The construct was expressed better than many other pBAD constructs and was purified by Ni-affinity chromatography. The resulting SDS-PAGE (Fig 7.2 (A)) indicates that although the expression of the fusion is reasonable, there are many other products that are co-eluted when Ni-affinity chromatography is performed.

The Chromatograph shows a relatively small peak of elution at around 250 mM imidazole (50% elution buffer) (Fig 7.2 (B). The resulting fraction was subjected to anion exchange (section 3.6.2.1), followed size exclusion chromatography (section 2.6.2.2) to further clean the MBP-LUX fusion. The cleaned fusion was digested, clearly separating MBP from LUX as shown by SDS-PAGE (Fig 7.3). Other digested fractions (Fig 7.3, lanes 2 and 3) show a completely digested fusion that results in only pure MBP. This would suggest instability of the LUX protein after digestion. Although the yield of LUX was poor, it was worth attempting to separate the MBP from the fraction.

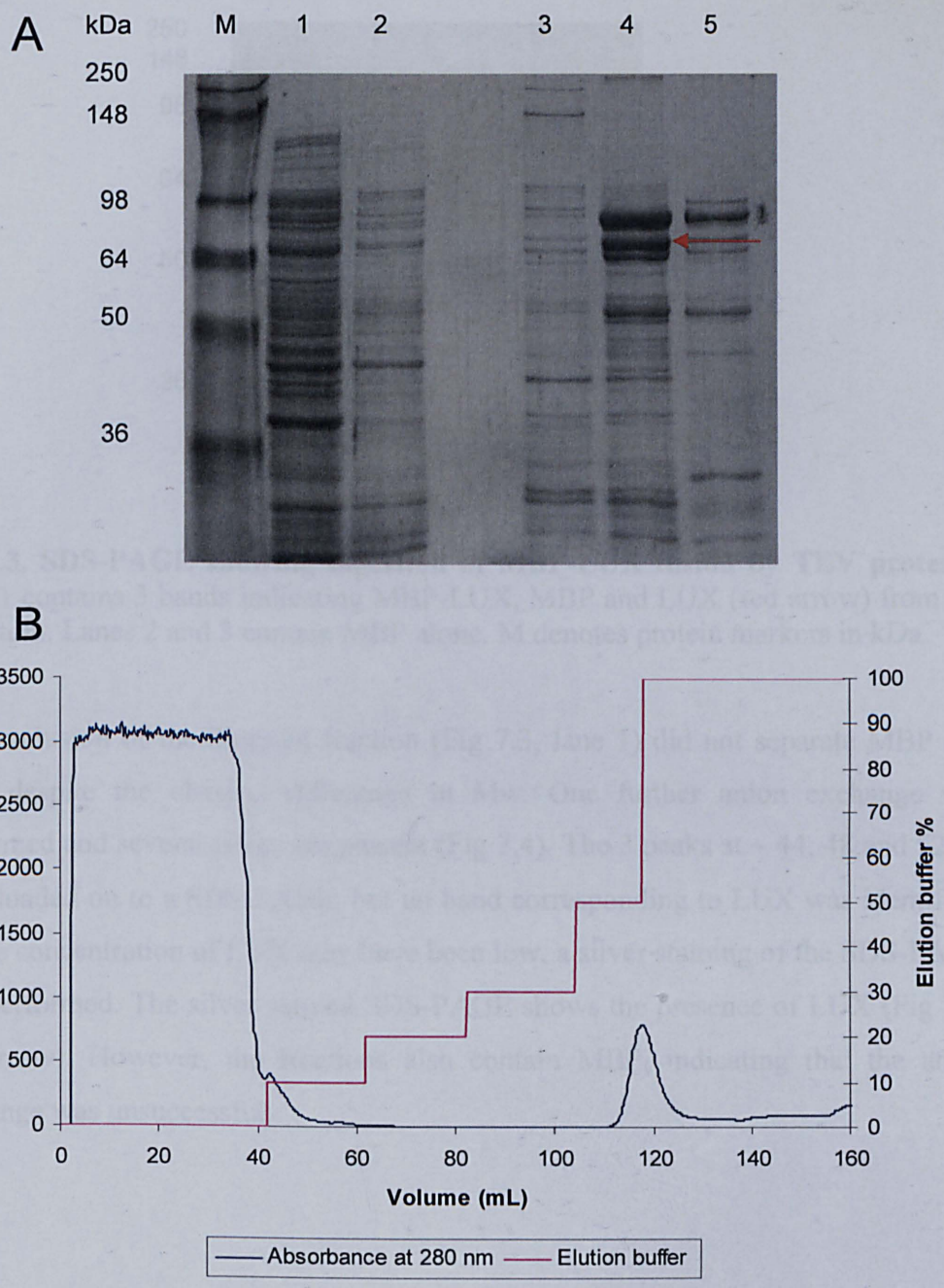


Fig 7.2. Purification of MBP-LUX using Ni-affinity chromatography. (A) SDS-PAGE, lane 1 contains crude fraction, lane 2 contains flow-through and lanes 3-5 contain the LUX elution fractions. M denotes protein markers in kDa (B) Chromatograph showing LUX elution around 120 ml at 50 % elution buffer containing 250 mM imidazole.

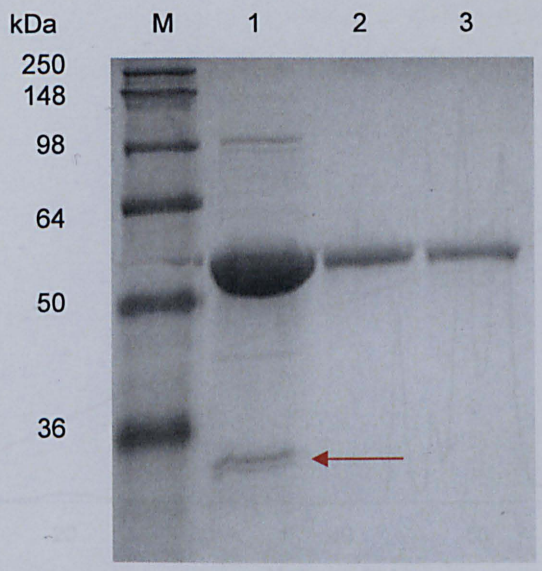


Fig 7.3. SDS-PAGE showing digestion of MBP-LUX fusion by TEV protease. Lane 1 contains 3 bands indicating MBP-LUX, MBP and LUX (red arrow) from top to bottom. Lanes 2 and 3 contain MBP alone. M denotes protein markers in kDa.

Size exclusion of the digested fraction (Fig 7.3, lane 1) did not separate MBP and LUX despite the obvious difference in Mw. One further anion exchange was performed and several peaks are present (Fig 7.4). The 3 peaks at ~ 44, 48 and 52 ml were loaded on to a SDS-PAGE, but no band corresponding to LUX was identified. As the concentration of LUX may have been low, a silver staining of the SDS-PAGE was performed. The silver stained SDS-PAGE shows the presence of LUX (Fig 7.5, red arrow). However, the fractions also contain MBP, indicating that the anion exchange was unsuccessful.

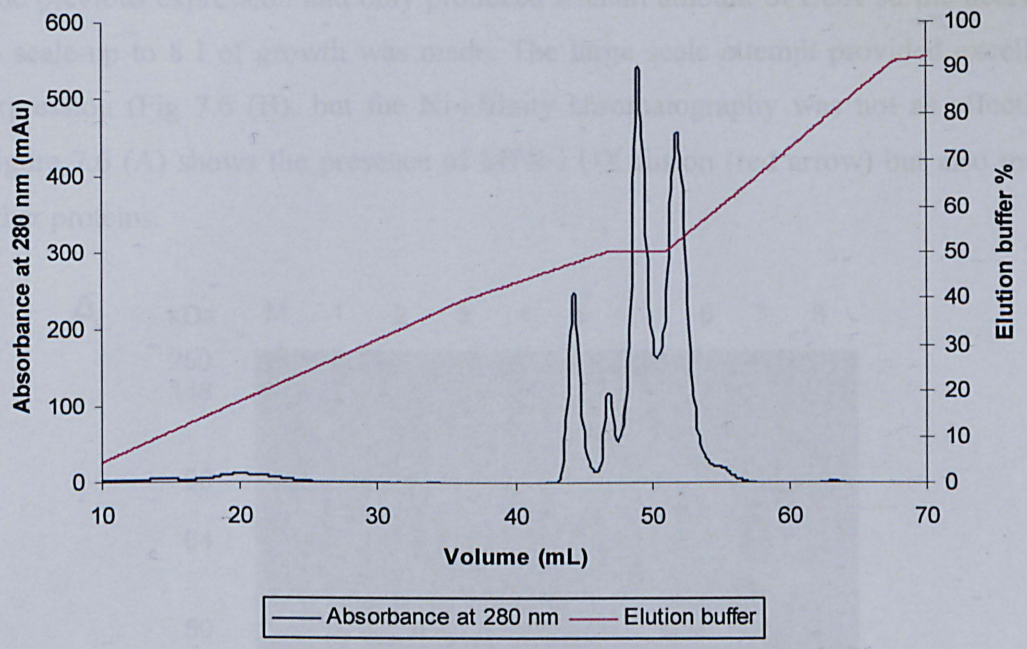


Fig 7.4. Chromatograph showing separation of MBP and LUX by anion exchange. 4 peaks in absorbance over a 12 ml volume (44-56 ml) are shown. The 3 largest peaks were taken for analysis by SDS-PAGE.

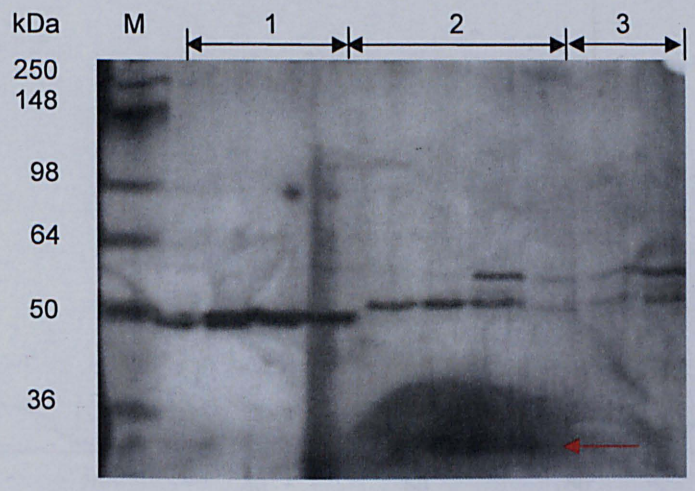


Fig 7.5. Silver-stained SDS-PAGE showing separation of MBP and LUX by anion exchange. Each lane contains 20 µl of 1 ml fractions taken from the anion exchange. Numbers 1-3 highlight the 3 largest peaks (as shown in Fig 5.6). MBP is present in all fractions and LUX is shown in the fraction 2 by a red arrow. The presence of a band at ~57 kDa in 2 and 3 is indicative of contamination. M denotes protein markers in kDa.

The previous expression had only produced a small amount of LUX so the decision to scale-up to 8 l of growth was made. The large scale attempt provided excellent expression (Fig 7.6 (B), but the Ni-affinity chromatography was not as effective. Figure 7.6 (A) shows the presence of MPB-LUX fusion (red arrow) but also many other proteins.

