Studies of light perception in marine diatoms and discovery of a novel blue light cryptochrome photoreceptor

Thesis

Studies of light perception in marine diatoms and discovery of a novel blue light cryptochrome photoreceptor

Manuela Mangogna

Laurea degree in Biological Sciences

Doctor of Philosophy

Sponsoring Establishment

Stazione Zoologica Anton Dohrn

Naples, Italy

July 2007

ProQuest Number: 13889967

ProQuest 13889967

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Director of Studies: Dr. Chris Bowler

                                        Laboratory of Cell Signaling

                                        Stazione Zoologica "Anton Dohrn"

                                        Naples, Italy

External Supervisors: Dr. Colin Brownlee

                                        Marine Biological Association

                                        Plymouth, United Kingdom

                                        Dr. Bernard Kloareg

                                        Station Biologique

                                        Roscoff, France

CONTENTS

CHAPTER III - RESULTS

CHAPTER IV - DISCUSSION

CHAPTER V - BIBLIOGRAPHY

ACKNOWLEDGMENTS

At the end of this incredible experience it is not easy to remember all the people that helped and encouraged me during these years.

First of all, I would like to sincerely thank my supervisor Dr. Chris Bowler who transmitted me his deep passion for science. Chris has always been constantly present although not physically. He helped me through stimulating discussions and suggestions and he gradually assisted me in developing the project by myself. In this way I have learned how it is important to believe in what you are doing in order to reach your own aim.

I wish to acknowledge the aid from my external supervisors: Dr. Colin Brownlee and Dr. Bernard Kloareg who improved my thesis project with their periodical assistance. Of course, many scientists of the Stazione Zoologica "Anton Dohrn" have contributed to the development of the project with *brainstorming*. In particular, Dr. Maurizio Ribera d'Alcalà was always open to hear my progress in the work. His suggestions have been extremely precious because full of his contagious passion for science.

I cannot forget to underline a fundamental daily event, the coffee break, shared together with oceanographic colleagues such as Serena, Vincenzo, Rosario, Mariella, Céline, Laurent, François, Daniele, and all the others, that was the source for stimulating discussions or even only pleasing moments.

Furthermore, I am grateful for all the people that spent their time in our laboratory in view of the existing pleasant atmosphere that stimulated the performance of our work. Many thanks especially to Sacha Coesel who assisted me directly in this project with passion and to Magali Siaut, with whom I shared the bench, many evenings and many discussions.

Fabio, Ganga, Anton, Mily, Marc, Uma will be impressed in my mind for their peculiar characters, and their helpfulness together with people who recently arrived: Valentina, Sabrina, Raffaella, Frédy, Daniele. Thank you also to Alessandro, our available technician and to Silvia the perfect secretary of the Open University who reminded me all the bureaucratic deadlines.

In addition, I would like to thank Giovanna Benvenuto who was always involved in solving both scientific and non-scientific problems having always the right answer or solution.

A special thank to Angela Falciatore, my "Naples supervisor" and, more important, a real friend. In particular, during the last period, she has been a "reference point" contributing to the coordination of the project with her positive energy.

Of course, I will never forget Margherita Groeben who helped me a lot with a substantial support and mostly with her smile and her friendly words. I will always remember those moments.

Finally, I must thank people who have always believed in me and without whom this adventure would not have been possible. For instance, my wonderful parents who have always supported me in the studies helping me to follow my passions; my lovely sister Viviana who was always ready to help me every time it was necessary; my sister Adriana who encouraged me despite we live far away.

A special thank for my wonderful husband who encouraged me and tolerated me all these years. He was always close to me and a part of the PhD title will be for him.

Dulcis in fundo, special thanks to my wonderful kids, Raimondo and Ludovica, which without any doubts have been my best experiments.

I know that it could appear mainly a boring list for the reader but for me it is a fragment of my life and there is a story in every sentence and in every name.

Finally, it is also very important for me to acknowledge my first scientific supervisors who followed me during my degree, Dr. Francesco Graziani and Dr. Graziella Persico. Both have contributed to my scientific formation with their passion and their advices.

I feel to dedicate this thesis to Dr. Graziella Persico who recently passed away. She was a rigorous scientist and she spent all her life in the scientific field always thinking of young people who were working in team with her.

ABSTRACT

The research project has focused on the structural and functional characterization of a gene denoted *PtCPF1* (Cryptochrome/Photolyase Family 1) that encodes a putative photoreceptor in the diatom *Phaeodactylum tricornutum*. Cryptochromes (cry) are blue light receptors that share sequence similarity with photolyases, flavoproteins that catalyze the repair of UV light-damaged DNA.

In order to characterize the diatom *PtCPF1* gene, expression at both transcriptional and translational levels have been performed in time course experiments designed to study circadian rhythmicity and acute light induction responses. From this analysis *PtCPF1* was shown to be strongly induced under blue light and to be expressed diurnally. In order to understand better the function of the gene product, the protein has been expressed and purified in *E. coli*. Spectral and biochemical analyses of the purified protein have shown that PtCPF1 is a blue-light-absorbing protein with DNA repair activity. On the other hand, localization studies in diatom cells have evidenced the constitutive nuclear localization of the protein.

Interestingly, comparative analysis of the diatom PtCPF1 protein has revealed it to be more similar to the animal cryptochromes than to plant counterparts. Since animal crys act as components of the circadian clock controlling daily physiological and behavioural rhythms and as photoreceptors that mediate entrainment of the circadian clock to light, it was important to elucidate the function of the PtCPF1 protein both in a heterologous system and in an *in vivo* system. Remarkably, transcription assays developed in mammalian cells have evidenced a repressor activity of the PtCPF1 protein within the clock machinery, mimicking the function of animal crys. Furthermore, gene expression studies of transgenic diatom

lines overexpressing *PtCPF1* have indicated that the protein acts as a blue light photoreceptor because it can modulate several blue light-dependent responses. Therefore, this research project has identified a novel protein that displays both blue light photoreceptor activity as well as DNA repair activity. This protein could, in fact, be considered the missing link in the evolutionary history of the Cryptochrome/Photolyase family.

CHAPTER I - INTRODUCTION

1.1 Diatoms

1.1.1 General Characteristics

Oceanic primary productivity is a major effector of global biogeochemical cycles. In fact, the oceanic biota can respond and affect the natural climatic variability by feedback from primary productivity, which influences the geochemistry of the Earth profoundly (Falkowski P.G. *et al.*, 1998). In some regions such as lakes or coastal oceans, phytoplankton blooms can fix approximately the same amount of carbon, a few grams per square meter per day, as a terrestrial forest (Smetacek V., 2001). Today, the oceans cover 70% of the Earth's surface, and on a global scale they are thought to contribute approximately one half of the total primary productivity of the planet. In contemporary oceans marine phytoplankton is composed of photosynthetic bacteria such as prochlorophytes and cyanobacteria, and eukaryotic microalgae such as chromophytes (brown algae), rhodophytes (red algae), and chlorophytes (green algae).

Diatoms are Bacillariophyceae within the division Heterokontophyta, a group of unicellular chromophyte algae that colonize the oceans down to depths to which photosynthetically available radiation can penetrate. They are thought to be the most important group of eukaryotic phytoplankton such that their ecological relevance is very impressive. In fact, diatoms are responsible for approximately 40% of marine primary productivity and contribute close to one quarter of global carbon fixation. They are widely distributed and highly diverse. There are well over 250 genera with perhaps as many as 100,000 existing species with a wide variety of shapes and sizes (Norton T.A. *et al.*, 1996). Giant diatoms can reach 2-5 mm in size whereas the small-celled species are in a range of 5-50 μm. Diatoms can exist as planktonic forms, found in all open water masses, and benthic forms that can grow on

sediments, attached to rocks or macroalgae. Some species can also be found in soil. Curiously, diatoms constitute a large portion of the algae associated with sea ice in the Antarctic and Arctic, and can sometimes form symbioses with nitrogen-fixing bacteria and cyanobacteria in warm oligotrophic seas (Villareal T.A., 1989).

In spite of their ecological relevance, very little is known about the basic biology of diatoms (Falciatore A. & Bowler, C. 2002). The molecular secrets that are behind their success are not known. One possibility is that diatoms have an extraordinary capacity to adapt to different environments. Moreover, planktonic evolution seems to be ruled by protection and not competition. In fact, Smetacek has argued that the many different shapes and sizes of diatoms may reflect defence responses to specific attack systems (1999; 2001).

It has also been proposed that the major factor determining ecological success is the diatom siliceous cell wall. This is also their characteristic feature. For example, a recent study in the diatom *Thalassiosira weissflogii* reported an increase in cell wall silification induced by grazing pressure. Such observations corroborate the idea that plant-herbivore interactions, beyond grazing sensu stricto, contribute to drive ecosystem structures and biogeochemical cycles in the ocean (Pondaven P. *et al.*, 2007).

The highly patterned siliceous external cell wall is composed of amorphous silica $[(SiO_2)n(H_2O)]$, and is known as the frustule (Fig. 1A). It is constructed of two almost equal halves, with the smaller fitting into the larger like a Petri dish. The larger of the two halves is denoted the epitheca whereas the inner one is denoted the hypotheca. Typically each theca is composed of two parts: the valve (which constitutes the larger outer surface) and a girdle (circular bands of silica attached to the edge of the valve). Pattern design of the frustule is reproduced from generation to generation, implicating a strict genetic control of this unknown process. The three

dimensional structure of the silica cell wall is also being widely studied for nanotechnological applications (Gordon R. & Parkinson J., 2005).

Diatoms are generally classified into two major groups depending on the symmetry of their frustule (Van Den Hoeck C. *et al.*, 1997). Centric diatoms are radially symmetrical and mostly planktonic species, whereas pennate diatoms have bilateral symmetry and many genera possess an elongated slit in one or both frustules called the raphe. This latter group is prevalently benthic, and live attached to sediments or other surfaces and is able to glide along surfaces thanks to the presence of the raphe (Fig. 1A). Motility is thought to be based on an actin/myosin system (Poulsen N.C. *et al.*, 1999), and polysaccharides are actively secreted from the raphe. How this process occurs is not well understood.
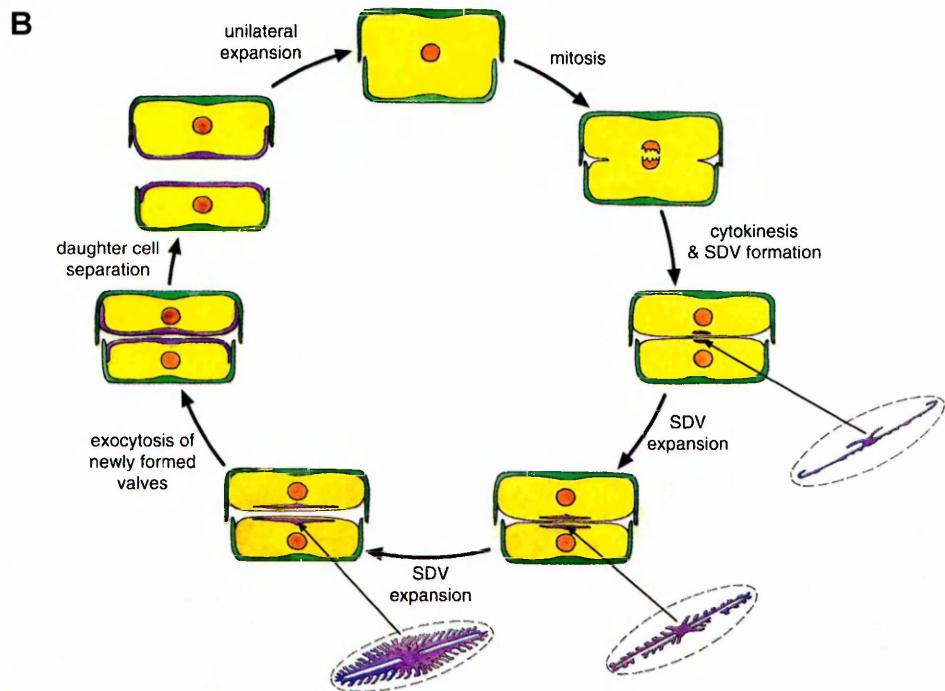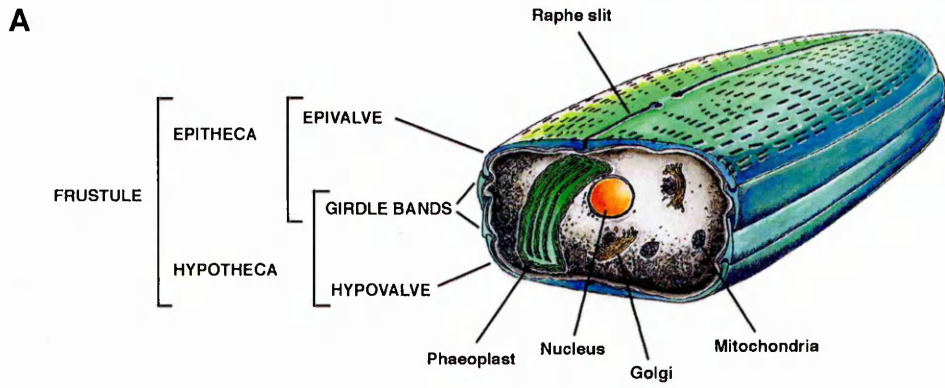
Fig. 1. A. Schematic of general structural features of a pennate diatom. B. Schematic overview of mitotic cell division and hypovalve formation in a pennate diatom (Falciatore A. & Bowler C., 2002).