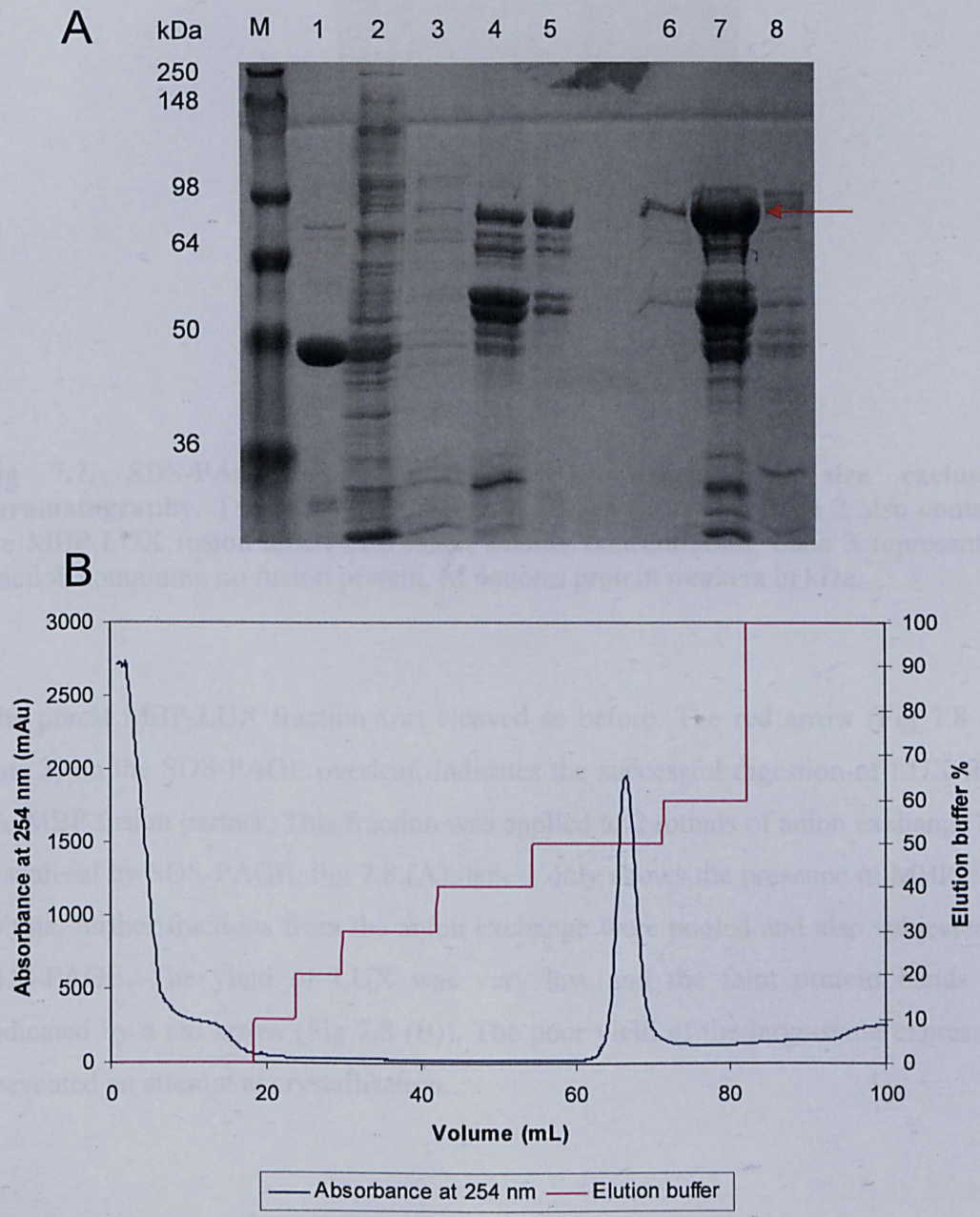


Fig 7.6. Purification of MBP-LUX by Ni-affinity chromatography. (A) SDS-PAGE showing purification of MBP-LUX fusion (red arrow). Lane 2 contains flow-through, lanes 3,4 and 5 contain elution with 25 mM imidazole and lanes 6,7 and 8 contain elution of MBP-LUX with 250 mM imidazole. M denotes protein markers in kDa (B) Chromatograph showing 1 peak with large absorbance at 280 nm. This peak corresponds to the MBP-LUX fusion.

The fraction from Fig 7.6 (A) lane 7, was used in a subsequent anion exchange followed by a size exclusion. The 3 resulting size-exclusion fractions are shown by SDS-PAGE (Fig 7.7).

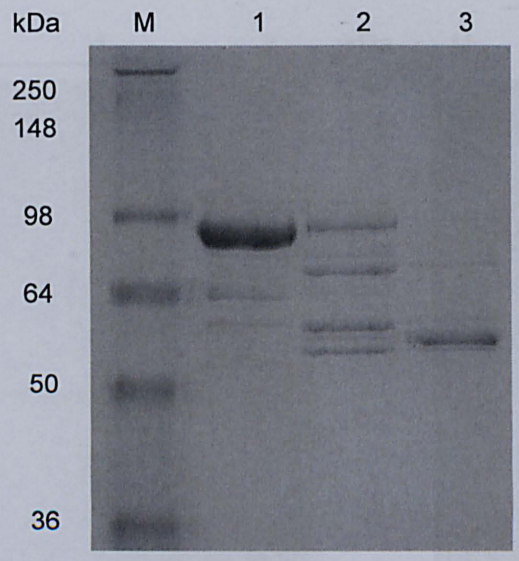


Fig 7.7. SDS-PAGE showing MBP-LUX fusion after size exclusion chromatography. The MBP-LUX fusion is shown in lane 1. Lane 2 also contains the MBP-LUX fusion albeit in a much smaller concentration. Lane 3 represents a fraction containing no fusion protein. M denotes protein markers in kDa.

The purest MBP-LUX fraction was cleaved as before. The red arrow (Fig 7.8 (A) lane 2) on the SDS-PAGE overleaf, indicates the successful digestion of LUX from the MBP fusion partner. This fraction was applied to 2 rounds of anion exchange and visualised by SDS-PAGE. Fig 7.8 (A), lane 1 only shows the presence of MBP. Due to this, further fractions from the anion exchange were pooled and also subjected to SDS-PAGE. The yield of LUX was very low and the faint protein bands are indicated by a red arrow (Fig 7.8 (B)). The poor yield of the large-scale expression prevented an attempt at crystallisation.

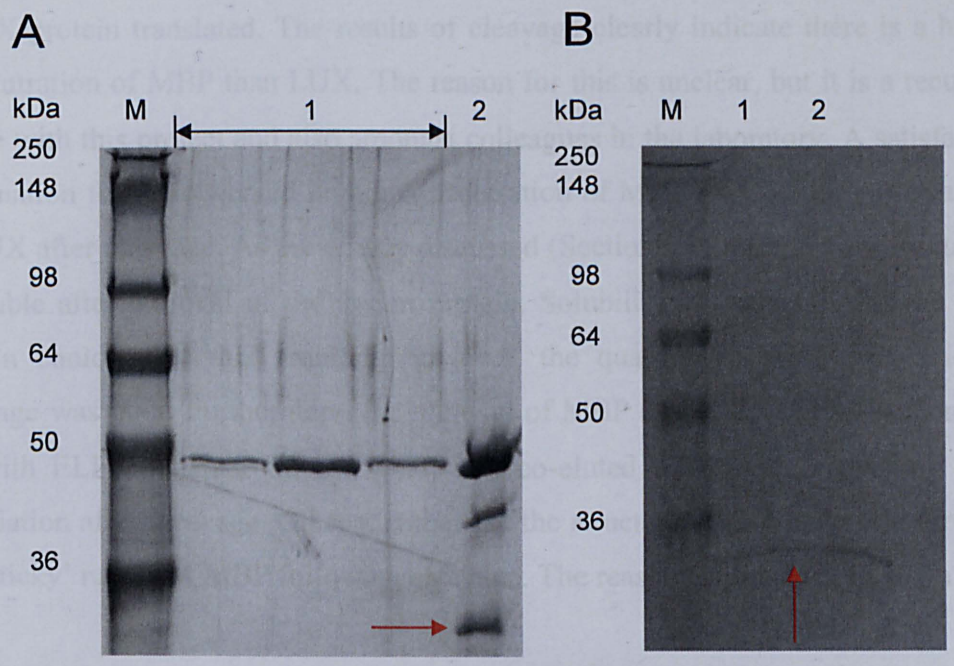


Fig 7.8. SDS-PAGE showing separation of MBP and LUX by 2 sucessive anion exchanges. (A) Lane 1 show the purification of MBP and lane 2 shows the cleavage of MBP-LUX fusion, with LUX highlighted by the red arrow. (B) Lanes 1 and 2 contain LUX (red arrow) after pooling of the 2 anion exchange fractions. M denotes protein markers in kDa.

7.4. Discussion

The initial expression of LUX was not as high as it had previously been for other pBAD contstucts (Fig 4.4), however there was enough expression to attempt purification. Unfortunately, the MBP-LUX fusion was co-eluted with many other proteins during Ni-affinity chromatography. The reason for this could be due to non-specific binding of other proteins to either MBP or LUX. The equilibration buffer (buffer A) and the elution buffer (buffer C) contain 0.5 M NaCl which would usually prevent non-covalent interactions. Another plausible reason for the contamination may be the presence of degraded fusion protein. Any degradation products containing the N-terminal of MBP would also bind the Ni-column and therefore be eluted with the fusion protein.

After cleaning the fusion by anion exchange and size exclusion chromatography, a significant amount was lost. With every purification procedure, the yield of protein

significantly decreased. As the fusion was cleaved, only a small amount of LUX was present (Fig 7.3 and 7.8 (B)). Theoretically, for every MBP translated there should be a LUX protein translated. The results of cleavage clearly indicate there is a higher concentration of MBP than LUX. The reason for this is unclear, but it is a recurring theme with this project and also amongst colleagues in the laboratory. A satisfactory explanation for the observed higher concentration of MBP may be due to instability of LUX after cleavage. As previously discussed (Section 6.8), proteins are frequently insoluble after removal of the fusion protein. Solubilisation screening of the LUX protein could solve this problem; however, the quantity of the fusion prior to cleavage was poor. Furthermore, the removal of MBP after cleavage is problematic. As with ELF4 (Chapter 6), the MBP was co-eluted with LUX suggesting a re-association after cleavage. Other members of the structural group have also reported the 'sticky' nature of MBP following cleavage. The reason for this is not known.

Although, the results presented show the successful purification of LUX (Fig 7.5 and 7.8 (B)), the concentration is so low, nothing can be done with the protein. The initial expression coupled with the quantity of purification steps made the production of LUX impractical. For the production of mg quantities of LUX, another expression system would need to be used. Time constraints did not allow any other vectors to be screened for increase in *LUX* expression, but this would be the logical progression. It would also be worth considering expression of individual domains. As shown with TOC1-PRR and LIP1Δ (Chapters 4 and 8 respectively), this can be successful in increasing the amount of protein produced and also the stability, once cleaved from their fusion partner.

Chapter 8. LIP1

8.1. Introduction

Several months before the start of this project, Laszlo Polgár's group in Budapest isolated a clock mutant that exhibited aberrant entrainment of the clock to light. The work on this mutant was not published at the time and the mutation was given the name B6 (personal communication). Later, B6 was renamed to LIP1 and was shown to represent a new *plant* specific GTPase which is the first GTPase to be implicated in the *Arabidopsis* clock.

LIP1 shows a degree of similarity to known GTPases as it contains a large Rab like domain (Fig 8.1), however this domain contains non-classical peptide inserts as well as the conserved Q_{94} alteration to a H (Kevei *et al.*, 2007). Furthermore, Rab-like GTPases are commonly membrane bound due to their role in vesicle fusion. The LIP1 sequence shows no specific motifs that are associated with lipid anchoring. Due to LIP1s novel negative role in light entrainment, its novel structure and its nature as an enzyme (Kevei *et al.*, 2007), it represents a very interesting protein to study from a crystallographic standpoint.