1.1.2 Cell Division and Cell Wall Biogenesis

Diatom cell division typically proceeds through asexual mitotic divisions to ensure the diploid vegetative state. As previously described, diatom cells are surrounded by a frustule which is a rigid siliceous cell wall composed of two valves. This shell precludes cell growth expansion; therefore the two daughter cells must generate inside the parent cell (Fig. 1B). The epitheca of the parent cell is used as a guide to build a new hypotheca, whereas the other daughter cell uses the parental hypotheca to create an inner theca, such that the parental hypotheca becomes the epitheca of the daughter cell. The consequence of this process is a reduction in size during successive mitotic divisions in one of the two daughter cells. Regeneration of the original size typically occurs via sexual reproduction (see later).

The biogenesis of silica valves is poorly understood, but largely studied based on microscopical observations (Pickett-Heaps J. *et al.*, 1990). Prior to cell division the cell elongates, pushing the epitheca away from the hypotheca, and nuclear division occurs by an "open" mitosis. The two daughter nuclei move slightly in opposite directions and a microtubule center positions itself between each nucleus and the plasma membrane regions where the new hypotheca will be generated. Remarkably, a specialized vesicle denoted silica deposition vesicle (SDV) forms in the central region that becomes the "pattern center". The SDV expands from the center along the rims of the valves, spreading a precise silica lattice network that is then coated with an organic matrix to prevent its dissolution. Once the new valve is complete, the SDV membrane is fused to the plasma membrane, forming the silicalemma, and finally the entire structure is exocytosed (Zurzolo C. & Bowler C., 2001).

As previously mentioned, gametogenesis occurs when cell size decreases to a species-specific critical size threshold: approximately 30-40% of the maximum size.

The resulting male and female gametes combine to create a diploid auxospore that is surrounded by a special organic or inorganic silica wall which allows expansion. Auxospore expansion is a well-controlled process (Mann D.G., 1993), during which the shape of the new enlarged cell is generated. In centric diatoms, sex is almost universally oogamous, with flagellated male gametes, whereas in the pennate diatoms there is more variability (anisogamy, isogamy, automixis). However, only fragmentary information is available because diatom sexuality is in fact limited to brief periods (minutes or hours) that may occur less than once a year in some species and that involve only a small number of vegetative cells within a population (Mann D.G., 1993).

1.1.3 Diatom Photosynthesis

Like in other photosynthetic eukaryotes, the photosynthetic apparatus of diatoms is organized within plastids inside the cell. However, a peculiar characteristic of diatom plastids is that they are enclosed within four membranes rather than two membranes as in land plants (see section 1.1.4). Diatom plastids have been denoted phaeoplasts to distinguish them from the rhodoplasts and chloroplasts of the red and green algae, respectively. In fact, diatoms are brown in color because the fucoxanthin, an accessory carotenoid pigment, is located together with chlorophyll a and c in their light-harvesting complexes. The thylakoid membranes display the typical structure of the Heterokontophyta, being grouped into stacks (lamellae) of three, all enclosed by a girdle lamella. Generally, centric diatoms have a large number of small discoid plastids, whereas pennate diatoms tend to have fewer plastids, sometimes only one, as is the case for the model species *Phaeodactylum tricornutum.*

Fucoxanthin and chlorophylls are bound within the light-harvesting antenna complexes by Fucoxanthin, Chlorophyll a/c–binding Proteins (FCP). The FCP proteins are integral membrane proteins localized on the thylakoid membranes and their primary function is to target light energy to chlorophyll a within the photosynthetic reaction centers. In the pennate diatom *P. tricornutum*, previously studies have identified two gene clusters containing, respectively, four and two individual *FCP* genes (Bhaya D. & Grossman A.R., 1993). They show sequence similarity to the Chlorophyll a/b–binding protein genes (*CAB*) of plants and green algae and, like CAB proteins, diatom FCP proteins are encoded in the nucleus. However, after the sequencing of the *P. tricornutum* and *Thalassiosira pseudonana* genomes (see section 1.1.5) it is possible to talk of 36 *FCP* members in *P. tricornutum* and at least 30 genes in *T. pseudonana*, that appear to belong to different categories. In addition, *FCP* genes appear more scattered all over the genome than within a cluster distribution (Beverley Green, personal communication).

In *Thalassiosira weissflogii*, semi-quantitative RT-PCR analysis of *FCP* gene expression revealed that the transcript levels decrease during prolonged darkness and are highly induced following a subsequent shift to white light (Leblanc C. *et al.*, 1999). Recently, an accurate expression study of the *P. tricornutum FCPB* gene has been described from our laboratory. This quantitative real time PCR-based approach confirmed the diel regulation of the transcript. More specifically, *FCPB* mRNA levels increase throughout the light period, peaking in the early afternoon, and decrease dramatically during the dark period (Siaut M. *et al.*, 2007).

Interestingly, C_4 photosynthesis has recently been proposed in diatoms (Reinfelder J.R. *et al.*, 2000). This specialized form of photosynthesis allows a more efficient utilization of available CO_2 and is restricted to a few land plants, such as

sugar cane and maize. In addition, the report of Reinfelder J.R. *et al.* (2000), suggested that C_4 carbon metabolism may be confined to the cytoplasm. If shown to be a universal feature of diatoms, C_4 photosynthesis may help to explain the ecological success of diatoms in the world's ocean. Unfortunately, the diatom whole genome sequences that have recently become available (see section 1.1.5) have not shed much light on this hypothesis.

1.1.4 Diatom Phylogeny

The plastids of all photosynthetic organisms are likely to be descendants of a primary endosymbiotic relationship in which a cyanobacterium was engulfed by (or invaded) a heterotrophic eukaryote, and eventually lost most of its genes by transfer to the host nucleus. This ancestral photosynthetic lineage diversified into all the modern groups with "primary" chloroplasts: the rhodophyte (red) algae, the glaucophytes, the green algae, and the higher plants. Current knowledge suggests that the initial endosymbiotic event occurred around 1.5 billion years ago and gave rise eventually to two major plastid lineages: chloroplasts and rhodoplasts. Green algae and their descendants, the higher plants, contain chloroplasts and use chlorophyll *a* and *b* in their light harvesting complexes, whereas red algae contain rhodoplasts and utilize chlorophyll *a* and phycobilisomes for the capture of light energy.

Subsequently, a secondary endosymbiotic event occurred. A non-photosynthetic eukaryotic host acquired a eukaryotic endosymbiont that already had a chloroplast, giving rise to the chromophyte algae (Gibbs S.P., 1981; Bhattacharya D. & Medlin L., 1995). Molecular phylogenetic studies and paleoclimatological reconstructions (Kooistra W.H.C.F. *et al.*, 2003) have suggested that this event occurred at around the Permian-Triassic boundary (245 million years ago), which

was a period of intense global changes such as an extreme shortage of dissolved inorganic carbon in ocean surface layers and essentially anoxic deep water (Lee R.E. & Kugrens P., 2000). The endosymbiosis events may have conferred advantages in relation to carbon acquisition for photosynthesis at a time when CO_2 was present at unprecedented low concentrations.

In fact, the chromophyte algae such as diatoms differ fundamentally from the majority of photosynthetic eukaryotes for the organization of their plastid membranes. The plastids of red and green algae and plants are normally surrounded by two membranes, whereas diatom plastids have four membranes. Thus, they required considerable adaptation at the cellular level in term of intracellular transport and the coordination of cellular activities (Cavalier-Smith T., 2000). In some chromophytes, it is possible to observe a second nucleus (the nucleomorph) between the outer and inner two membranes. The sequencing of the nucleomorph of *Guillardia theta* showed the lack of almost all genes for metabolism, which have been transferred to the nucleus of the secondary host, and the presence of many genes for plastid-localized proteins, explaining why the nucleomorph has persisted during evolution (Douglas S. *et al.*, 2001). Comparison of plastid genomes strongly indicates that diatoms acquired their chloroplast from a red algal endosymbiont (Oudot-Le Secq M.P. *et al.*, 2006). However, diatoms have not retained phycobilins for light harvesting and instead use chlorophyll *a* and *c* together with the brown carotenoid fucoxanthin.

The evolutionary history of diatoms can be reconstructed from the fossil record. Diatoms possess a marvelous paleontoligical record. The oldest diatom fossils were found in Lower Cretaceous sediments and they clearly represent centric species. On the contrary, pennate diatoms have never been observed in the Lower

Cretaceous; they appear only around 90 million years ago (Koiistra W.H.C.F. *et al.*, 2003).

1.1.5 Diatom Genomics

In these last years a major focus of the marine phytoplankton community has centred on whole genome sequencing projects, with the hope of improving our understanding of the physiology and cell biology of these organisms. In particular, between 2003 and 2007, the complete nuclear, mitochondrial, and plastid genome sequences have become available for the diatoms *Thalassiosira pseudonana* (Armbrust E.V. *et al.*, 2004) and *Phaeodactylum tricornutum* (see Results section), the red alga *Cyanidioschyzon merolae* (Matsuzaki M. *et al.*, 2004), and the green algae *Ostreococcus tauri* (Derelle E. *et al.*, 2006), *Ostreococcus lucimarinus* (Palenik B. *et al.*, 2007) and *Chlamydomonas reinhardtii* (http://genome.jgi-psf.org).

The laboratory in which this thesis was conducted has been directly involved in the development of genome sequencing projects of the two diatoms, *T. pseudonana* and *P. tricornutum*. Thanks to this information, it has been possible to reveal some molecular features of diatom biology. The *T. pseudonana* nuclear genome is 34.5 mega base pairs distributed in 24 chromosomes with a prediction of around 11,000 genes (Armbrust E.V. *et al.*, 2004). The *P. tricornutum* genome is slightly smaller (27 mega base pairs), with approximately the same number of chromosomes and genes.

Interestingly, almost half of diatom proteins have similar alignment scores to their closest homologs in plant, red algal, and animal genomes, underscoring their evolutionarily history and, most notably, the diatom genomes encode multiple transporters for nitrate and ammonium, as well as a complete urea cycle, previously known only in animals. The role and degree of conservation of the urea cycle in

marine diatoms is still unclear, although it is possible to suggest several important roles, such as signaling for defense or cell death, a way to produce intermediate molecules for silica precipitation, and as a mechanism to store energy (Allen A.E. *et al.*, 2006).