In addition to attempting full-length expression and purification of LIP1 in pMAL and pBADM-41+, one further construct was used. This construct contained LIP1 minus 108 aa from the C-terminal (LIP1Δ) (Fig 8.1). By removing the relatively disordered proline rich C-terminal, it is expected that the protein would form a more compact ordered structure which would be beneficial for crystallisation

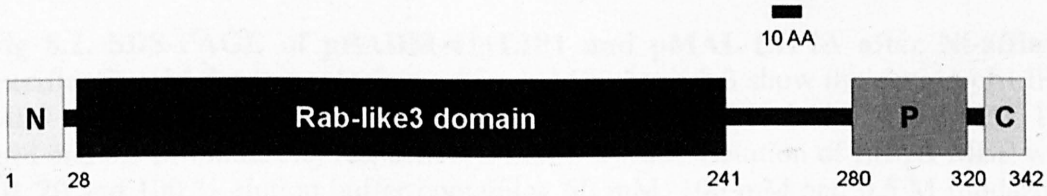


Fig 8.1. Schematic diagram of LIP1 protein structure. Residues 28-241 is a Rab-like domain which contains the GTP binding residues and LIP1 specific features including several peptide inserts and a serine rich domain. P represents a conserved proline rich domain. In this Chapter, a truncated form of LIP1 (LIP1Δ) minus 108 residues from the C-terminus was also investigation. (Reproduced from Kevei *et al.*, 2005)

8.2. Small scale expression trials

Small scale expression trials were performed on pMAL-LIP1, pBADM-41+-LIP1 and pMAL-LIP1Δ as described in section 3.8.1. The pMAL-LIP1 construct failed to express after several attempts, so this was discarded. The pBADM-41+-LIP1 showed a reasonable expression with the isolation of a relatively pure LIP1-MBP fusion after the application of Ni-affinity chromatography Fig 8.3 (A). However, the expression and purification of pMAL-LIP1Δ was greater than either full-length construct as shown in Fig 8.3 (B). The relative expressions of both are highlighted by SDS-PAGE in Fig 8.2.

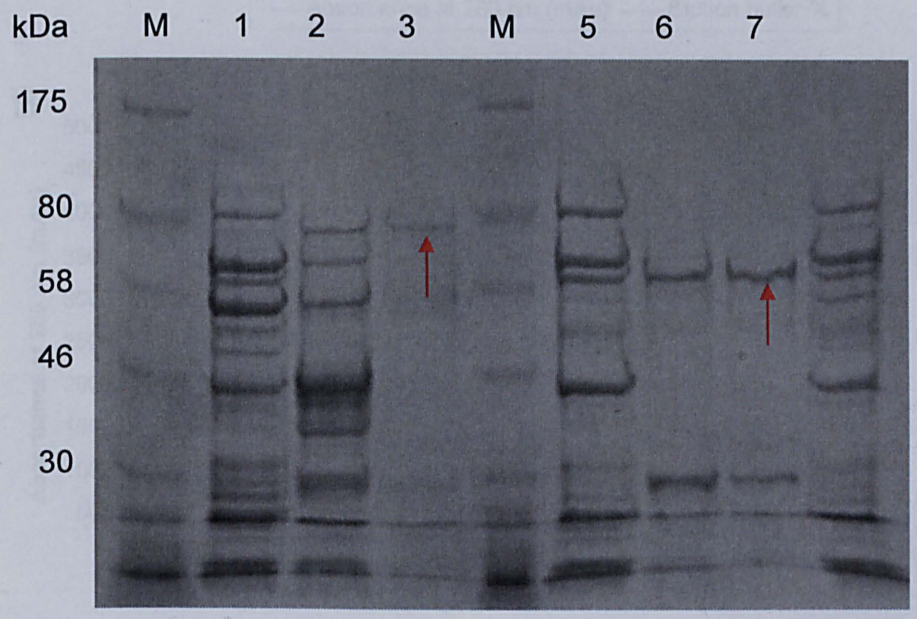


Fig 8.2. SDS-PAGE of pBADM-41+LIP1 and pMAL-LIP1Δ after Ni-affinity purification. M denotes protein markers in kDa. Lane 1-3 show the elution of LIP1-MBP with increasing elution buffer concentration 10, 35 and 100 % (50 mM, 175 mM and 0.5 M imidazole) respectively. Lanes 5-7 show elution of LIP1Δ-MBP with 10, 20 and 100 % elution buffer containing 50 mM, 100 mM and 0.5 M imidazole, respectively. Red arrows indicate the purest fractions of LIP1-MBP and LIP1Δ-MBP.

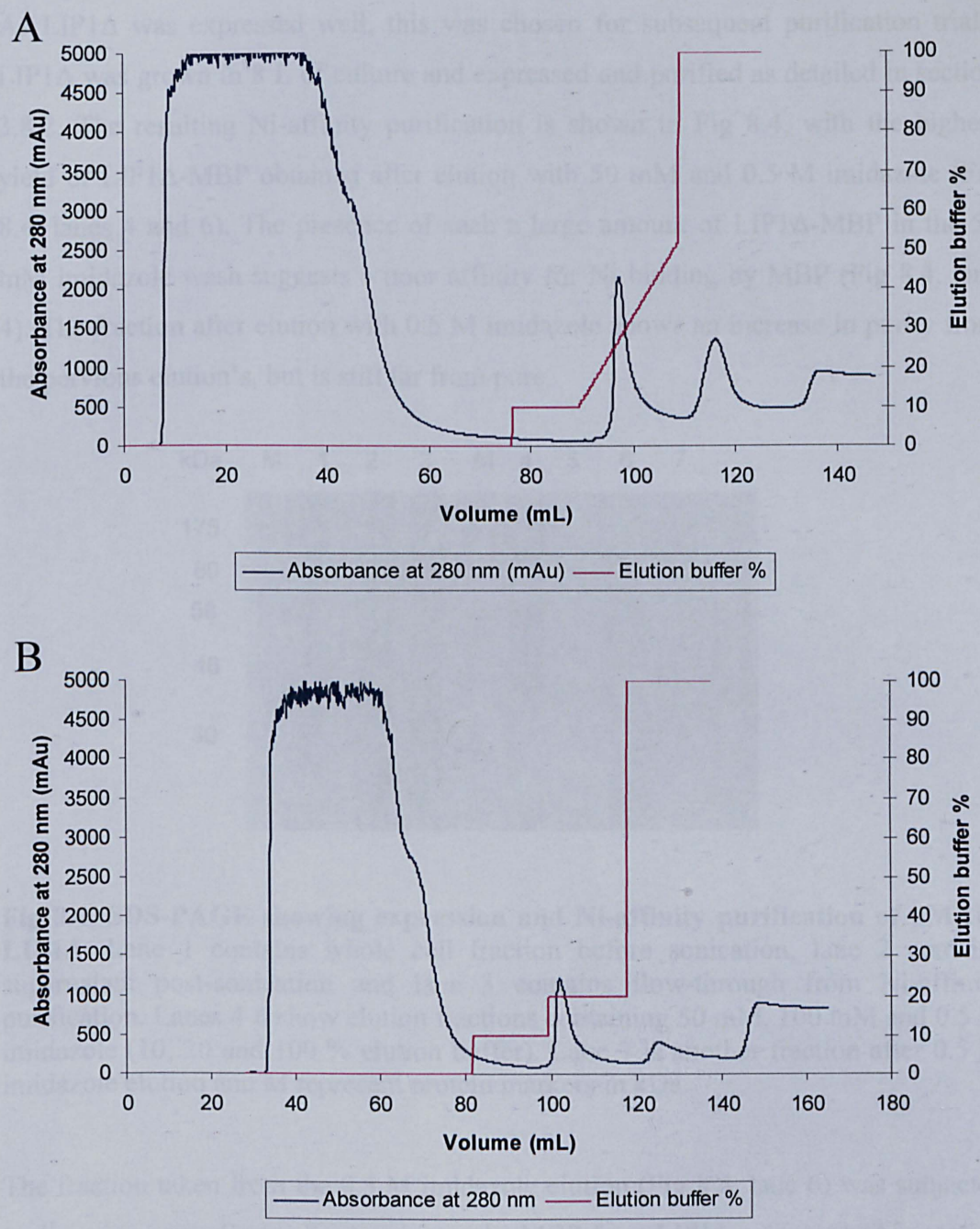


Fig 8.3. Chromatographs showing pBADM-41+LIP1 and pMAL-LIP1Δ after Ni-affinity purification. (A) shows 3 imidazole washes (50 mM, 100 mM and 0.5 M) with the peak at 100 % elution buffer representing the purest fraction of MBP-LIP1. (B) shows the elution of MBP- LIP1Δ in a similar manner, with peaks at 100 ml and 123 ml containing relatively pure fractions.

8.3. Large scale expression of LIP1Δ

As LIP1Δ was expressed well, this was chosen for subsequent purification trials. LIP1Δ was grown in 8 L of culture and expressed and purified as detailed in section 3.8.2. The resulting Ni-affinity purification is shown in Fig 8.4, with the highest yield of LIP1Δ-MBP obtained after elution with 50 mM and 0.5 M imidazole (Fig 8.4, lanes 4 and 6). The presence of such a large amount of LIP1Δ-MBP in the 50 mM imidazole wash suggests a poor affinity for Ni binding by MBP (Fig 8.4, lane 4). The fraction after elution with 0.5 M imidazole shows an increase in purity from the pervious elution's, but is still far from pure.

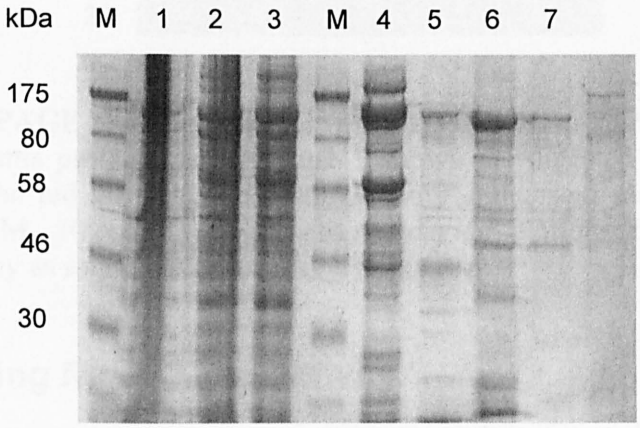


Fig 8.4. SDS-PAGE showing expression and Ni-affinity purification of pMAL-LIP1Δ. Lane 1 contains whole cell fraction before sonication, lane 2 contains supernatant post-sonication and lane 3 contains flow-through from Ni-affinity purification. Lanes 4-6 show elution fractions containing 50 mM, 100 mM and 0.5 M imidazole (10, 20 and 100 % elution buffer). Lane 7 is another fraction after 0.5 M imidazole elution and M represent protein markers in kDa.

The fraction taken from the 0.5 M imidazole elution (Fig 8.4, lane 6) was subjected to digestion using Factor Xa, to remove the MBP from LIP1Δ as described in section 3.8.3. By concentrating the protein in centrifugal concentrators prior to digestion, almost half of the protein precipitated. Due to this, there was simply not enough protein to attempt crystallisation trials after digestion. The digestion with Factor Xa produced only small amounts of LIP1Δ (25.5.kDa) in comparison with the amount of MBP (~ 40 kDa) present (Fig 8.5, lane 1). With time constraints, it was impractical to attempt other purification protocols and so the resulting fraction was used for assaying the GTPase activity of LIP1Δ.

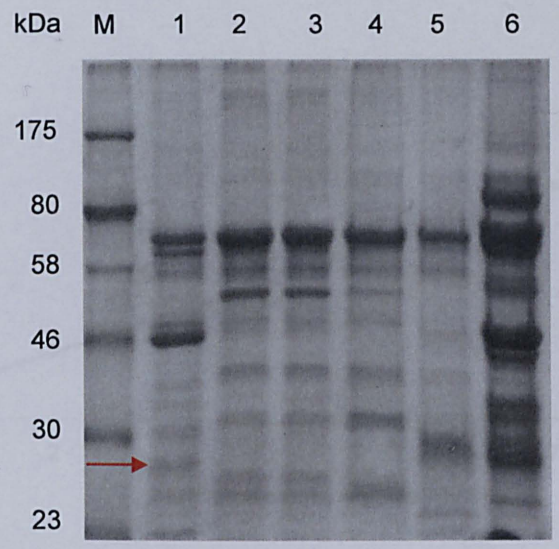


Fig 8.5. SDS-PAGE showing the cleavage of MBP from LIP1Δ-MBP by Factor Xa. M represents protein markers, lane 1 shows the digested sample with LIP1Δ indicated by the red arrow. Lane 2 and 3 show undigested samples and lanes 4-6 contain 50 mM, 100 mM and 0.5 M imidazole elutions from the Ni-affinity chromatography as reference.

8.4. Assaying for GTPase activity

As with TOC1, the trunacated form of LIP1 was easier to express and purify. As LIP1Δ contains the GTPase domain, we could still assay for GTPase activity. The Malachite green assay can accurately measure inorganic phosphate using light spectroscopy at 620-640 nm. The assay detects the inorganic phosphate released by the turnover of GTP to GDP by a nucleoside triosphophatase (LIP1), via the formation of a green solution containing a complex of malachite green, molybdate and free orthophosphate.

A phosphate standard curve was constructed (Fig 8.6) to estimate the amount of phosphate released in the GTPase assays, and thus a rate of reaction

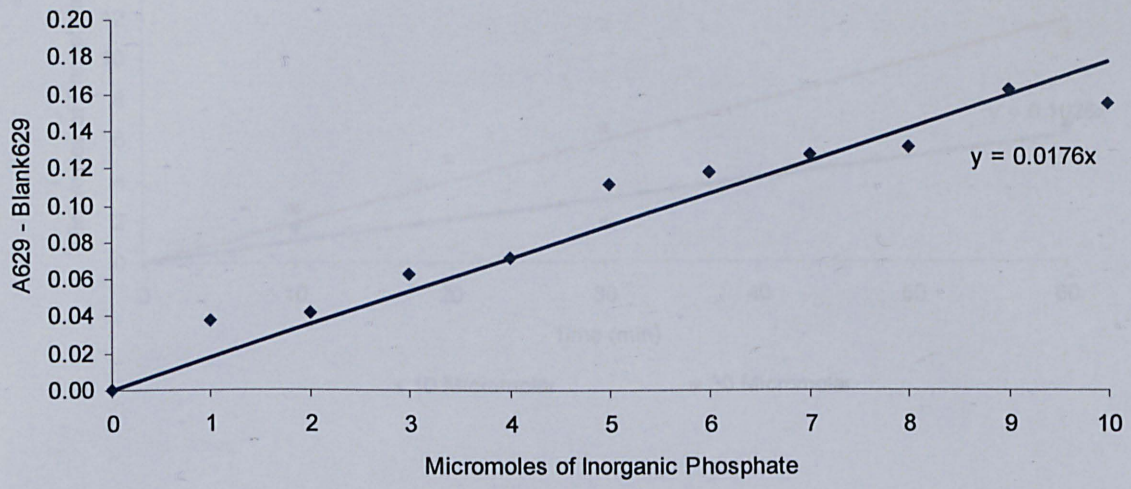


Fig 8.6. A graph showing a phosphate standard curve. Concentrations from 0-10 μM inorganic phosphate (80 μl) were added to 20 μl of malachite green reagent and the reaction was followed at 630 nm for 60 min.

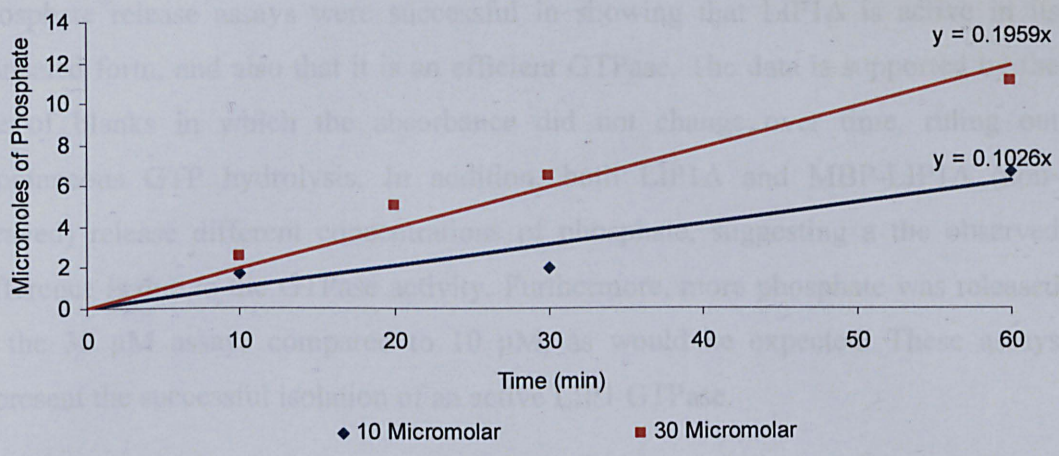


Fig 8.7. A graph showing LIP1Δ GTPase activity. The concentration of inorganic phosphate released was followed by the absorption of light at 630 nm, via the formation of a green complex (malachite green, molybdate and free orthophosphate) over 60 min. Absorbance has been rebased by subtracting blank absorbance values and converted to μM of phosphate using the standard curve (Fig 8.6). The gradient represents the rate of phosphate release in μmol / min.

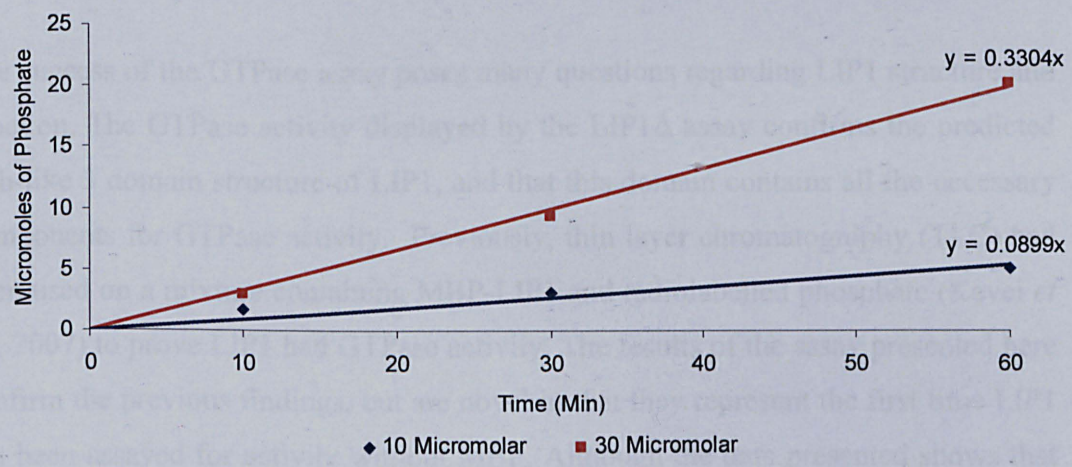


Fig 8.8. A graph showing MBP-LIP1Δ GTPase activity. Absorbance has been rebased by subtracting blank absorbance values and converted to μM of phosphate using the standard curve (Fig 8.6). The gradient represents the rate of phosphate release in μmol / min.

Figures 8.7 and 8.8 clearly show the release of phosphate over a 60 min period. The phosphate release assays were successful in showing that LIP1Δ is active in its truncated form, and also that it is an efficient GTPase. The data is supported by the use of blanks in which the absorbance did not change over time, ruling out spontaneous GTP hydrolysis. In addition, both LIP1Δ and MBP-LIP1Δ (non-cleaved) release different concentrations of phosphate, suggesting a the observed difference is due to the GTPase activity. Furthermore, more phosphate was released in the 30 μM assays compared to 10 μM, as would be expected. These assays represent the successful isolation of an active LIP1 GTPase.

8.5. Discussion

As with TOC1 (Chapter 4), the truncated form of LIP1 proved easier to express and purify than the full length. That said, the expression and purification protocol still requires modification as the quantity and purity of LIP1Δ produced was not enough for crystallisation trials. A similar problem of stability after removal from the fusion partner is one that requires attention.

The success of the GTPase assay poses many questions regarding LIP1 structure and function. The GTPase activity displayed by the LIP1Δ assay confirms the predicted Rab-like 3 domain structure of LIP1, and that this domain contains all the necessary components for GTPase activity. Previously, thin layer chromatography (TLC) had been used on a mixture containing MBP-LIP1 and radiolabelled phosphate (Kevei *et al.*, 2007) to prove LIP1 had GTPase activity. The results of the assay presented here confirm the previous findings, but are novel in that they represent the first time LIP1 has been assayed for activity without MBP. Although the data presented shows that activity is higher in the MBP- LIP1Δ (Fig 8.7 and 8.8), it is highly unlikely that MBP is responsible for any of the phosphate turnover. It is more likely that removal of the MBP results in an adverse effect on LIP1Δ stability and hence a decrease in GTPase activity.

One further interesting point is the rate of GTP hydroysis. The TLC assay was performed overnight and only recorded small levels of hydrolysis (Kevei *et al.*, 2007), whereas the data presented here suggests that LIP1 is a very active GTPase

capable of an average rate of reaction of 6.0 μmol / min / mg. Such a high rate was not expected; as all studied Rab GTPases require accessory factors to turnover GTP. Guanine nucleotide exchange factors (GEFs) facilitate the interchange between GDP and GTP bound forms of Rab allowing the Rab to perform its molecular switch function. Furthermore, GTPase activating proteins (GAPs) assist in the hydrolysation of GTP to GDP. As there were no accessory proteins present, LIP1 represents a highly novel Rab-like GTPase with high GTPase activity independent of accessory factors.

Taken together, these results support the evidence that LIP1 is a truly *plant*-specific novel class of proteins, not only as a GTPase but within the circadian oscillator. Indeed, as LIP1 does not contain the membrane binding motif (important in vesicle formation) present in Rab proteins, which may be consistent with the unusual GTPase activity.

To further characterise the GTPase activity and to determine Michaelis Menton enzyme kinetics, a coupled enzyme would have to be performed. This would result in continuous monitoring of GTPase activity and would allow better comparison with other GTPases. However, optimising the purification and separation of LIP1Δ would be worthwhile if further biochemical characterisation is required.