The marine pennate diatom *Phaeodactylum tricornutum* was chosen as the second diatom for whole genome sequencing for several reasons. Primarily this species has been used in laboratory-based studies of diatom physiology for several decades. Furthermore, scientists have progressively established molecular tools aimed at using *P. tricornutum* as a model organism (Falciatore A. & Bowler C., 2002). Unlike other diatoms, this species can exist in different morphotypes and the changes in cell shape can be stimulated by environmental conditions. Moreover it can grow in the absence of silicon because the biogenesis of silicified frustules is facultative. These features can be used to explore the molecular basis of cell shape control and morphogenesis, and the exploration of silica-based nanofabrication, by modulating gene expression by reverse genetic approaches.

Another interesting aspect is the availability of a large EST (expressed sequence tag) collection, generated by our laboratory in recent years. Around 120,000 ESTs have now been assembled and organized in the Diatom EST Database (Maheswari U. *et al.*, 2005), and digital gene expression profiles from cells grown in 14 different conditions can be readily accessed (http://www.biologie.ens.fr/diatomics/EST2).

1.2 Light and Photoreceptors

1.2.1 General Introduction

Sunlight is a primary source of energy for life on Earth and is also a critical information carrier for most organisms. Plants and animals perceive the light environment to gain information about their external world (local surroundings, time of day, season of the year, etc...).

As a result of selection during evolution, organisms have optimised mechanisms for the perception of light signals from the natural environment by utilizing light–sensing proteins known as photoreceptors. A key feature of photoreceptors is the presence of a chromophore able to capture selective light wavelengths. On a spectral basis most biological responses can be confined solely or in concert to the red, blue, or near-UV region. The UV region of the spectrum is subdivided into UV-C (200-280 nm), UV-B (280-320 nm) and UV-A (320-400 nm) and the term "near-UV" typically describes the region above 300 nm. Many different photoreceptor proteins have now been described in the literature, although they can all be classified into a limited number of families, based on the chemical structure of the light-absorbing chromophores that are utilized.

Six photoreceptor families have been characterized to date: rhodopsins, phytochromes, xanthopsins, cryptochromes, phototropins, and BLUF proteins (Van Der Horst M.A. & Hellingwerf K.J., 2004, Falciatore A. & Bowler C., 2005). The primary photochemistry of activation of these photoreceptor proteins changes the configuration of their chromophore. This change in configuration then initiates formation of a signalling state of sufficient stability to communicate the process of photon absorption to a downstream signal transduction partner. The general characteristics of these photoreceptor classes are summarized in Table 1.1, but it is

important to remember that the photobiology of many of these complex molecules remains to be clarified.

Rhodopsin photoreceptors are a family of membrane bound, heptahelical G-Protein-Coupled Receptors (GPCRs) characterized by their ability to covalently bind a retinaldehyde chromophore using a lysine residue located in the seventh transmembrane α-helix. Absorption of photons by the chromophore causes its isomerization, with a consequent conformational change in the molecule that allows the activation of the phototransduction cascade (Bellingham J. & Foster R.G., 2002).

Table 1.1
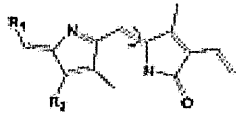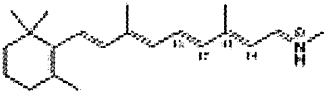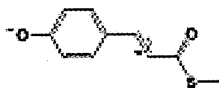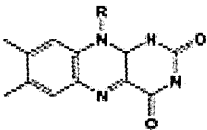
Photosensor Family	Chromophores	Key structural element
Phytochromes	phytochromobilin	
Rhodopsins	retinal	
Xanthopsins	coumaric acid	
Cryptochromes Phototropins BLUF proteins	flavin	

Table 1.1. Classes of Photosensor Families. For descriptions see text.

This photoreceptor family includes the visual rhodopsins of Eukarya, those from Archaea and the ion-translocating prokaryotic rhodopsins. Interestingly, the family still continues to expand. New members have been discovered in *Chlamydomonas* (Sineshchekov O.A. *et al.*, 2002; Nagel G. *et al.*, 2002), in *Neurospora* (Bieszke J.A. *et al.*, 1999), in proteobacteria (i.e., proteorhodopsin; Beja O. *et al.*, 2000), in cyanobacteria (Jung K.H. *et al.*, 2003), and in the vertebrate retina (Provencio I. *et al.*, 1998).

A second photosensor family is made up of the phytochrome molecules, discovered as the receptors responsible for red/far-red light reversible responses in plants. They exist as dimers in the eukaryotic cytoplasm and translocate to the nucleus upon light activation (Sakamoto K. & Nagatani A., 1996). Their light-sensitive chromophore is a linear tetrapyrrole covalently attached to a cysteinyl residue of each apoprotein. Red light triggers a change in chromophore configuration, converting the phytochrome red light (R)-absorbing form (Pr) into the far-red light (FR)-absorbing form (Pfr). Subsequently, it slowly reverts back in the dark or almost instantaneously upon absorption of far-red light. During these transitions, structural changes take place in the protein, and the C-terminal region, through the histidine kinase domain, interacts with signalling partners (Quail P.H., 2002).

Xanthopsins are a family of blue-light induced photoreceptors in which the chromophore is constituted by coumaric acid. The Photoactive Yellow Protein (PYP) has been isolated from the purple bacterium *Ectothiorhodospira halophila* and its crystallographic structure has been solved (Genick U.K. *et al.*, 1998). PYP is a water-soluble protein that displays a typical α/β fold, with a central five-stranded β-sheet and helical segments on either side. The PYP photoreceptor determines a blue-light-induced avoidance response (Sprenger W.W. *et al.*, 1993).

The remaining three families of photoreceptors are blue light responsive and their photochemistry is based on the flavin chromophore. The cryptochrome family is widely distributed in bacteria and eukaryotes. Cryptochromes are involved in many processes ranging from synchronization of the circadian clock in animals and plants, to hypocotyl elongation, seed germination, and pigment accumulation in plants. Moreover, cryptochromes work together with phytochromes to regulate photomorphogenic responses, including the regulation of cell elongation and photoperiodic flowering. The following section 1.3 will explain this photoreceptor family in detail.

Phototropins use the flavin derivative flavin mononucleotide (FMN) as light-sensitive chromophore and mediate several light responses in plants (Christie J.M., 2007), such as phototropism, chloroplast movement, stomatal opening, and the rapid inhibition of hypocotyl growth. The light sensitive domain that generates signaling in this photoreceptor family is referred to as the LOV domain (Light, Oxygen, Voltage). Actually, all known phototropins contain two of these domains, of which the second is the most important for their light-regulated serine/threonine kinase activity (see section 1.3.9).

The last photoreceptor family is called BLUF for "sensors of blue-light using FAD" (Gomelsky M. & Klug G., 2002). Members of this family are involved in photophobic responses in *Euglena gracilis* and in transcriptional regulation in *Rhodobacter sphaeroides* (see section 1.3.9). The BLUF domains bind FAD non covalently, as do the LOV domains with FMN, but the initial photochemistry and structural transitions remain to be resolved.

1.2.2 Light signals underwater

Light is subjected to momentary, diurnal, seasonal and global changes both in irradiance and in spectral distribution. Under water, light irradiance and spectral composition can vary depending on: (1) the incident solar radiation (angular distribution), (2) the inherent optical properties of the water body (absorption and scattering processes), (3) the presence of dissolved organic matter (DOM), and (4) the depth of the water column (Kirk J.T.O., 1994).

Irradiance and light quality also change drastically with depth. Most of the visible light is absorbed within the first 10 metres of the water column, and almost none penetrates below 150 metres, even when the water is very clear. The longer wavelengths of the light spectrum (red, yellow, and orange) can penetrate to approximately 15, 30, and 50 meters, respectively, while shorter wavelengths (violet, blue, and green) can penetrate further, to the lower limits of the euphotic zone, generally considered to extend to a depth of 100 metres. More generally, the photic zone covers the oceans from surface level to 200 meters down. In oceanic water blue light predominates in deep-waters, whereas in coastal water the light is enriched in the green-orange region of the spectrum (see Fig. 2a from Levine J.S. & MacNichol Jr E.F, 1982).

The oceans can be divided into several zones differentiated by depth or light regime (Fig. 2b). The two major parts of the marine environment are the benthic and the pelagic zone. The benthic zone is the part of the ocean associated with the bottom and the pelagic zone is the water column above the benthos. The benthic zone can be divided into several parts depending on depth. The littoral zone extends between the highest and lowest tidal levels. Part of this zone is periodically exposed to air, depending on position within the intertidal range. The sublittoral zone is situated beneath the littoral, from the low tide mark to the continental shelf edge,

about 200 meters deep. Beneath the shelf edge down to about 2,000 meters, the bathyal zone is located. This area coincides with the continental slope. Deeper down the abyssal zone covers the abyssal plains between 4,000 and 6,000 meters. Lastly, the hadal zone includes the trenches, the deepest part of the sea floor, deeper than 6,000 meters. The pelagic area is also divided into different zones. The neritic zone is the water column situated above the continental shelf, between the lowest tidal level to the shelf edge. It is often a region of high productivity because the sunlit surface layers of the water are not far removed from the regeneration of nutrients in the sediment below. Beyond the neritic zone, where the seafloor drops rapidly to great depths, the oceanic zone is situated. The productivity is usually restricted in this zone due to the low availability of nutrients. The mesopelagic is the uppermost region, with its lowermost boundary at the thermocline of 10°C, which in the tropics generally lies between 700 and 1,000 meters. After that is the bathypelagic zone lying between 10°C and 4°C, or between 700 or 1,000 m and 2,000 m or 4,000 m. Lying along the top of the abyssal plain is the abyssalpelagic, whose lower boundary lies at about 6,000 m. The final zone falls into the ocean trenches, and is known as the hadalpelagic. This lies between 6,000 m and 10,000 m and is the deepest oceanic zone.

The spatial, temporal and spectral variability of light experienced by marine phytoplankton differs significantly from that experienced by terrestrial plants, due to the attenuation of solar irradiance in the aquatic medium. Phytoplankton are also incapable of sustained directional movement and are therefore subjected to the environmental conditions in their parent body of water. Moreover, phytoplankton must adapt to relatively rapid changes in both the intensity and the spectral quality of light as they move vertically in a water column. Photosynthesis is responsive to changes in nutrient availability on short time scales and the complementary

chromatic adaptation is a process that allows optimizing absorption of excitation energy thanks to the modulation of different pigments (Grossman A.R., 2003). Changes in pigment content are characteristic of the relatively slow response of "sun-shade" photoacclimation that occurs on timescales typical of mixing in the open ocean. In estuaries, the variations are much faster and subsequently induce rapid changes in the activity of different components of the photosynthetic apparatus. These components modulate light harvesting and Calvin cycle activity, or protect the pigments from excess light (i.e. energy) absorption. When protective capacity is exceeded, photoinhibition occurs (MacIntyre H.L. *et al.*, 2000).

There are several other responses regulated by light, such as vertical migration (Villareal T.A. & Carpenter E.J., 2003), phototaxis (Sineshchekov O. & Govorunova E.G., 2001), chloroplast movement and reorientation (Briggs W.R. & Christie J.M., 2002; Wada M. *et al.*, 2003), circadian rhythms (Rensing L. & Ruoff P., 2002), and the dynamics of phytoplankton blooms (Berger S.A. *et al.*, 2007). Furthermore, the spectral variation of light carries information about the time of day, the vertical position, and the presence of very close neighbours. This is consistent with the widespread occurrence of photoreceptors in marine algae, likely necessary to capture and to transform light signals into intracellular molecular signals. Genomic and functional studies of marine photoreceptors will provide novel information for the understanding of light perception and the mechanisms of light-adaptive responses in this largely unexplored environment.
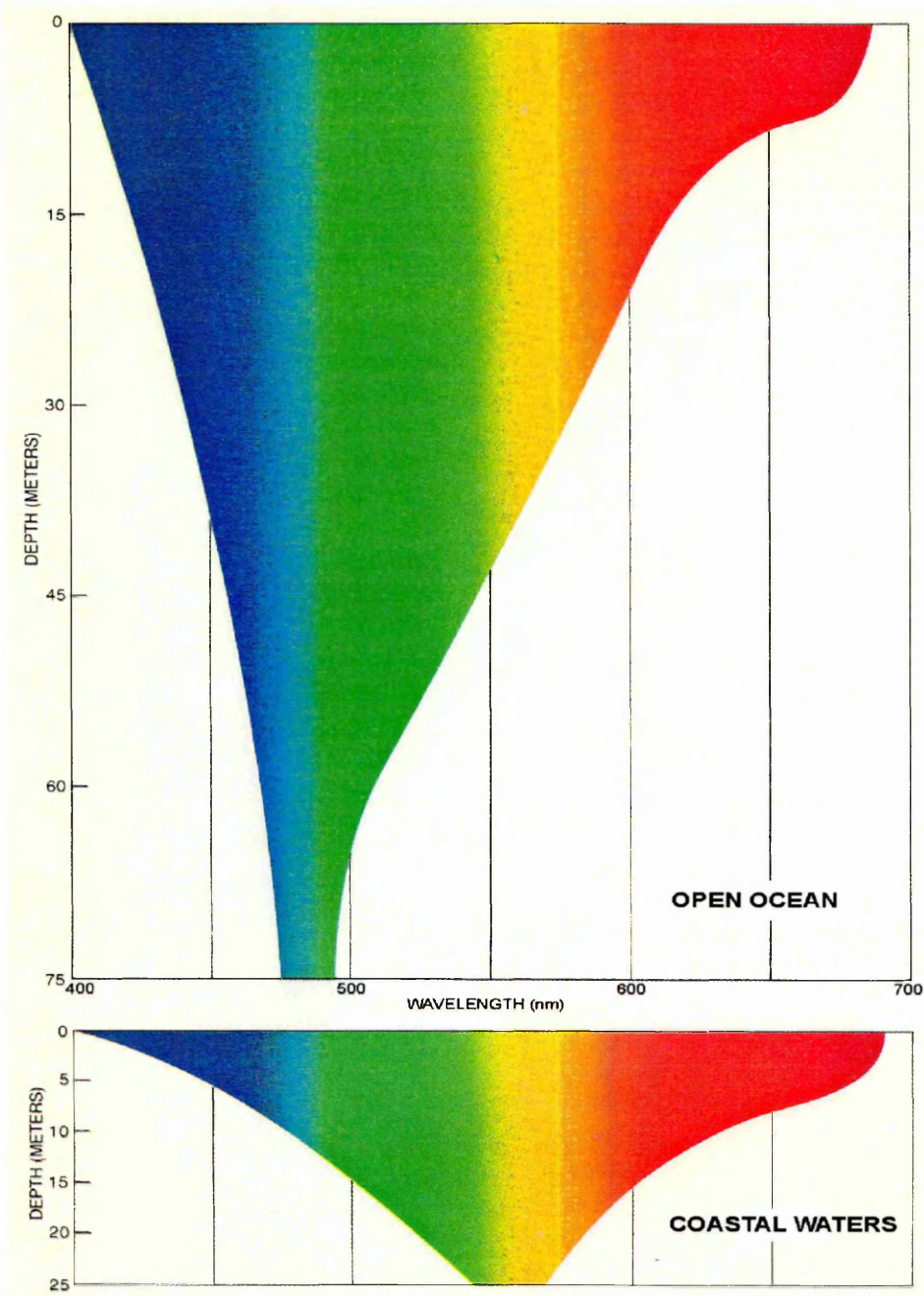
Fig. 2a. Light absorption in oceanic and coastal waters. The attenuation of incident sunlight is more pronounced in red light than blue light in oceanic water, in contrast with estuarine water, where blue light is attenuated more rapidly than red light (Levine J.S. & MacNichol Jr E.F., 1982).
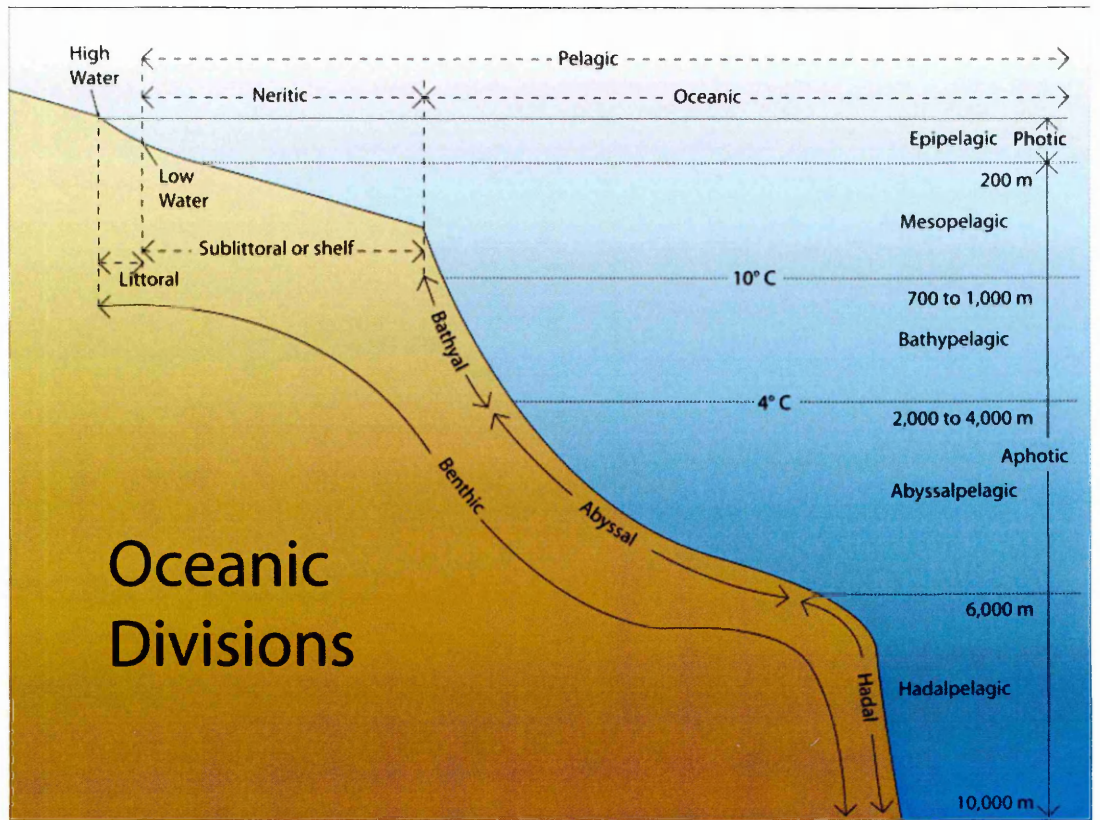
Fig. 2b. The major oceanic divisions. For descriptions see text.

1.3 Blue light Perception and Signaling

1.3.1 Structure of the Cryptochrome/Photolyase Family

The blue region of the light spectrum lies between 350 and 500 nm. It was originally hypothesized that blue light receptors were flavoproteins because the action spectra of many blue light-mediated responses was found to be similar to the absorption spectrum of flavins (Galston A.W., 1950). However, because the nature of the photoreceptor remained unknown and controversial for many years, the cryptic blue light photoreceptor was denoted cryptochrome (Senger H., 1980).

Cryptochromes (cry) were first discovered following the characterization of plant mutants with altered responses to blue light. In particular, the first gene was identified from studies of the *hy4* mutant of *Arabidopsis thaliana*, that showed an impaired inhibition of hypocotyl elongation under blue light (Ahmad M. & Cashmore A.R., 1993). The affected gene was subsequently denoted *CRY1* and a second member (*CRY2*) was subsequently identified (Lin C. *et al.*, 1996). More recently a third gene, denoted *CRY3*, has been identified that seems to cluster in the cry-DASH subclass (Kleine T. *et al.*, 2003; see section 1.3.5). Plant cryptochromes are now known to act as blue/UV–A light receptors which regulate several processes during plant development. Cryptochromes are also widely distributed in the animal kingdom playing an essential role in the regulation of the circadian clock (Lin C. & Todo T., 2005; see section 1.3.6).

In addition, these proteins share sequence similarity to DNA photolyases, a class of enzymes that catalyse blue/UV–A light-dependent repair of DNA damage following exposure to ultraviolet light. Two types of photolyases have been extensively characterized that selectively repair two different lesions: the CPD photolyases that repair the cyclobutane pyrimidine dimers (CPD) and the (6–4)

photolyases that repair pyrimidine (6–4) pyrimidone photoproducts (Sancar A., 2003).

On the contrary, cryptochromes from both plants and animals lack DNA repair activity. All members of the cryptochrome/photolyase family share similarity in the amino-terminal domain, generally known as the PHR domain for Photolyase Homology Region, that is responsible for light-absorption. The catalytic chromophore is the Flavin Adenine Dinucleotide (FAD) in both proteins, whereas a second cofactor can be either a pterin (methenyltetrahydrofolate, MTHF) or a deazaflavin (8-hydroxy-7,8-didemethyl-5-deaza-riboflavin, 8-HDF) in the majority of photolyases, whereas only a pterin has been found in the cry proteins characterized until now.

Another known difference between cryptochromes and photolyases is the presence of a so-called cry carboxy-terminal extension (CCT) that lies beyond the PHR domain, and which is absent in the photolyases (Chen M. *et al.*, 2004). However, little sequence similarity is observed in the CCT domain between different crys. Most plant crys contain a longer CCT extension than animal cryptochromes, while cry–DASH proteins lack this domain. It is therefore difficult to establish a general role for the cry carboxy–terminal domain in cryptochrome signaling. Nevertheless, in the carboxy–terminal extension of plant crys, it is possible to recognize three sequence motifs: (a) DQXVP at the N-terminus, (b) a region containing several acidic residues (E or D), and (c) STAES and GGXVP at the carboxy-terminus separated by a short non-conserved spacer. Collectively these motifs are referred to as the DAS domain (Lin C. & Shalitin D., 2003). A general scheme of the structure of cryptochrome/photolyase family members is represented in Fig. 3.

1.3.2 Evolution of the Cryptochrome/Photolyase Family

Cryptochrome/Photolyase family (CPF) members have been found in all three kingdoms: archaebacteria, eubacteria, and eukaryotes. Notwithstanding, their distribution among species is not uniform. For example, CPD photolyase is found in *Escherichia coli*, *Saccharomyces cerevisiae*, *Drosophila melanogaster*, opossum, and some animal viruses (where it is incorporated in the virion), but it is absent from *Bacillus subtilis*, *Schizosaccharomyces pombe*, and placental mammals. Conversely, the (6-4) photolyase is present in *D. melanogaster*, *Xenopus laevis*, rattlesnake, zebrafish, and *Arabidopsis thaliana*, but not in *Caenorhabditis elegans* or humans and has not yet been found in eubacteria or archaea. In contrast, cryptochrome has been found in several eubacteria, including *Vibrio cholerae* (two members), *A. thaliana* (three members), zebrafish (six members), *X. laevis* (three members), *D. melanogaster* (one member), humans and mice (two members each), but not in archaea or in *C. elegans*. Thus, *D. melanogaster* contains all three members of the cry/photolyase family, *C. elegans* lacks all three, humans have cryptochrome but no photolyase, and *E. coli* has CPD photolyase but no (6-4) photolyase or cryptochrome.

Phylogenetic trees based on the hundreds of cry/photolyase sequences present in public databases have been constructed (Brudler R. *et al.*, 2003; Cashmore A.R., 2003; Kleine T. *et al.*, 2003; Partch C.L. *et al.*, 2005). However one of the most intriguing aspects is to try to reconstruct the evolutionary relationships between these proteins and their acquisition of several functions in different organisms. Because cryptochromes were initially thought to be absent in bacteria, a central dogma up until recently was that a bacterial DNA photolyase was the common ancestor of plant and animal crys. In addition, because phylogenetic analysis revealed that animal crys were more similar to (6-4) photolyases, whereas plant crys

were more related to CPD photolyases, the common idea has been that animal and plant crys derived from two independent evolutionary events after the plant-animal divergence (Cashmore A.R. *et al.*, 1999). Recently, the discovery of a cry-DASH in the cyanobacterium *Synechocystis* sp. PCC 6803 indicates that cryptochromes evolved before the origin of eukaryotic organisms (Brudler R. *et al.*, 2003). However there is also the possibility of a convergent evolution within the cry family. The discovery of *CRY3* in *Arabidopsis* led to speculation that plant crys have evolved from two independent horizontal transfer events, the *Arabidopsis CRY1* and *CRY2* genes having originated from an endosymbiotic α–proteobacteria–like ancestry, while *CRY3* originated from an endosymbiotic cyanobacteria–like ancestor (Kleine T. *et al.*, 2003), i.e., that *CRY1* and *CRY2* are derived from the mitochondrial ancestor and that *CRY3* is derived from the ancestor of the chloroplast.