Chapter 9. General Discussion and Conclusions

9.1. Expression, purification and structural studies of circadian-related proteins

The work presented in this thesis highlights the difficulty in large scale expression and purification of eukaryotic proteins in prokaryotic expression systems. It has been shown that attempts to express TOC1 and ELF3 (Chapter 4) were unsuccessful in a variety of expression vectors, *E. coli* strains and even yeast strains. It appears that utilising fusion proteins Trx and MBP was ineffective at increasing the expression of TOC1 and ELF3. As these proteins are *plant*-specific, it is possible that they are toxic to *E. coli* and yeast which could explain the low levels of expression. Other attributes that may effect the expression of recombinant eukaryotic proteins in *E. coli* include codon bias and post-translational modification. Both proteins were expressed in *E. coli* strains that contained accessory plasmids providing rare tRNA's that would aid the translation of eukaryotic proteins. These did not appear to aid the expression of either TOC1 or ELF3. It is possible that some post-translational modifications are required for the production of fully stable and functional TOC1 and ELF3. These modifications may not be made in *E. coli* and therefore represent another explanation for the difficulty of expression. One important discovery of the work presented is the successful overexpression of truncated forms of TOC1 (Chapter 4) and LIP1 (Chapter 8). Expressing individual domains may represent the best option for large scale protein production. Indeed, a precedent has be set for this approach by the recently published partial structure of the *Drosophila* clock protein PERIOD (PER) (Landskron *et al.*, 2009) Individual domains will reduce the overall size of the protein (and therefore processing required) resulting in independently folded units of structure lacking the flexible inter-domain regions. This is likely to increase the solubility and stability which may be beneficial for crystallisation. That said, as many of these circadian proteins are novel in structure (based on primary sequence) (Chapter 2) it is very difficult to infer where domain boundaries exist. To this end, an approach where random sections of the gene are cloned into expression vectors may be one method that allows regions of the protein to be expressed and purified. The problem with this approach is that it would be labour intensive and there would be no

guarantee of successful crystallisation. In the best case scenario where a crystal structure is obtained, other regions of the protein would still need to be purified in order to build a structure of the whole protein in order to fully investigate structure-function relationships to infer protein function. That said, the information provided in Chapter 2 does highlight regions of predicted structure which may allow the construction of stable, structured truncated forms of the protein.

Many of the clock proteins that were not perused after initial expression trials (Chapter 4) could be purified by expressing individual domains. This includes LUX (Chapter 7) which contains a single Myb-domain for DNA binding and a domain. If time had permitted, this would have been trialled, as full length expression and purification resulted in extremely low quantities of LUX (Chapter 7). The reason for not attempting this on proteins such as LHY and CCA1 are simple. The literature contains many examples of Myb-domain transcription factor structures which are likely to be similar to these proteins. In addition, we already understand that LHY and CCA1 regulate the positive arm of the negative feedback loop by binding to the EE located in the promoters of positive regulators. They are also capable of binding to each others promoters by binding the CBS. A crystal structure of these two would confirm the postulated function, but is unlikely to provide new insights into protein function. The other protein targets presented in this thesis; TOC1, ELF3, ELF4 and GI do represent exciting structural targets, as they have no sequence homology to any other protein and the exact function are unknown (Chapter 2). This makes such a project very high risk, but if successful would provide valuable information regarding the function of the proteins within the regulatory clock network.

The CESG protocol used at the start of the thesis suggested that a fusion using MBP results in the solubilisation of ~80 % of *Arabidopsis* proteins tested (496 out of 632) (Jeon *et al.*, 2004). The results shown in Chapter 4 confirm that MBP is a good solubilisation tag. However, there are problems associated with the removal of MBP. The results in Chapter 7 clearly show the difficulty in separating MBP and LUX after cleavage, and also the high insolubility of LUX without MBP. If a protein precipitates after release from MBP then it is probably improperly folded or prone to aggregation in its native state. This can be overcome by altering conditions that maintain solubility, including varying pH, addition of metal ions and other additives,

and the variation of buffer composition. Although several obvious conditions were screened in this thesis, the time required to solubilise a protein could be a project in its own right and even then there is no guarantee of success. This point is further confirmed by the purification of ELF4 (Chapter 6). Once Trx was removed from ELF4, the solubility of the protein decreased dramatically. Rather than alter buffer composition and procedure (time-consuming) it was easier to keep the purification procedures to a minimum and work at low temperatures, quickly. To avoid this problem, it is possible to screen whether a passenger protein is properly folded and whether it will be soluble after cleavage (Kapust and Waugh, 1999). The authors co-express the MBP fusion with a vector containing the TEV protease (pRK603). TEV protease is produced following the addition of anhydrotetracycline and the cells harvested after time has been allowed for digestion of the MBP fusion. Soluble fractions can then be easily visualised by SDS-PAGE. If domains were to be expressed in *E. coli* as MBP fusions, it would be worthwhile screening in this manner to determine potential purification targets.

The overall purification success rate according to the CESG protocol was 42 % (Jeon *et al.*, 2004). The authors cite weak expression, low solubility and protein precipitation during concentration of the protein, as reasons for the failure of the purification. Indeed, all 3 of these problems have been witnessed throughout this thesis. Proteins that failed to express well were discarded fairly early on and the solubility issues have been discussed. The precipitation problem was evident in TOC1-PRR and ELF4 purification. These proteins precipitated when attempting to concentrate them to at least 10 mg / ml for crystallisation screens. This was not a problem for the S-tagged ELF4, which maintained solubility, but this does not present a solution to the problem. Tagging with a protein / peptide that maintains solubility is more likely to affect the structure of the passenger protein and therefore represents a problem for crystallisation. With few exceptions, fusion tags must be removed before successful crystallisation.

This thesis has clearly shown the difficulty in trying to express these eukaryotic proteins in *E. coli*. Understandably, there have been no reported crystal structures of *Arabidopsis* circadian clock proteins as yet. Many of the proteins mentioned here have been prepared in small quantities sufficient for use in antibody production, but

quantities required for structural studies have been rarely reported. The research carried out for this thesis corroborates this point. Some progress has been made towards structural studies of proteins involved in circadian clocks of other organisms. PERIOD (PER) and KaiC from the *Drosophila* and cyanobacteria central oscillators, respectively, have had their structures solved by x-ray crystallography (Landskron *et al.*, 2009; Pattanayek *et al.*, 2004), highlighting the emerging interest in structural studies of circadian clock proteins. The fact that no structures exist for the *plant* clock is more likely due to aforementioned difficulties of expression and purification of these challenging proteins than lack of interest. Indeed, Chapter 2 provides evidence for circadian proteins being interesting structural targets, and this thesis provides the first attempt to bridge the 'gap' between experimental evidence and structure / function (Bioinformatics) of the circadian proteins.

9.2. Future directions

9.2.1. Leading on from this project

The expression of TOC1-PRR (Chapter 4) requires further attention to minimise precipitation during concentration of the protein. As the expression and purification yield good amounts of protein, this could be further extended to include the cloning and expression of the other PRR domains from PRR9, 7, 5 and 3 for crystallisation screening. Comparisons between the PRR domains could shed light on the function of this complex set of proteins.

Crystal trials for ELF4 (Chapter 6) have proved unsuccessful. This is perhaps unsurprising given the predicted intrinsically unstructured nature of this protein. Indeed, the background literature, Bioinformatics (Chapter 2) and CD data (Chapter 6, Fig 6.12) all indicate that ELF4 is a novel protein, without much ordered secondary structure. It would not be conducive to continue crystal screening in the case of ELF4. Without a crystal hit in the initial ~480 conditions it is unlikely that ELF4 will crystallise. A better strategy would be to perform NMR spectroscopy to the study of the structure. However, the intrinsic disorder would probably pose problems for NMR as well. This would be a PhD project in its own right.

Work on LUX (Chapter 7) showed the problems that can occur during purification even if high levels of soluble expression can be achieved. The solubility is very low after cleavage from MBP and the removal of MBP is problematic. Other expression vectors may be worth attempting, however, it may not be desirable enough as a structural target to justify the painstaking experimental procedure it would require for obtaining crystals. It could be preferential to investigate the function of LUX rather than pursuing the crystal structure.

LIP1 studies have shown that it is a highly active GTPase (Chapter 8). The results presented need to be confirmed using an entirely pure LIP1 sample and also needs to have coupled assays performed to determine Michaelis Menten enzyme kinetics of this GTPase. The protein appears to be difficult to purify, but the structure is likely to be interesting due to its novel properties. A project focusing on assaying for activity and screening for crystallisation could carry on from the work presented in this thesis.

This thesis has provided a protocol for the production of aptamers (Chapter 5). Whilst the target in this case was the clock-associated SRR1 protein, the protocol extends far beyond the clock. In theory, any protein target could be used with this protocol to obtain specific DNA aptamers. As previously discussed, aptamers can be used in a range of biochemical applications and would of use in nearly any field of protein research. The thesis has highlighted the potential use of apatamers in the field of circadian biology and therefore constitutes a novel approach for protein research in this area.

9.2.2. Investigating the function of circadian proteins

During this project, an attempt to express TOC1 and ELF3 *in planta* was made. The genes were cloned into a 35S cassette which could then be cloned into pGREEN29. The 35S cassette is regulated by the Cauliflower Mosaic Virus promoter (CaMV) which results in constitutive expression of the target gene. We could then transfect *Arabidopsis plant*s using *Agrobacterium tumefaciens* containing the pGREEN29 construct for expression trials in *plant*s. The cloning of the 35 S cassettes into

pGREEN29 was not completed due to time constraints. However, this strategy could be applied to the expression of other circadian genes. This would remove the post-translational modification and codon bias problems discussed previously. However, the generation of a transgenic *plant* can take time and the quantity of protein obtained from the expression would probably not be enough for crystallisation. If this is the case, the small amount of expression could be used to raise antibodies against the proteins for functional studies. For proteins that show acceptable expression (Chapter 4) in *E. coli*, existing constructs could be used for antibody production.

If antibodies could be raised against all of the circadian proteins, it would be possible to use them in Western Blots over time courses to find out at what times and relative levels these proteins are accumulating. This would give insight into which proteins may be acting together or as part of complexes. Potential interactions could be further characterised by gel shift assays with combinations of the purified proteins.

Another possibility utilising antibodies includes co-immuno-precipitations over time courses. For example, LHY and CCA1 have been shown to interact *in vitro* (Lu *et al.*, 2009). It should be possible to show such interactions *in vivo* at specific circadian times. This could be extended to include chromatin-immuno-precipitations to determine how proteins such as TOC1, ELF4 and GI up-regulate the expression of *LHY* and *CCA1*. Such experiments may result in the identification of novel binding partners (especially transcription factors) which would be of interest in deciphering control of *LHY* and *CCA1*.

9.3. Conclusions

This project represents an initial effort to develop a protocol for the expression and purification of *plant* circadian proteins. With hindsight, the project was ambitious, as many of these proteins proved difficult to purify in any reasonable quantity. To follow on from the work presented, it would be worth taking a more focused approach. Further projects could either concentrate solely on a structural genomics project, on a number of proteins or a combined biochemical and structural study on individual or discreet families of proteins. These strategies would be more likely to

yield results which may help deduce the structure and function of these novel, *plant*-specific proteins.

In addition to experimental design, this thesis summarises our current understanding of the *Arabidopsis* circadian clock from a genetic and structural (protein) perspective. This is the first instance of such a wide review and therefore represents an attempt to bridge the current gap in the understanding of circadian-associated protein function.

Bibliography

Ahmad, M. and Cashmore, A. R. (1993). HY4 gene of A. thaliana encodes a protein with characteristics of a blue-light photoreceptor. *Nature* **366**(6451): 162-6.

Alabadi, D., Yanovsky, M. J., Mas, P., Harmer, S. L and Kay, S. A. (2002). Critical role for CCA1 and LHY in maintaining circadian rhythmicity in Arabidopsis. *Curr Biol* **12**(9): 757-61.