Although there is no consensus conclusion emerging from these analyses, the general view is that photolyase and cryptochrome have a single progenitor. A primordial organism may have employed a photosensory pigment to detect light and regulate its physiology with a 24 hour periodicity (circadian = about a day) of the geophysical light–dark cycle. The same pigment may have been used to repair DNA damage. Later mutation and selection, in some cases accompanied by gene duplication, could have produced enzymes that perform more specifically one function or the other, thus giving rise to the current dogma that photolyases repair DNA and that cryptochromes regulate the daily oscillations in organismal and/or cellular physiology.

Cryptochrome-like proteins have also been identified in two species of marine diatoms, the centric diatom *T. pseudonana* and the pennate diatom *P. tricornutum* (Falciatore A. & Bowler C., 2005). The identification of putative blue light receptors in these unicellular algae was not surprising because blue light is

abundant in the ocean (see section 1.2.2). A complete phylogenetic analysis of the diatom Cryptochrome/Photolyase Family is provided in this thesis (see section 3.2.2).
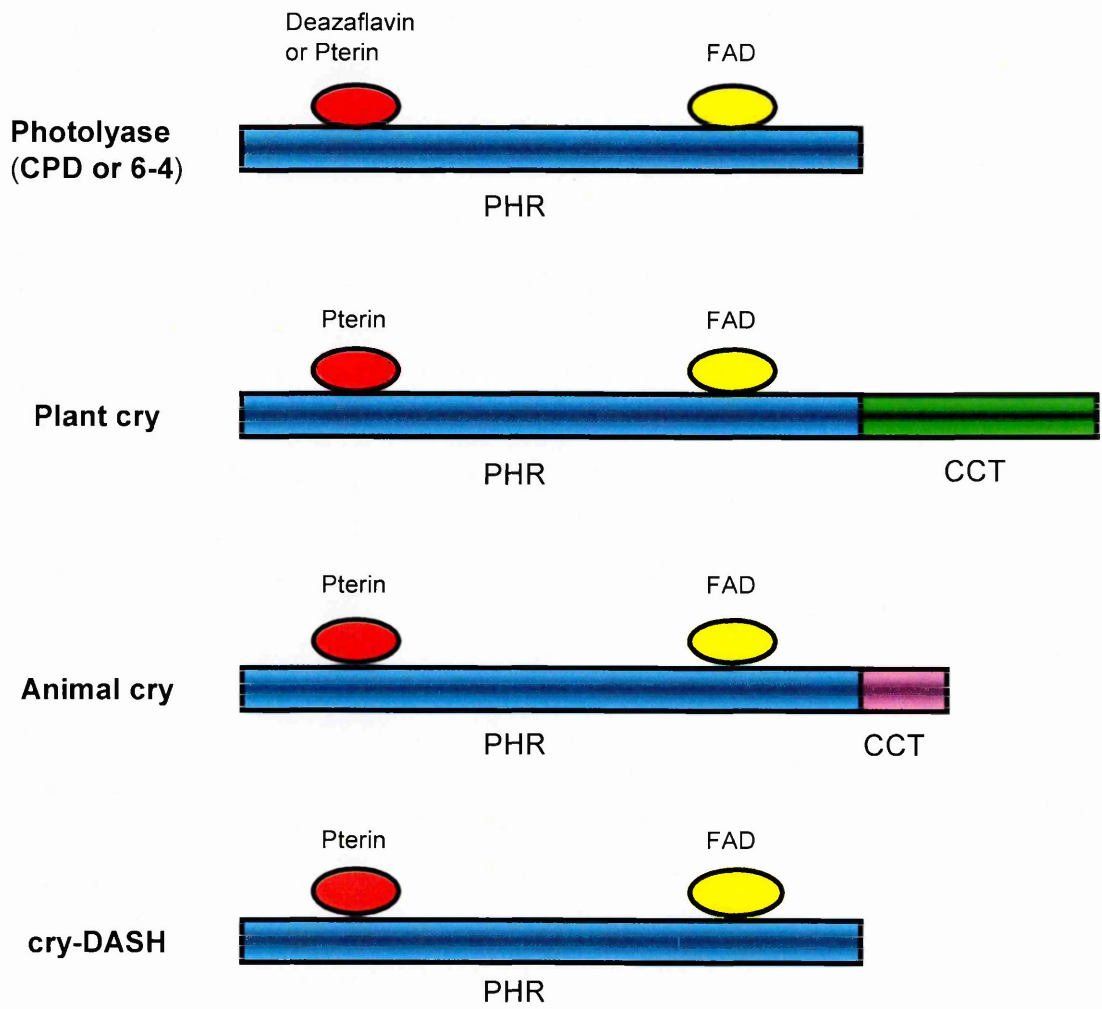
Fig. 3. Schematic representation of the Cryptochrome/Photolyase Family.

1.3.3.1 Structures of chromophores

Photolyases are monomeric proteins of 450-550 amino acids that utilize visible light as an energy source to cleave photoproducts in DNA. These proteins carry two non-covalently bound chromophore cofactors, a catalytic cofactor and a light-harvesting cofactor.

As previously mentioned, the catalytic cofactor is always FAD, whereas the second cofactor can be different. Two prosthetic groups are well known as light–harvesting cofactors: the MTHF for the photolyases classified into the folate class, which is found in the CPD photolyase from *E. coli*, and the 8-HDF for enzymes of the deazaflavin class, which is found in the CPD photolyase from *Anacystis nidulans*. However, recent experimental evidence has shown the existence of other antenna chromophores. For example, flavin mononucleotide (FMN) has been found as a light–harvesting cofactor in photolyase from *Thermos thermophilus* HB8 (Ueda T. *et al.*, 2005) and more surprisingly, the crystal structure of archaeal CPD photolyase from *Sulfolobus tokodaii* has shown another FAD molecule at the position of the light–harvesting cofactor (Fujihashi M. *et al.*, 2007), making the classification more complex.

The flavin can be considered the most commonly used cofactor in nature and FAD is the most common form of flavin found in enzymes. The flavin can be reduced and oxidized by one- and two-electron-transfer reactions utilizing a redox switch between NADH and heme groups, which can carry out two- and one-electron-transfer reactions, respectively (Walsh C.T., 1986). For photolyase activity FAD is the essential chromophore both for specifically binding to damaged DNA and for catalysis. The active form of flavin is the two-electron-reduced form and

deionized form, FADH⁻ (Kim S.-T. *et al.*, 1993). This cofactor is bound non-covalently but very tightly to *E. coli* and *A. nidulans* CPD photolyases and can be released only after mild denaturation of the enzymes (Jorns M.S. *et al.*, 1990). The purified CPD photolyase from *E. coli* contains the FAD at virtually 1:1 stoichiometry with respect to the apoenzyme (Sancar A. & Sancar G.B., 1984). However, many cryptochrome/photolyase family proteins expressed in heterologous systems do not contain stoichiometric amounts of FAD. For example, *Drosophila melanogaster* (6-4) photolyase and cryptochrome proteins overproduced in *E. coli* contain only 1-5% FAD (Zhao X. *et al.*, 1997; Selby C.P. & Sancar A., 1999, respectively), as do human cryptochromes 1 and 2 (Hsu D.S., *et al.*, 1996; Zhao S. & Sancar A., 1997, respectively). Regarding the flavin cofactor of the photolyases, it is important to point out the redox status of the cofactor. FAD can be found in three redox states: oxidized (FAD_{ox}), one-electron-reduced (neutral blue radical or anionic red radical, FADH⁻), and two-electron-reduced forms (neutral or anionic, FADH*). Under physiological conditions FAD is synthesized and incorporated into the appropriate apoenzymes in the FAD_{ox} form. The catalytic cycle converts the oxidized form into one- and two-electron-reduced forms, but how this mechanism occurs for the photolyase flavin is unknown. In addition, at present there is no evidence for light-independent redox reactions carried out by photolyases.

The second chromophore present in the photolyases is not essential for activity, but may increase the rate of repair by 10-100 fold because they have a higher extinction coefficient than FADH⁻ in the near-UV/blue region, thus they are responsible for absorbing > 90% of the photoreactivation photons in sunlight. The MTHF cofactor is the photoantenna most frequently found in photolyases. In contrast to flavin, the folate generally dissociates readily from the apoprotein. The purified *E. coli* CPD photolyase contains substoichiometry (20-30%) folate (Hamm-

Alvarez S.F. *et al.*, 1990). On the contrary, the 8-HDF cofactor was first discovered in anaerobic methanogenic bacteria and was considered as an ancient molecule because of its relative abundance in archaea. The binding to the apoenzyme is tight and 8-HDF is present in stoichiometic amounts in all well-characterized deazaflavin class enzymes.

1.3.3.2 Structure of photolyases

The major chemical lesion caused by UV light wavelengths in sunlight in DNA is a cyclobutane pyrimidine dimer formed by the cyclo–addition of two adjacent pyrimidine rings, generally between two thymine residues. This modification can be specifically repaired by CPD photolyases.

The amino acid sequences of about 50 CPD photolyases are currently known. The sequences of these proteins reveal varying degrees of homology ranging from 15% to more than 70% sequence identity. Several points of interest emerge from sequence alignments. The C–terminal 150 amino acids exhibit the highest degree of homology among the folate and the deazaflavin classes and this region was predicted to be the FAD binding domain. Second, plant and animal CPD photolyases show a limited degree of homology to microbial photolyases, and hence, CPD photolyase can be classified into class I (found mainly in microbes) and class II (found principally in animals and plants) on the basis of sequence similarity (Yasui A. *et al.*, 1994). However, more classes were proposed recently as a result of the growing number of available gene sequences.

The crystal structure of photolyases from four different organisms have been reported: *E. coli* (Park H.W. *et al.*, 1995), *A. nidulans* (Tamada T. *et al.*, 1997), *Thermus thermophilus* (Komori H. *et al.*, 2001) and, more recently, *Sulfolobus tokodaii* (Fujihashi M. *et al.*, 2007). Remarkably, although these enzymes show only

20-25% sequence identity, the traces of C_α backbone atoms are similar among all family members.

The *E. coli* CPD photolyase (class I) is composed of two domains: an N-terminal α/β domain (residues 1-131) and a C-terminal α-helical domain (residues 204-471). The two domains are connected by a long and structured interdomain loop of 72 residues. The light-harvesting cofactor MTHF binds in a cleft between the two domains, whereas the FAD adopts a U-shaped conformation between two helix clusters in the center of the helical domain and is accessible through a hole in the surface of this domain (Fig. 4A). Dimensions and polarity of the hole match those of a cyclobutane pyrimidine dimer, suggesting that the photoproduct "flips out" of the helix to fit into this hole, and that electron transfer between the flavin and the CPD occurs over van der Waaals contact distance (Park H.W. *et al.*, 1995).

Site-specific mutagenesis studies of the positively charged residues show that the distorted DNA backbone is an important contributor to specificity. Mutations of either aromatic residues or polar residues lining the side walls of the hole drastically reduce the affinity of the enzyme for the cyclobutane pyrimidine dimer (Vande Berg B.J. & Sancar G.B., 1998). In particular, mutation of Trp277 in *E. coli* CPD photolyase to a non-aromatic residue virtually eliminates specific binding, suggesting that Trp277 plays a crucial role in specific binding (Li Y.F. & Sancar A., 1990).