Alabadi, D., Oyama, T., Yanovsky, M. J., Harmon, F. G., Mas, P and Kay, S. A. (2001). Reciprocal regulation between TOC1 and LHY/CCA1 within the Arabidopsis circadian clock. *Science* **293**(5531): 880-3.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25** : 3389-3402.

Baker, D and Sali, A. (2001). Protein structure prediction and structural genomics. *Science* **294**: 93-96.

Baskin, J. M and Baskin, C. C. (1976). Effect of photoperiod on germination of *cyperus inflexus* seeds. *Bot Gaz* **137**: 269-273.

Bennett-Lovesey, R. M., Herbert, A. D., Sternberg, M. J. E and Kelley, L. A. (2008). Exploring the extremes of sequence / structure space with ensemble fold recognition in the programme Phyre. *Proteins* **70**: 611-625.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N and Bourne, P. E. (1999).The protein data bank. *Nucleic Acids Res* **28**: 235-242.

Bren, A and Eisenbach, M. (1998). The N terminus of the flagellar switch protein, FliM, is the binding domain for the chemotactic response regulator, CheY. *J Mol Biol* **278**: 507-514.

Briggs, W. R. and Christie, J. M. (2002). Phototropins 1 and 2: versatile *plant* blue-light receptors. *Trends Plant Sci* **7**(5): 204-10.

Bryant, T. R. (1972). Gas exchange in dry seeds: circadian rhythmicity in the absence of DNA replication, transcription, and translation. *Science* **178**(61): 634-6.

Cardozo, T and Pagano, M. (2004). The SCF ubiquitin ligase: insights into a molecular machine. *Nat Rev. Mol. Cell. Biol* **5**: 739-751.

Cashmore, A. R., Jarillo, J. A., Wu, Y. J and Liu, D. (1999). Cryptochromes: blue light receptors for *plant*s and animals. *Science* **284**(5415): 760-5.

Christie, J. M. (2007). Phototropin blue-light receptors. *Annu Rev Plant Biol* **58**: 21-45.

Cho, H. Y., Tseng, T. S., Kaiserli, E., Sullivan, S., Christie, J. M and Briggs W. R. (2007). Physiological roles of the light, oxygen, or voltage domains of phototropin1 and phototropin2 in *Arabidopsis. Plant Physiol* **143**:517-529.

Cole, C., Barber, J. D and Barton, G. J. (2008). The JPred 3 secondary structure prediction server. *Nucleic Acids Res* **36** (Web server issue): W197-W201.

Covington, M. F., Panda, S., Liu, X. L., Strayer, C. A., Wagner, D. R and Kay, S. A. (2001). ELF3 modulates resetting of the circadian clock in Arabidopsis. *Plant Cell* **13**(6): 1305-15.

Craig, K. L. and Tyers, M. (1999). The F-box: a new motif for ubiquitin dependent proteolysis in cell cycle regulation and signal transduction. *Prog Biophys Mol Biol* **72**(3): 299-328.

Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y and McPherson, A. (1994). *Acta Cryst* **D50**: 414-423.

Daniel, X., Sugano, S and 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**(9): 3292-7.

David, K. M., Armbruster, U., Tama, N and Putterill, J. (2006). Arabidopsis GIGANTEA protein is post-transcriptionally regulated by light and dark. *FEBS Lett* **580**(5): 1193-7.

de Mairan, J (1729). Observation botanique. *Hist Symp Quant Biol* **25**: 159-184.

Demarsy, E and Fankhauser, C. (2009). Higher *plant*s use LOV to perceive blue light. *Curr Opin in Plant Biology* **12**: 69-74.

Densmore, R. V. (1997). Effect of daylength on germination of seeds collected in Alaska. *Am J Bot* **84**: 274-278.

Deutscher, M. (1990). Guide to protein purification. *Methods in Enzymology* **182**: London. Academic press.

Devlin, P. F and Kay, S. A. (2001). Circadian photoperception. *Annu Rev Physiol* **63**: 677-94.

Dodd, A. N., Salathia, N., Hall, A., Kevei, E., Toth, R., Nagy, F., Hibberd, J. M., Millar, A. J and Webb, A. A. (2005). *Plant* circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science* **309**(5734): 630-3.

Dowson-Day, M. J. and Millar, A. J. (1999). Circadian dysfunction causes aberrant hypocotyl elongation patterns in Arabidopsis. *Plant J* **17**(1): 63-71.

Doyle, M. R., Davis, S. J., Bastow, R. M., McWatters, H. G., Kozma-Bognar, L., Nagy, F., Millar, A. J and Amasino, R. M. (2002).The ELF4 gene controls circadian rhythms and flowering time in Arabidopsis thaliana. *Nature* **419**(6902): 74-7.

Du, H., Zhang, L., Liu, L., Tang, X., Yang, W.J., Wu, Y., Huang, Y and Tang, Y. (2008). Biochemical and molecular characterization of *Plant* MYB Transcription factor family. *Biochemisty* (Moscow) **74** (1): 1-11.

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

Dyer, C. M and Dahlquist, F. W. (2006). Switched or not? : the structure of unphosphorylated CheY bound to the N-terminus of FliM. *J. Bacteriol* **188**: 7354-7363.

Edwards, K. D., Anderson, P. E., Hall, A., Salathia, N. S., Locke, J. C. W., Lynn, J. R., Straume., M., Smith, J. Q and Millar, A. J. (2006). FLOWERING LOCUS C Mediates Natural Variation in High-Temperature Response of the Arabidopsis circadian clock. *Plant cell* **18**(3): 639-50.

Ellington, A. D and Szostak, J. W. (1990). In vitro selection of RNA molecules that bind specific ligands. *Nature* **346** (6287):818-822.

Englemann, W. and Johnson, A. (1998). Rhythms in organ movement: In Biological rhythms and photoperiodism in *plant*s. (Lumsden, P. J. and Millar, A. J. eds): 35-50. Oxford. Bios Scientific publishers.

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

Fowler, S., Lee, K., Onouchi, H., Samach, A., Richardson, K., Morris, B., Coupland, G and Putterill, J. (1999). GIGANTEA: a circadian clock-controlled gene that regulates photoperiodic flowering in Arabidopsis and encodes a protein with several possible membrane-spanning domains. *EMBO J* **18**(17): 4679-88.

Fujiwara, S., Wang, L., Han, L., Suh, S. S., Salome, P. A., McClung, C. R and Somers, D. E. (2008). Post-translational regulation of the Arabidopsis circadian clock through selective proteolysis and phosphorylation of pseudo-response regulator proteins. *J Biol Chem* **283**(34): 23073-83.

Gardner, M. J., Hubbard, K. E., Hotta, C. T., Dodd, A. N and Webb, A. A. (2006). How *plant*s tell the time. *Biochem J* **397**(1): 15-24.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D and Bairoch, A. (2005). Protein identification and Analysis Tools on the ExPASy server. John M Walker (Ed). The proteomics protocols handbook, Humana Press : 571-607.

Guo, H., Yang, T., Mockler, C and Lin, C. (1998). Regulation of flowering time by Arabidopsis photoreceptors. *Science* **279**: 1360-1363.

Geiger, A., Burgstaller, P., Von der Eltz, H., Roeder, A and Famulok, M. (1996). RNA Aptamers that bind L-Arginine with Sub-Micromolar Dissociation Constants and High Enantioselectivity. *Nucleic Acids Res* **24**(6):1029-1036.

Gietz, R.D and Woods, R. A. (2002). TRANSFORMATION OF YEAST BY THE Liac/SS CARRIER DNA/PEG METHOD. *Methods in Enzymology* **350:** 87-96.

Gould, P. D., Locke, J. C., Larue, C., Southern, M. M., Davis, S. J., Hanano, S., Moyle, R., Milich, R., Putterill, J., Millar, A. J and Hall, A. (2006). The molecular basis of temperature compensation in the Arabidopsis circadian clock. *Plant Cell* **18**(5): 1177-87.

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 and Millar, A. J. (2003).The TIME FOR COFFEE gene maintains the amplitude and timing of Arabidopsis circadian clocks. *Plant Cell* **15**(11): 2719-29.

Han, L., Mason, M., Risseeuw, e. p., Crosby, W. L and Somers, D. E. (2004). Fromation of an SCF(ZTL) complex is required for proper regulation of circadian timing. *Plant J* **40**: 291-301.

Harmer, S. L. (2009). The circadian system in higher *plant*s. *Annu Rev Plant Biol* **60**: 357-77.

Harmer, S. L., Panda, S and Kay, S. A. (2001). Molecular bases of circadian rhythms. *Annu Rev Cell Dev Biol* **17**: 215-53.

Harmer, S. L., Hogenesch, J. B., Straume, M., Chang, H. S., Han, B., Zhu, T., Wang, X., Kreps, J. A and Kay, S. A. (2000). Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. *Science* **290**(5499): 2110-3.

Harmon, F., Imaizumi, T and Gray, W. M. (2008). CUL1 regulates TOC1 protein stability in the Arabidopsis circadian clock. *Plant J* **55**(4): 568-79.

Hazen, S. P., Schultz, T. F., Pruneda-Paz, J. L., Borevitz, J. O., Ecker, J R and Kay, S. A. (2005). LUX ARRHYTHMO encodes a Myb domain protein essential for circadian rhythms. *Proc Natl Acad Sci U S A* **102**(29): 10387-92.

Hicks, K. A., Albertson, T. M and Wagner, D. R. (2001).EARLY FLOWERING3 encodes a novel protein that regulates circadian clock function and flowering in Arabidopsis. *Plant Cell* **13**(6): 1281-92

Hicks, K. A., Millar, A. J., Carre, I. A., Somers, D. E., Straume, M., Meeks-Wagner, D. R and Kay, S. A. (1996). Conditional circadian dysfunction of the Arabidopsis early-flowering 3 mutant. *Science* **274**(5288): 790-2.

Ho, M. S., Ou, C., Chan, Y., Chien, C-T and Pi, H. (2008). The utility F-box for protein destruction. *Cell Mol Life Sci* **65**: 1977-2000.

Huang, T., Bohlenius, H., Eriksson, S., Parcy, F and Nilsson, O. (2005). The mRNA of the Arabidopsis gene FT moves from leaf to shoot apex and induces flowering. *Science* **309**(5741): 1694-6.

Huq, E., Tepperman, J. M and Quail, P. H. (2000). GIGANTEA is a nuclear protein involved in phytochrome signalling in Arabidopsis. *Proc Natl Acad Sci USA* **97**: 9789-9794.

Imamura, A., Hanaki, N., Umeda, H., Nakamura, A., Suzuki, T., Ueguchi, C and Mizuno, T. (1998). Response regulators implicated in His-to-Asp phototransfer signalling in *Arabidopsis*. *PNAS USA* **95**: 2691-2696.

James, A. B., Monreal, J. A., Nimmo, G. A., Kelly, C. L., Herzyk, P., Jenkins, G. I and Nimmo, H. G. (2008). The circadian clock in Arabidopsis roots is a simplified slave version of the clock in shoots. *Science* **322**(5909): 1832-5.

Jarillo, J. A., Capel, J., Tang, R. H., Yang, H. Q., Alonso, J. M., Ecker, J. R and Cashmore, A. R. (2001). An Arabidopsis circadian clock component interacts with both CRY1 and phyB. *Nature* **410**(6827): 487-90.

Jancarik and Kim, J. (1991). *J Appl Cryst* **24**: 409-411.

Jenal, U and Galperin, M. Y. (2009). Single-domain response regulators: molecular switches with emerging roles in cell organisation and dynamics. *Curr Opin Microbiol* **12**(2): 152-160.

Jenison, R. D., Gill, S. C., Pardi, A. Polisky, B. (1994). High-resolution discrimination by RNA. *Science* **263** (5152):1425-1429.

Jeon, W. B., Aceti, D. J., Bingman C. A., Vojtik, F. C., Olson, A.C., Ellefson, J. M., McCombs, J. E., Sreenath, H. K., Blommel, P. G., Seder, K. D., Burns, B.T., Geetha, H. V., Harms, A. C., Sabat, G., Sussman, M. R., Fox, B.G and Phillips G. N. Jr. (2005). High-throughput purification and quality assurance of *Arabidopsis thaliana* proteins for eukaryotic structural genomics. *J Struct Funct Genomics* **6**(2-3): 143-7.

Jeong, H., Mason, S. P., Barabási, A. L and Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature* **411**: 41-42.

Johnson, C. H. (1999). Forty years of PRCs--what have we learned? *Chronobiol Int* **16**(6): 711-43.

Johnson, C. H., Knight, M. R., Kondo, T., Masson, P., Sedbrook, J., Haley, A and Trewavas, A. (1995). Circadian oscillations of cytosolic and chloroplastic free calcium in *plants*. *Science* **269**(5232): 1863-5.

Kapust, R. B and Waugh, D. S. (1999). *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci* **8**: 1668-1674.

Kagawa, T., Kasahara, M., Abe, T., Yoshida, S and Wada, M. (2004). Function analysis of phototropin2 using fern mutants deficient in blue light induced chloroplast avoidance movement. *Plant Cell Physiol* **45**:416-426.

Kelley, L. A and Sternberg, M. J. E. (2009). Protein structure prediction on the web: a case study using the Phyre server. *Nature Protocols.* **4**: 363 – 371.

Kelly, S. M., Jess, T. J and Price, N. C. (2005). How to study proteins by circular dichroism. *Biochim. Biophys. Acta* **1751**: 119-139.

Kern, D., Volkman, B. F., Luginbuhl, P., Nohaile, M. J., Kustu, S and Wemmer, D. E. (1999). Strucutre of a transiently phosphorylated switch in bacterial signal transduction. *Nature* **402**: 894-898.

Kevei, E., Gyula, P., Feher, B., Toth, R., Viczian, A., Kircher, S., Rea, D., Dorjgotov, D., Schafer, E., Millar, A. J., Kozma-Bognar, L and Nagy, F. (2007). Arabidopsis thaliana circadian clock is regulated by the small GTPase LIP1. *Curr Biol* **17**(17): 1456-64.

Khanna, R., Kikis, E. A and Quail, P. H. (2003). EARLY FLOWERING 4 functions in phytochrome B-regulated seedling de-etiolation. *Plant physiol* **133**: 1530-1538.

Kiba, T., Henriques, R., Sakakibara, H and Chua, N. H. (2007). Targeted degradation of PSEUDO-RESPONSE REGULATOR5 by an SCFZTL complex regulates clock function and photomorphogenesis in Arabidopsis thaliana. *Plant Cell* **19**(8): 2516-30.

Kikis, E. A., Khanna, R and 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**(2): 300-13.

Kim, W. Y., Fujiwara, S., Suh, S. S., Kim, J., Kim, Y., Han, L., David, K., Putterill, J., Nam, H. G and Somers, D. E. (2007). ZEITLUPE is a circadian photoreceptor stabilized by GIGANTEA in blue light. *Nature* **449**(7160): 356-60.

Kiyosue, T. and Wada, M. (2000). LKP1 (LOV kelch protein 1): a factor involved in the regulation of flowering time in arabidopsis. *Plant J* **23**(6): 807-15.

Koornneef, M., Hanhart, C. J and van der Veen, J. H. (1991). A genetic and physiological analysis of late flowering mutants in Arabidopsis thaliana. *Mol Gen Genet* **229**: 56-66.

Kozma-Bognar, L. and Kaldi, K. (2008). Synchronization of the fungal and the *plant* circadian clock by light. *Chembiochem* **9**(16): 2565-73.

Laemmli, U. K. (1970).Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**(5259): 680-5.

Landskron, J., Chen, K. F., Wolf, E and Stanewsky, R. (2009). A role for the PERIOD:PERIOD Homodimer in the Drosophila circadian clock. *PLoS Biol* **7**(4): e1000003.

Lee, S. Y., Cho, H. S., Pelton, J. G., Yan, D., Berry, E. A and Wemmer, D. E. (2001). Crystal structure of activated CheY. Comparison with other activated receiver domains. *J Biol Chem* **276** (19): 16425-16431.

Lin, C., Yang, H., Guo, H., Mockler, T., Chen, J and Cashmore, A. J. (1998). Enhancement of blue-light sensitivity of Arabidopsis seedlings by a blue light receptor cryptochrome 2. *Proc Natl Acad Sci U S A* **95**(5): 2686-90.

Liu, Y., Garceau, N. Y., Loros, J. J and Dunlap, J. C. (1997). Thermally regulated translational control of FRQ mediates aspects of temperature responses in the neurospora circadian clock. *Cell* **89**(3): 477-86.

Locke, J. C., Kozma-Bognar, L., Gould, P. D., Feher, B., Kevei, E., Nagy, F., Turner, M. F., Hall, A and 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 and Turner, M. S. (2005a). Modelling genetic networks with noisy and varied experimental data: the circadian clock in Arabidopsis thaliana. *J Theor Biol* **234**(3): 383-93.

Locke, J. C., Southern, M. M., Kozma-Bognar, L., Hibberd, V., Brown, P. E., Turner, M. S and Millar, A. J. (2005b). Extension of a genetic network model by iterative experimentation and mathematical analysis. *Mol Syst Biol* **1**: 2005 0013.

Lu, S. X., Knowles, S. M., Andronis, C., Ong, M. S and Tobin, E. M. (2009). CIRCADIAN CLOCK ASSOCIATED1 and LATE ELONGATED HYPOCOTYL function synergistically in the circadian clock of Arabidopsis. *Plant physiology* **150**: 834-843.

Lui, X. L., Covington, M. F., Fankhauser, C., Chory, J and Wagner, D. R. (2001). ELF3 encodes a circadian clock-regulated nuclear protein that functions in an Arabidopsis PHYB signal transduction pathway. *Plant Cell* **13** (6): 1293-304.

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

Martinez-Garcia, J. F., Huq, E and Quail, P. H. (2000). Direct targeting of light signals to a promoter element-bound transcription factor. *Science* **288**(5467): 859-63.

Mas, P., Alabadi, D., Yanovsky, M. J., Oyama, T and Kay, S. A. (2003). Dual role of TOC1 in the control of circadian and photomorphogenic responses in Arabidopsis. *Plant Cell* **15**(1): 223-36.

Mas, P., Kim, W. Y., Somers, D. E and Kay, S. A. (2003). Targeted degradation of TOC1 by ZTL modulates circadian function in Arabidopsis thaliana. *Nature* **426**(6966): 567-70.

Mas, P., Devlin, P. F., Panda, S and Kay, S. A. (2000). Functional interaction of phytochrome B and cryptochrome 2. *Nature* **408**(6809): 207-11.

Mathews, S. (2006). Phytochrome-mediated development in land *plants*: red light sensing evolves to meet the challenges of changing light environments. *Mol Ecol* **15**(12): 3483-503.

Matsushika, A., Murakami, M., Ito, S., Nakamichi, N., Yamashino, T and Mizuno, T. (2007). Characterization of Circadian-associated pseudo-response regulators: I. Comparative studies on a series of transgenic lines misexpressing five distinctive PRR Genes in Arabidopsis thaliana. *Biosci Biotechnol Biochem* **71**(2): 527-34.

Matsushika, A., Imamura, A., Yamashino, T and Mizuno, T. (2002). Aberrant expression of the light-inducible and circadian-regulated APRR9 gene belonging to the circadian-associated APRR1/TOC1 quintet results in the phenotype of early flowering in Arabidopsis thaliana. *Plant Cell Physiol* **43**(8): 833-43.

Matsushika, A., Makino, S., Kojima, M and 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**(9): 1002-12.

McClung, C. R. (2006). *Plant* circadian rhythms. *Plant Cell* **18**(4): 792-803.

Mc Cormick, M and Mierendorf, R. (1994). *Novations* **1**: 4.

McGuffin, L. J., Bryson, K and Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* **16**: 404-405.

McWatters, H. G., Bastow, R. H., Hall, A and Millar, A. J. (2000). The ELF3 zeitnehmer regulates light signalling to the circadian clock. *Nature* **408**(6813): 716-20.

Millar, A. J. (2004). Input signals to the *plant* circadian clock. *J Exp Bot* **55**(395): 277-83.

Millar, A. J., Carre, I. A., Strayer, C. A., Chua, N. H and Kay, S. A. (1995a). Circadian clock mutants in Arabidopsis identified by luciferase imaging. *Science* **267**(5201): 1161-3.

Millar, A. J., Straume, M., Chory, J., Chua, N. H and Kay, S. A. (1995b). The regulation of circadian photoperiod by phototransduction pathways in Arabidopsis. *Science* **267** (5201): 1163-6.

Mizoguchi, T., Wright, L., Fujiwara, S., Cremer, F., Lee, K., Onouchi, H., Mouradov, A., Fowler, S., Kamada, H., Putterill, J and Coupland, G. (2005). Distinct roles of GIGANTEA in promoting flowering and regulating circadian rhythms in Arabidopsis. *Plant Cell* **17**(8): 2255-70.

Mizoguchi, T., Wheatley, K., Hanzawa, Y., Wright, L., Mizoguchi, M., Song, H. R., Carre, I. A and Coupland, G. (2002). LHY and CCA1 are partially redundant genes required to maintain circadian rhythms in Arabidopsis. *Dev Cell* **2**(5): 629-41.

Mizuno, T and Nakamichi, N. (2005). Pseudo-Response Regulators (PRRs) or True Oscillator Components (TOCs). *Plant Cell Physiol* **46**(5): 677-85.

Mockler, T., Yang, H., Yu, X., Parikh, D., Cheng, Y. C., Dolan, S and Lin, C. (2003).Regulation of photoperiodic flowering by Arabidopsis photoreceptors. *Proc Natl Acad Sci U S A* **100**(4): 2140-5.

Murakami-Kojima, M., Nakamichi, N., Yamashino, T and Mizuno, T. (2002). The APRR3 component of the clock-associated APRR1/TOC1 quintet is phosphorylated by a novel protein kinase belonging to the WNK family, the gene for which is also transcribed rhythmically in Arabidopsis thaliana. *Plant Cell Physiol* **43**(6): 675-83.

Murzin, A. G., Brenner, S. E., Hubbard, T and Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol* **247**: 536-540.

Nakamichi, N., Kita, M., Ito, S., Yamashino, T and Mizuno, T. (2005). PSEUDO-RESPONSE REGULATORS, PRR9, PRR7 and PRR5, together play essential roles close to the circadian clock of Arabidopsis thaliana. *Plant Cell Physiol* **46**(5): 686-98.

Nelson, D. C., Lasswell, J., Rogg, L. E., Cohen, M. A and Bartel, B. (2000). FKF1, a clock-controlled gene that regulates the transition to flowering in Arabidopsis. *Cell* **101**(3): 331-40.

Niinuma, K., Someya, N., Kimura, M., Yamaguchi, I and Hamamoto, H. (2005). Circadian rhythm of circumnutation in inflorescence stems of Arabidopsis. *Plant Cell Physiol* **46**(8): 1423-7.

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S and Nishimura, Y. (1994). Solution structure of specific DNA complex of

the Myb DNA-Binding domain with cooperative recognition helices. *Cell* **79**: 639-648.