The crystal structure of class I CPD photolyase from the cyanobacterium *A. nidulans* (also known as *Synechococcus* sp PCC6301) showed the typical backbone structure composed of an α/β domain and an α-helical domain but revealed a completely different binding site for the light-harvesting cofactor (Fig. 4B). In fact, the antenna cofactor, identified as 8-HDF in *A. nidulans*, is bound in the α/β domain at a quite different position compared to the corresponding MTHF cofactor in *E. coli*

CPD photolyase, whereas the amino acid residues which interact with FAD are entirely conserved in the corresponding positions in both photolyases (Tamada T. *et al.*, 1997).

(Park H.W. *et al.*, 1995)

(Tamada T. *et al.*, 1997)

(Komori H. *et al.*, 2001)

(Fujihashi M. *et al.*, 2007)
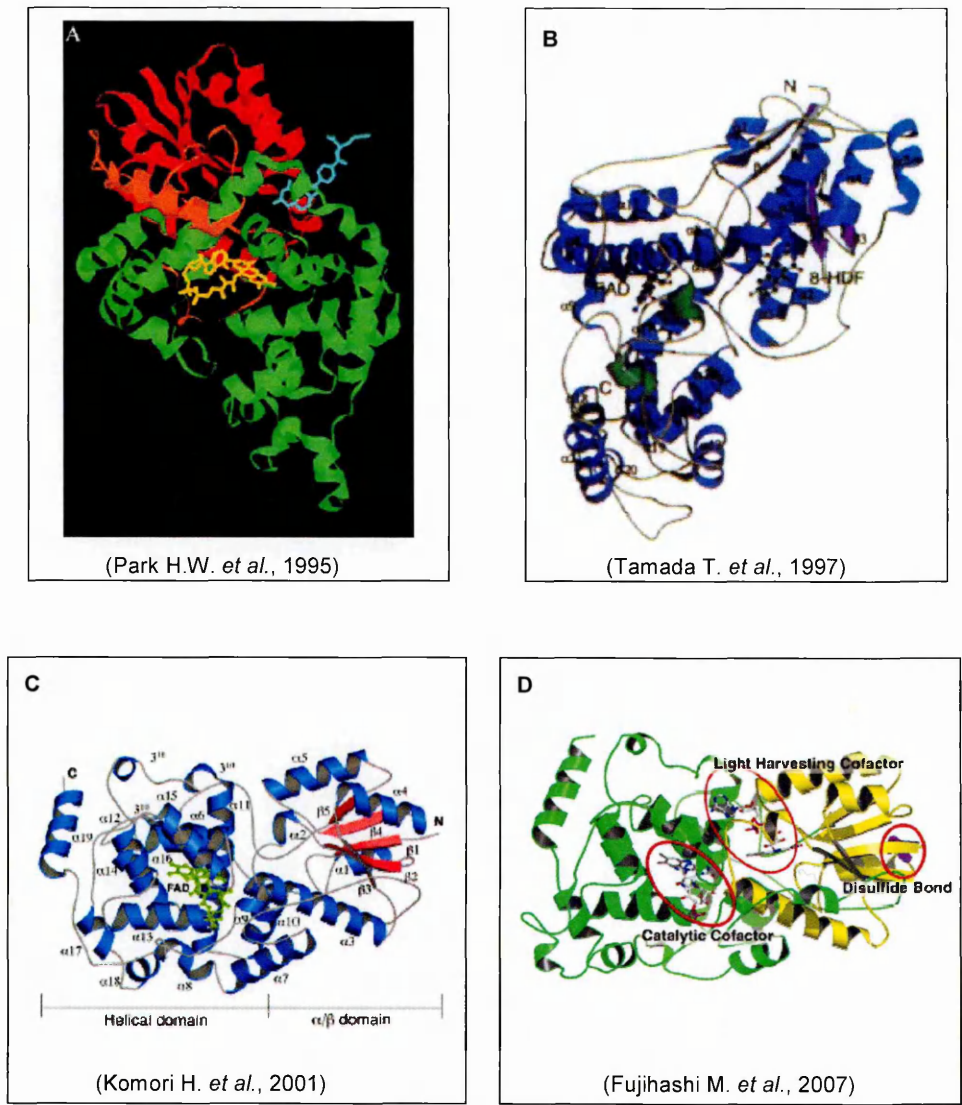
Fig. 4. Crystal structures of CPD photolyases from *E. coli* (**A**), *A. nidulans* (**B**), *T. thermophilus* (**C**), *S. tokodaii* (**D**). Details are described in the text.

- 36 -

The photolyase from *T. thermophilus* is a thermostable class I CPD photolyase. The crystal structure revealed the conserved topology: an α/β domain in the N-terminal region and an α–helical domain in the C-terminal region in which the FAD is located (Fig. 4C, Komori H. *et al.*, 2001). At that time, the protein overproduced in *E. coli* did not contain the light-harvesting cofactor. Subsequently, Ueda T. *et al.* (2005) developed a method for the purification of this enzyme from *E. coli* that retained its second chromophore, and identified it as FMN. However, this does not rule out the possibility that the *T. thermophilus* CPD photolyase does not contain the FMN *in vivo*.

More recently, the crystal structure of the first archaeal photolyase from *Sulfolobus tokodaii* has been determined (Fujihashi M. *et al.*, 2007). The overall structure has the two domains found in the three known photolyases, and the two cofactors are found at essentially the same sites. Surprisingly, another FAD molecule is found at the position of the light-harvesting cofactor, well accommodated in the crystallographic structure, suggesting the existence of a different mechanism to recognize CPD dimers in genomic DNA (Fig. 4D). No crystal structures of (6-4) photolyases have been determined until now, but based on the high similarity to CPD enzymes, it is possible to envisage a common tridimensional structure.

1.3.3.3 Reaction Mechanism

Photolyases carry out catalysis by Michaelis-Menten kinetics. They bind S (substrate) to form ES (enzyme-substrate), which performs catalysis to yield EP (enzyme-product), and then P dissociates. However, the mechanism differs from classic Michaelis-Menten kinetics in one important aspect: the transition ES into EP

is strictly light dependent (hv = light energy), as explained by the following reaction formula:

$$E + S \rightleftharpoons ES \xrightarrow{h\nu} E + P$$

Photolyase binds the photoproduct generated by UV-light in DNA independently of light, and flips the dimer out of the double helix into the active site cavity to make a stable ES complex.

CPD photolyases catalyze the repair of cyclobutane pyrimidine dimers (Pyr<>Pyr). The photochemical reaction (schematically represented in Fig. 5A) is initiated when the photoantenna chromophore (MTHF or 8-HDF) absorbs a near-UV/blue-light photon and transfers the excitation energy to the FADH$^-$ cofactor by Förster resonance energy transfer (via a dipole-dipole interaction). The excited singlet state flavin (FADH^{-*}) transfers an electron to the Pyr<>Pyr dimer to generate FADH* neutral radical and Pyr<>Pyr$^{-\bullet}$ anionic radical. The cyclobutane ring of the dimer radical then spontaneously rearranges to yield two canonical pyrimidines, and an electron is concomitantly transferred back to the nascently formed FADH* to generate the FADH$^-$ form. The repaired dinucleotide then flips out of the enzyme and into the DNA duplex and the enzyme dissociates from the DNA.

The mechanism for (6-4) photolyases (Fig. 5B) is thought to be similar except that thermal conversion (kT) of the photoproduct to the oxetane intermediate occurs upon formation of the ES complex (Zhao X. *et al.*, 1997). Examined overall, the reaction is a photon-powered cyclic electron transfer that does not result in a net gain or loss of an electron and hence it is not possible to define it as a redox reaction. Studies of the mechanism of photoactivation (light-induced reduction of the flavin adenine dinucleotide) of *E. coli* photolyase showed that three tryptophan residues are involved in the electron transfer chain from the protein surface to the FAD cofactor

for its activation (Trp306, Trp359, Trp382; Aubert C. *et al.*, 2000). Indeed, these residues are highly conserved within cryptochrome family members suggesting that the electron transfer chain is involved in the function of cryptochromes as well as photolyases.
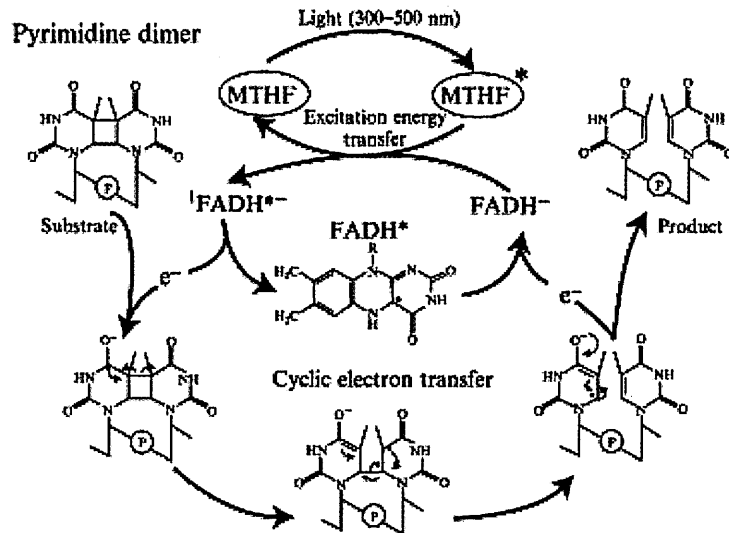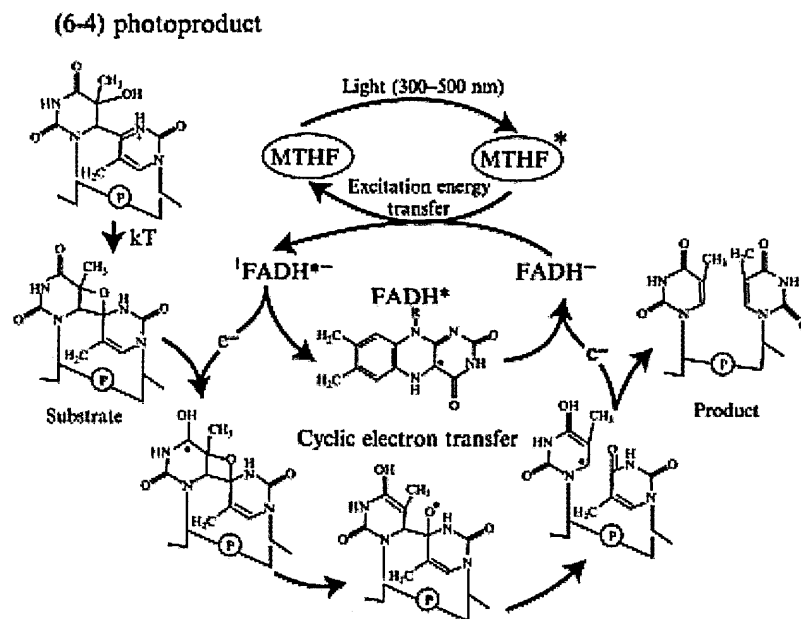
Fig. 5. Reaction mechanisms of pyrimidine dimer photolyases (A) and (6-4) photolyases (B). The antenna chromophores absorb a photon and transfer the excitation energy to FADH$^-$, producing ^{1}(FADH^{*-}). An electron is transferred from FADH^{*-} to the dimer, initiating a cycloreversion reaction that restores both the pyrimidines to their original undamaged state. FADH$^-$ is regenerated by back electron transfer from the dimer. The mechanism for the (6-4) photolyases is thought to be similar except that thermal conversion (kT) of the photoproduct to the oxetane intermediate occurs upon formation of the ES complex. The oxetane intermediate is the substrate for the photochemical reaction (Sancar G.B. & Sancar A., 2006).

1.3.4.1 Structures of cryptochromes

Although initially vague, the term Cryptochrome has now assumed a precise meaning: a photolyase sequence homologue with no DNA repair activity but with blue-light-activated enzymatic function (Sancar A., 2003). Most cryptochromes, with the exception of cry-DASH proteins, are composed of two domains, an amino-terminal photolyase-related (PHR) region and a carboxy-terminal domain of varying size. The PHR region of cryptochromes appears to bind two chromophores: the FAD and the 5,10-methenyltetrahydrofolate (denoted pterin or MTHF). As previously described, the carboxy-terminal domain of cryptochromes is generally less conserved than the PHR region. This extension in plant crys is long and contains the DAS domain (see section 1.3.1), animal crys have a shorter C-terminal extension, whereas cry-DASH proteins lack this domain.