Onai, K. and Ishiura, M. (2005). PHYTOCLOCK 1 encoding a novel GARP protein essential for the Arabidopsis circadian clock. *Genes Cells* **10**(10): 963-72.

Orlicky, S., Tang, X., Willems, A., Tyers, M and Sicheri, F. (2003). Structural basis for phosphodependent substrate selection and orientation by the SCF^{Cdc4} ubiquitin Ligase. *Cell* **112** : 243-256.

Overland, L. (1960). Endogenous rhythm in opening and odour of flowers of cestrum nocturnm. *Am J Bot* **47**: 378-382.

Para, A., Farre, E. M., Imaizumi, T., Pruneda-Paz, J. L., Harmon F. G and Kay, S. A. (2007).PRR3 is a vascular regulator of TOC1 stability in the Arabidopsis circadian clock. *Plant Cell* **19**(11): 3462-73.

Park, D. H., Somers, D. E., Kim, Y. S., Choy, Y. H., Lim, H .K., Soh, M. S., Kim, H. J., Kay, S. A and Nam, H. G. (1999). Control of circadian rhythms and photoperiodic flowering by the Arabidopsis GIGANTEA gene. *Science* **285**(5433): 1579-82.

Pattanayek, R., Wang, J., Mori, T., Xu, Y., Johnson, C. H and Martin, E. (2004). Visualizing a circadian clock protein: Crystal structure of KaiC and functional insights. *Mol cell* **15**: 375-388.

Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P and Saedler, H. (1987). The regulatory CI locus of *Zea mays* encodes a protein with homology to myb-proto-oncogene products and with structural similarities to transcriptional activators. *EMBO J* **6**: 3553-3558.

Pelton, J. T and McClean, L. R. (2000). Spectroscopic methods for analysis of protein secondary structure. *Anal Biochem* **15**; 277(2): 167-76.

Pittendrigh, C. S. (1960a). On temporal organization in living systems. *Harvey Lect* **56**: 93-125.

Pittendrigh, C. S. (1960b). Circadian rhythms and the circadian organization of living systems. *Cold Spring Harb Symp Quant Biol* **25**: 159-84.

Pollastri, G., Przybylski, D., Rost, B and Baldi, P. (2002). Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins* **47**: 228-235.

Pruneda-Paz, J. L., Breton, G., Para, A and Kay, S. A. (2009). A functional genomics approach reveals CHE as a component of the Arabidopsis circadian clock. *Science* **323**(5920): 1481-5.

Reed, J. W., Nagpal, P., Bastow, R. M., Solomon, K. S., Dowson-Day, M. J., Elumalai, R. P and Millar, A. J. (2000). Independent action of ELF3 and phyB to control hypocotyl elongation and flowering time. *Plant Physiol* **122**(4): 1149-60.
Rockwell, N. C., Y. S. Su and J. C. Lagarias (2006). Phytochrome structure and signaling mechanisms. *Annu Rev Plant Biol* **57**: 837-58.

Richards, R. S. (1992). *Protein Sci* **1**:1721.

Robertson, D. L and Joyce, G. F. (1990). Selection *in vitro* of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* **344** (6265):467-468.

Salome, P. A. and McClung, C. R. (2005). PSEUDO-RESPONSE REGULATOR 7 and 9 are partially redundant genes essential for the temperature responsiveness of the Arabidopsis circadian clock. *Plant Cell* **17**(3): 791-803.

Samach, A., Onouchi, H., Gold, S. E., Ditta, G. S., Schwarz-Sommer, Z., Yanofsky, M. F and Coupland, G. (2000). Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis. *Science* **288**(5471): 1613-6.

Sawa, M., Nusinow, D. A., Kay, S. A and Imaizumi, T. (2007). FKF1 and GIGANTEA complex formation is required for day-length measurement in *Arabidopsis*. *Science* **318**(5848): 261-5.

Sawyer, L. A., Hennessy, J. M., Peixoto, A. A., Rosato, E., Parkinson, H., Costa, R and Kyriacou, C. P. (1997). Natural variation in a Drosophila clock gene and temperature compensation. *Science* **278**(5346): 2117-20.

Schaffer, R., Landgraf, J., Accerbi, M., Simon, V., Larson, M and Wisman, E. (2001). Microarray analysis of diurnal and circadian-regulated genes in Arabidopsis. *Plant Cell* **13**(1): 113-23.

Schaffer, R., Ramsay, N., Samach, A., Corden, S., Putterill, J., Carre, I. A and Coupland, G. (1998). The late elongated hypocotyl mutation of Arabidopsis disrupts circadian rhythms and the photoperiodic control of flowering. *Cell* **93**(7): 1219-29.

Schöning, C and Staiger, D. (2005). At the pulse of time: protein interactions determine the pace of circadian clocks. *FEBS Letters* **579**: 3246-3252.

Schultz, T. F., Kiyosue, T., Yanovsky, M., Wada, M and Kay, S. A. (2001). A role for LKP2 in the circadian clock of Arabidopsis. *Plant Cell* **13**(12): 2659-70.

Searle, I and Coupland, G. (2004). Induction of flowering by seasonal changes in photoperiod. *EMBO J* **23**(6): 1217-22.

Sharrock, R. A. and Clack, T. (2002). Patterns of expression and normalized levels of the five Arabidopsis phytochromes. *Plant Physiol* **130**(1): 442-56.

Smith, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P and Kobe, B. (2003). Crystal structures of fusion proteins with large affinity tags. *Protein Sci* **12**: 1313-1322.

Somers, D. E., Webb, A. A., Pearson, M and Kay, S. A. (1998). The short-period mutant, toc1-1, alters circadian clock regulation of multiple outputs throughout development in Arabidopsis thaliana. *Development* **125**(3): 485-94.

Song, H. R. and Carre, I. A. (2005). DET1 regulates the proteasomal degradation of LHY, a component of the Arabidopsis circadian clock. *Plant Mol Biol* **57**(5): 761-71.

Sothern, R. B., Tseng, T. S., Orcutt, S. L., Olszewski, N. E and Koukkari, W. L. (2002). GIGANTEA and SPINDLY genes linked to the clock pathway that controls circadian characteristics of transpiration in Arabidopsis. *Chronobiol Int* **19**(6): 1005-22.

Sreerama, N and Woody, R. W. (2004). Computation and analysis of protein circular dichroism spectra. *Methods Enzymol* **383**: 318-351.

Staiger, D., Allenbach, L., Salathia, N., Fiechter, V., Davis, S. J., Millar, A. J., Chory, J and Fankhauser, C. (2003). The Arabidopsis SRR1 gene mediates phyB signaling and is required for normal circadian clock function. *Genes Dev* **17**(2): 256-68.

Strayer, C., Oyama, T., Schultz, T. F., Raman, R., Somers, D. E., Mas, P., Panda, S., Kreps, J. A and Kay, S. A. (2000). Cloning of the Arabidopsis clock gene TOC1, an autoregulatory response regulator homolog. *Science* **289**(5480): 768-71.

Sugano, S., Andronis, C., Ong, M. S., Green, R. M and Tobin, E. M. (1999). The protein kinase CK2 is involved in regulation of circadian rhythms in Arabidopsis. *Proc Natl Acad Sci U S A* **96**(22): 12362-6.

Sullivan, S., Thomas, C. F., Lamont, D. J., Jones, M. A and Christie, J. M. (2008). In vivo phosphorylation site mapping and functional characterisation of Arabidopsis phototropin 1 . *Mol Plant* **1**:178-194.

Sweeny, B., M (1969). *Rhythmic phenomena in Plants*. London. Academic press.

Sweeny, B., M (1987). *Rhythmic phenomena in Plants*. San Diego. Academic press.

Thomas, B. (2006). Light signalling and Flowering. *Journal of Experimental Botany* **57**(13): 3387-3393.

Thompson, J. D., Higgins, D. G and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acid Res* **22**: 4673-4680.

Tokutomi, S., Matsuoka, D and Zikihara, K. (2008). Molecular structure and regulation of phototropin kinase by blue light. *Biochim Biophys Acta* **1784**(1): 133-42.

Tóth, R., Kevei, É., Hall, A., Millar, A. J., Nagy, F. and Kozma-Bognár, L. (2001). Circadian clock-regualted expression of phytochrome and cryptochrome genes in Arabidopsis. *Plant physiol* **127**: 1607-1616.

TRAFO protocol. http://home.cc.umanitoba.ca/~gietz/Quick.html.

Tseng, T. S., Salomé, P. A., McClung, C. R and Olszewski, N. E. (2004). SPINDLY and GIGANTEA interact and act in Arabidopsis thaliana pathways involved in light responses, flowering, and rhythms in cotyledon movements. *Plant cell* **16**(6): 1550-63.

Tuerk, C and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249** (4968):505-510.

Turck, F., Fornara, F and Coupland, G. (2008). Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. *Annu Rev Plant Biol* **59**: 573-94.

Valverde, F., Mouradov, A., Soppe, W., Ravenscroft, D., Samach, A and Coupland, G. (2004). Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. *Science* **303**(5660): 1003-6.

van Doorn, W. G. and Van Meeteren, U. (2003). Flower opening and closure: a review. *J Exp Bot* **54**(389): 1801-12.

Van Gelder, R. N. (2002). Tales from the crypt(ochromes). *J Biol Rhythms* **17**(2): 110-20.

Volkman, B. F., Lipson, D., Wemmer, D. E and Kern, D. (2001). Two-state allosteric behaviour in a single domain signalling protein. *Science* **291**: 2429-2433.

Wang, Z. Y and Tobin, E. M. (1998). Constitutive expression of the CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) gene disrupts circadian rhythms and suppresses its own expression. *Cell* **93**(7): 1207-17.

Webb, A. A. R. (1998). Stomatal rhythms. In Lumsden PJ, Millar AJ (eds) *Biological rhythms and photoperiodism in Plants*, BIOS Scientific Publishers, Oxford : 69-79.

Wells, L., Vosseller, K and Hart, G .W. (2001).Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc. *Science* **291**(5512): 2376-8.

Whitmore, L and Wallace, B. A. (2008). Protein Secondary structure analysis from Circular Dichroism Spectroscopy: methods and reference databases. *Biopolymers* **89**: 392-400.

Yakir, E., Hilman, D., Harir, Y and Green, R. M. (2007). Regulation of output from the *plant* circadian clock. *FEBS J* **274**(2): 335-45.

Yanovsky, M. J., Mazzella, M. A and Casal, J. J. (2000). A quadruple photoreceptor mutant still keeps track of time. *Curr Biol* **10**(16): 1013-5.

Young, M. W. and Kay, S. A. (2001). Time zones: a comparative genetics of circadian clocks. *Nat Rev Genet* **2**(9): 702-15.

Zeilinger, M. N., Farre, E. M., Taylor, S. R., Kay, S. A and Doyle, F. J. 3rd (2006). A novel computational model of the circadian clock in Arabidopsis that incorporates PRR7 and PRR9. *Mol Syst Biol* **2**: 58.

Zheng, N., Schulman, B. A., Song, L., Miller, J. J., Philip, D. J., Wang, P. Chu, C., Koepp, D. M., Elledge, S. J., Pagano, M., Conaway, R. C., Conaway, J. W., Harper, J. W and Pavletich, P. P. (2002). Structure of the Cul1-Rbx1-Skp1-F-box^{Skp2} SCF ubiquitin ligase complex. *Nature* **416**: 703-709.

Zhong, H. H., Painter, J. E., Salome, P. A., Straume, M and McClung, C. R. (1998). Imbibition, but not release from stratification, sets the circadian clock in Arabidopsis seedlings. *Plant Cell* **10**(12): 2005-17.