Because there is strong similarity in cry/photolyase family members, at the amino acid level, it appears extremely important to determine the structure of different members in order to elucidate the unknown photochemical mechanism of the cryptochrome protein. Based on this consideration, several research groups have studied the tridimensional structures of cryptochrome family members.

In particular, crystal structures of two members of the cry-DASH subfamily have been determined: the cry-DASH protein from the cyanobacterium *Synechocystis* sp. PCC6803 (Brudler R. *et al.*, 2003) and the CRY3 protein from *A. thaliana* (Huang Y. *et al.*, 2006; Klar T. *et al.*, 2007). Both cryptochromes display a backbone structure similar to photolyases. The photolyase homology region (PHR) is constituted by an α/β domain and a helical domain, which are connected by a variable loop that wraps around the α/β domain. The FAD cofactor is located

between the two lobes of the helical domain in a U-shaped conformation, with its adenine and isoalloxazine rings positioned at the bottom of the cavity (Fig. 6). The antenna cofactor in the cry-DASH protein from *Synechocystis* sp. PCC6803 has not been reported, whereas the crystal structure of *A. thaliana* CRY3 revealed MTHF as light-sensitive cofactor in addition to the catalytic cofactor FAD, both non-covalently bound. The residues responsible for binding MTHF are not conserved in *E. coli* CPD photolyase but are strongly conserved in the cry-DASH subfamily of cryptochromes (Huang Y. *et al.*, 2006). Interestingly, the crystal structure of *A. thaliana* CRY3 in a dimeric state with the antenna chromophore MTHF bound along a dimer interface has recently been reported (Klar T. *et al.*, 2007).

The crystal structure of the photolyase-like domain of CRY1 (PHR-CRY1) from *A. thaliana* has also been determined (Brautigam C.A. *et al.*, 2004). This structure revealed a fold very similar to photolyases with a single molecule of FAD non-covalently bound to the protein. It is possible to recognize the two typical domains: α/β N-terminal domain and α C-terminal domain (Fig. 6D). Interestingly, despite the overall structural similarity, the PHR-CRY1 region has several structural characteristics that differ from photolyase and cry-DASH. Photolyase has a generally positively charged groove running through the FAD-access cavity, which is where the DNA interacts, whereas cry-DASH has a similar positive electrostatic potential on the surface around the cavity. Unlike cry-DASH, the surface of PHR-CRY1 is predominantly negatively charged. Moreover, since experimental evidence has previously demonstrated that CRY1 binds ATP in the presence of Mg $^{2+}$ (Bouly J.P. *et al.*, 2003), Brautigam C.A. *et al.* (2004) succeeded to determine the crystal structure of the photolyase-like domain of CRY1 containing a single molecule of an ATP analog. They demonstrated that the location of this ATP-binding site is equivalent to the putative pyrimidine-dimer binding site in photolyases (Brautigam

C.A. *et al.*, 2004), whereas neither photolyase nor cry-DASH have been reported to bind ATP.

1.3.5 Cryptochrome-DASH

Recently, phylogenetic analysis allowed the identification of a novel class of cryptochromes most closely related to the animal cryptochromes and (6-4) photolyases. Brudler, R. *et al.*, 2003 decided to name this class "cryptochrome DASH", to underscore the phylogenetic relationship of the first two cry-DASH members identified in *Arabidopsis* and *Synechocystis* with the animal cryptochromes, first identified in *Drosophila* and *Homo sapiens* (although cry-DASH itself is not found in *Drosophila* and humans). Cry–DASH proteins have now also been identified in marine bacteria, algae, fungi, and in other vertebrates and so the name is not entirely appropriate (Daiyasu H. *et al.*, 2004).

A general function for this subfamily has not yet been clarified. Localization studies showed that *Arabidopsis* CRY3 is the only plant cry targeted to chloroplasts and mitochondria and is also distinct from the other plant crys because it has DNA-binding activity (Kleine T. *et al.*, 2003). Interestingly, binding to DNA in a sequence-independent fashion was also observed for *Synechocystis* cry-DASH (Brudler R. *et al.*, 2003) and mouse CRY1 (Kobayashi K. *et al.*, 1998). Recently, it has been demonstrated that cry–DASH proteins from bacterial (*V. cholerae* CRY1), plant (*A. thaliana* CRY3), and animal sources (*X. laevis* cry-DASH) show photolyase activity with a high degree of specificity for cyclobutane pyrimidine dimers in single stranded DNA, suggesting the identification of a new class of photolyases (Selby C.P. & Sancar A., 2006).

On the other hand, comparison of microarray expression profiles between *Synechocystis* wild-type and cry-DASH knockout mutant strains was used as

evidence to suggest that cry-DASH functions as a transcriptional repressor (Brudler R. *et al.*, 2003). Therefore, the function of cry-DASH proteins is still an open question. Based on the localization of *Arabidopsis* CRY3, it is intriguing to speculate that the protein could function as a transcriptional regulator of chloroplast and mitochondrial-encoded genes, while Selby C.P. and Sancar A. (2006) suggested that the single strand DNA repair activity may contribute to protect ssDNA viruses.

1.3.6 Circadian rhythms

A circadian rhythm is defined as the oscillation of a physiological behavioural function with a periodicity of about 24 hours. The circadian system is presumed to confer a selective advantage because it allows organisms to anticipate the daily changes occurring in the environment that are a consequence of the Earth's rotation around the sun. The rhythm is an innate property of the organism and the amplitude and the period length are typically maintained under constant environmental conditions for several days.

Schematically, the clock consists of three components: an input pathway that constitutes the external signal to entrain the process, a central oscillator that generates the oscillation, and an output pathway that couples the oscillator to circadian-regulated responses (Fig. 7). Generally, the regulation of the clock involves a transcription/translation feedback loop with positive or negative regulatory elements (Devlin P.F., 2002). In plants, some examples of this process are leaf movements, the opening and closing of stomatal pores, and flowering. In animals, examples include the control of body temperature and sleep-wake cycles (Millar A.J., 2004; Sancar A., 2004).

A central property of circadian rhythms is their ability to be synchronized with the environment by light. Even though heat and other environmental inputs can affect

the phase, the amplitude and the period length of the rhythm, by far the most predominant and perhaps the only physiologically relevant environmental cue (or zeitgeber, from German zeit = time and geber = giver) is light.
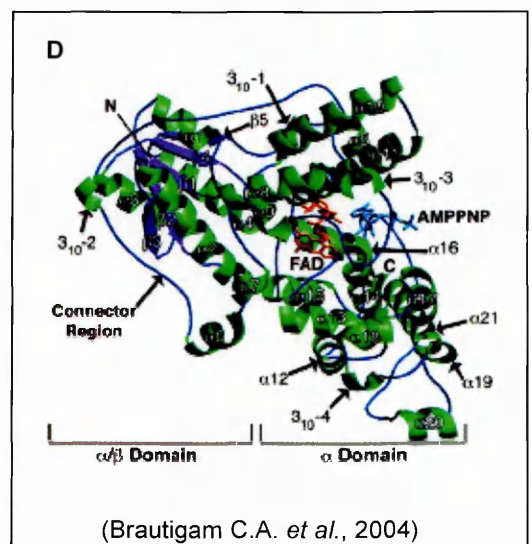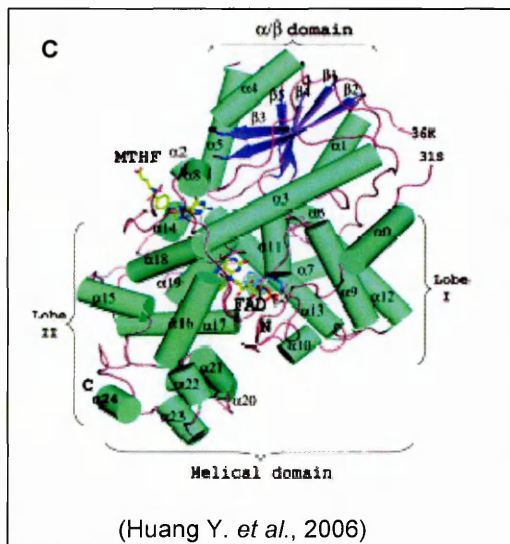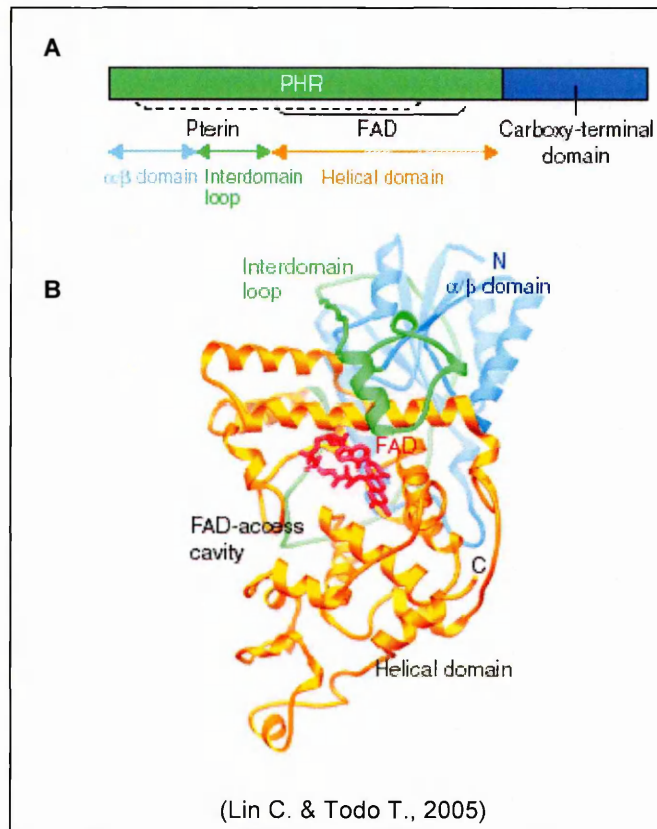
Fig. 6. A. Schematic representation of typical Cryptochrome/Photolyase superfamily proteins. **B.** Crystal structure of CRY-DASH from *Synechocystis* sp. PCC6803. **C.** Crystal structure of CRY3 from *A. thaliana* and (**D**) crystal structure of the PHR domain of *A. thaliana* CRY1 protein. Details are described in the text.

The Circadian Clock
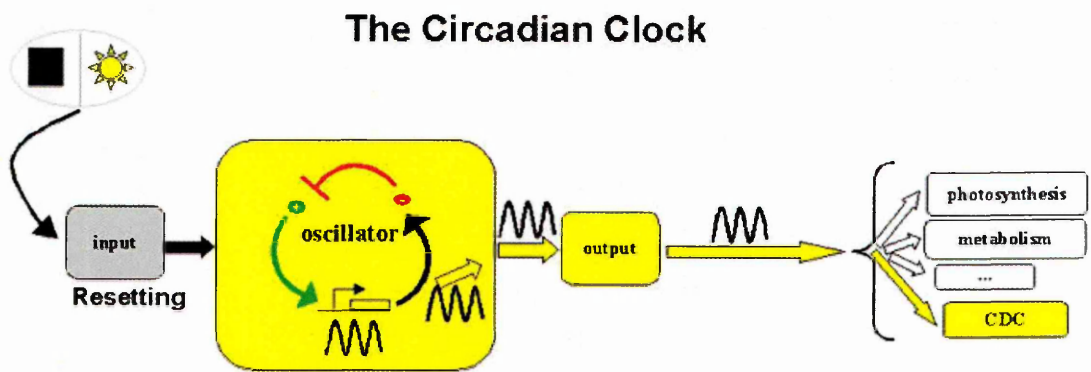
Fig. 7. Circadian system allows an organism to anticipate the dark/light changes in its surrounding environment. The central oscillator maintains an endogenous approximately 24 h rhythm even under constant environmental conditions and this controls a series of overt rhythms within an organism via a series of output pathways. The central oscillator is entrained to the rhythmic cycle via input pathways by which environmental signals are transmitted to the oscillator.

1.3.7 Animal Cryptochromes and the Circadian Clock

The first evidence of a cryptochrome-like protein in animals is derived from the discovery of a human gene encoding a protein related to the *Drosophila* (6–4) photolyase (Todo T. *et al.*, 1996). Two putative blue-light photoreceptors, denoted *hCRY1* and *hCRY2*, were eventually identified in humans (Hsu D.S. *et al.*, 1996). However, knowledge of cryptochrome function in animals derived from studies in *Drosophila melanogaster*.

1.3.7.1 Drosophila *cryptochrome*

Drosophila cryptochrome (dCRY) is a predominantly nuclear protein that acts as a photoreceptor for the entrainment of the circadian clock by physical interaction with central oscillator components. The central oscillator components of *Drosophila* include PER (period), TIM (timeless), CLK (clock) and CYC (cycle; called also BMAL or MOP3) (Dunlap J.C., 1999). CLK/CYC and TIM/PER are positive and negative regulators, respectively, for the transcription of clock genes such as *PER, TIM* and *CRY* (Fig. 8A). CLK and CYC are basic helix-loop-helix proteins that act together to activate transcription through binding to "E box" promoter elements of clock-regulated genes. The E-box (Enhancer-box: CACGTG) is the key component of the circadian clock because it seems to represent a specific binding site for positive regulators.

Moreover the transcription of clock genes is negatively controlled by their own gene products. PER and TIM form heterodimers in the cytosol and then enter the nucleus to suppress their own transcription. It was found that cryptochrome interacts with TIM in a light-dependent manner. The carboxy-terminal domain of *Drosophila* cry is important for protein stability, interaction with TIM, and sensitivity of the photoreceptor to circadian light signals (Busza A. *et al.*, 2004). The

dCRY-TIM interaction results in sequestration of TIM, which promotes ubiquitination and proteasome-dependent degradation of TIM with the consequent suppression of TIM-dependent inhibition of transcription (Ceriani M.F. *et al.*, 1999) (see scheme in Fig. 8A).

A *Drosophila* cry mutant, *cry^b^*, was identified on the basis of its defect in regulating the circadian rhythm of activity of the *PER* promoter (Emery P., *et al.*, 1998; Stanewsky R., *et al.*, 1998). The *cry^b^* mutant abolished cycling of *PER* and *TIM* expression and abrogated the effect of constant illumination on circadian behaviour. Transgenic flies overexpressing cryptochrome showed increased circadian photosensitivity (Emery P. *et al.*, 2000). However, cryptochrome is apparently not the only photoreceptor that entrains the circadian clock in *Drosophila*, because the *cry^b^* mutant fly still displays an entrainment of behavioural rhythmicity in blue light unless signal transduction for the visual pigment is also eliminated (Stanewsky R. *et al.*, 1998). Therefore appears to be a functional redundancy between cryptochromes and other photoreceptor systems.

1.3.7.2 Mammalian cryptochromes

In mammals the circadian system is more complex due to the duality of the photosensory systems. The input component is represented by a photoreceptor/phototransducer pathway located in the eye. The master circadian clock is localized in the midbrain (hypothalamus) in a cluster of neurons called the suprachiasmatic nucleus (SCN) and the output system consists of neuropeptides released from the master circadian clock (Reppert S.M. & Weaver D.R., 2001). Thus, mammals have two photosensory systems that are divergent at the molecular and anatomical levels: the visual system for 3-D vision and the circadian photosensory system for sensing the fourth dimension, time. The two photosensory

systems function more or less independently. For example, certain retinal degeneration diseases in humans and mice that destroy the outer retina and cause total visual blindness leave the circadian phototransduction system intact (Rönneberg T. & Foster R.G., 1997) and, conversely, mutations in the circadian photosensory pathway that seriously compromise the master circadian clock do not affect the animal's vision (Suter B. *et al.*, 1997).

In the last years several studies have been performed in order to understand the role of mammalian cryptochromes as photoreceptors entraining the circadian clock, but the nature of the receptor responsible for this process is still largely controversial (Lin C. & Todo T., 2005; Sancar A., 2004).

As in humans, two cryptochrome genes have been identified in mouse, denoted *mCRY1* and *mCRY2*. The mCRY1 protein is localized in mitochondria whereas mCRY2 is found predominantly in the nucleus (Kobayashi K. *et al.*, 1998). It was shown that mouse cryptochromes are expressed in many tissues and organs, including the front part of the retina. In particular, the expression of *mCRY2* was relatively high in the retinal ganglion cells (Miyamoto Y. & Sancar A., 1998). Moreover, a subset of these retinal ganglion cells (RGCs), known to be the photosensitive cells required for entrainment of the circadian clock (Berson D.M. *et al.*, 2002), project to the SCN. Consistent with its function, the gene encoding mCRY1 was found to be expressed at a high level in the SCN where its levels oscillate with a 24 h periodicity (Rönneberg T. & Foster R.G., 1997). Thus, cryptochromes appear to be produced in those cells and tissues required for entrainment and functioning of the master clock. As a consequence of this intimate involvement in clock function, animals lacking *mCRY1* have short periods, whereas those lacking *mCRY2* have long periods. Remarkably, the knockout mice missing both *mCRY1* and *mCRY2* retained near-normal behavioural rhythmicity in light/dark

cycles, but showed an instantaneous and complete loss of rhythmicity in free-running conditions (van der Horst G.T. *et al.*, 1999). These observations indicate that the cry proteins play an essential and light-independent function in the mammalian central oscillator, and that cryptochromes are not the only photoreceptors mediating light control of the clock. The fact that cryptochromes are integral parts of the mouse central oscillator makes it almost impossible to test directly their role in light entrainment of the clock.

Nevertheless, several observations demonstrate light-dependent roles of mammalian cry proteins. Knockout mice lacking one or both genes have a reduced or abolished ability to induce expression of genes such as *PER* and the protooncogene *c-fos* in response to light (Selby C.P. *et al.*, 2000). Moreover, the pupils of mutant mice lacking both *CRY* genes have a reduced reflex response to light (Van Gelder R.N. *et al.*, 2003). It has also been found somewhat analogous to the situation in the *Drosophila cry^b* mutant, that the mouse *cry1cry2* double mutant retained its ability to mediate light input unless signal transduction from the visual pigments was also disrupted at the same time. Only triple-mutant mice carrying both cryptochrome mutations and a retinal degenerative mutation were nearly arrhythmic in light/dark cycles, with a marked reduction in light-induced gene expression (Vitaterna M.H. *et al.*, 1999). These results indicate that mammalian cry proteins are indeed involved in regulation of the circadian clock by light, but that their role is carried out redundantly with other photoreceptors.

It should be noted that the proposed photoreceptor function of cry in animals, particularly in mammals, is not universally accepted. Even though there is a consensus that the vitamin A (retinal)-based rhodopsins and color opsins located in rods and cones (outer retina) are not required for circadian photoreception, many investigators believe that another opsin that is expressed in the inner retina is

responsible for circadian photoreception, rather than a cryptochrome. Indeed, a novel opsin called melanopsisn has been discovered in all vertebrates examined to date and this molecule is exclusively expressed in the inner retina in man and mouse (Provencio I. *et al.*, 2000). In summary, it is consistent to define cryptochrome as a primary circadian photoreceptor in mammals, with redundant or complementary photoreception provided by classical opsins in rods and cones and melanopsin and perhaps yet-to-be-discovered minor opsins in the inner retina.

Like *Drosophila* cry, mammalian cryptochromes interact physically with clock proteins, including the promoter-binding transcriptional regulators PER, CLK, and BMAL1 (Brain and Muscle Arnt-like protein) (Fig. 8B). However, differently from *Drosophila* cry, mammalian cryptochrome proteins are components of the *negative-feedback loop* of the circadian clock (Lin C. & Todo T., 2005). The CLK/BMAL1 positive regulator that binds the E-box promoter element of clock genes is inhibited by the PER/CRY heterodimer (Fig. 8B). The physical interaction of cryptochrome with other clock components can affect their activity, interaction, degradation, or nuclear trafficking, and consequently they also alter the transcriptional regulation of the clock genes (Reppert S.M. & Weaver D.R., 2002).

In yeast two-hybrid assays it has been found that human cryptochromes bind to human PER1, PER2, CLOCK, and TIM proteins independently of light (Griffin E.A. Jr *et al.*, 1999). However, these light-independent interactions suggest the involvement of a mechanism of photo-entrainment for the circadian clock, different from that found in *Drosophila*. In addition to the direct interaction with the promoter-binding transcription regulators, cryptochromes may also affect the circadian clock by participating in the regulation of histone modifications (Naruse Y. *et al.*, 2004) but how this process works remains to be elucidated.
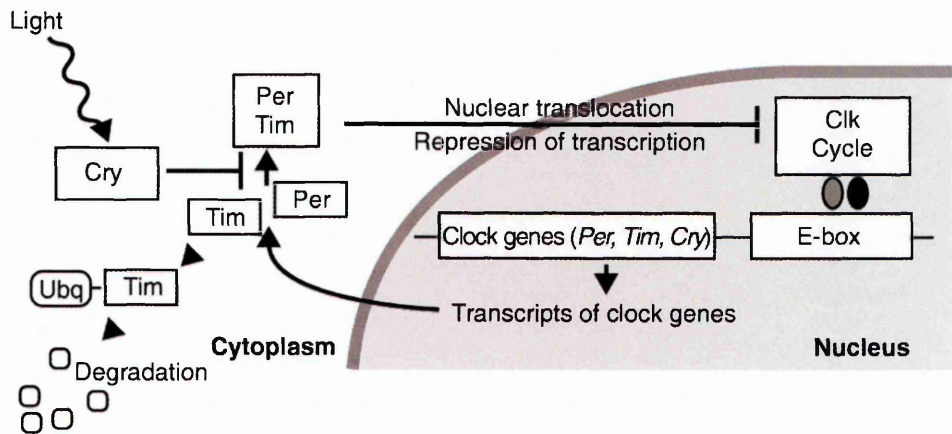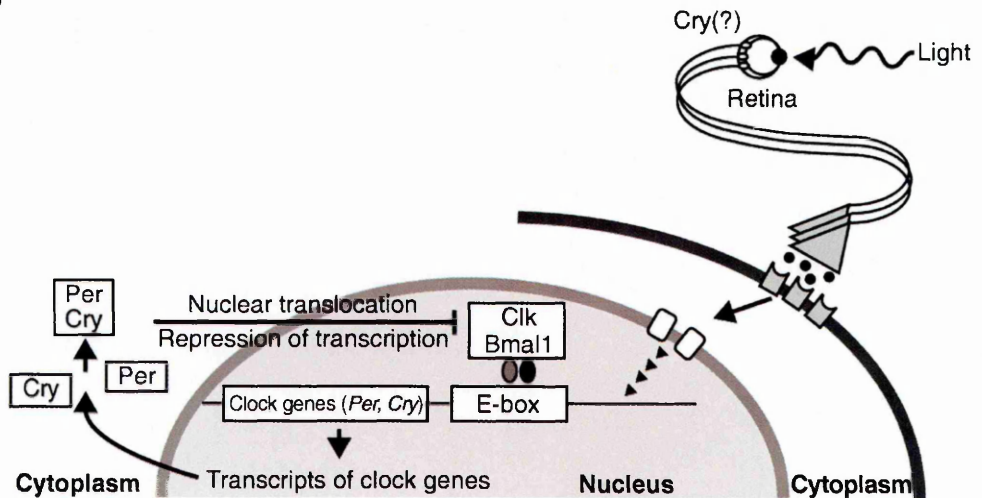
Fig. 8. Regulation of the circadian clock by animal cryptochrome. **A.** In *Drosophila*, Cry binds Tim in a light-dependent manner with the consequent proteasome-dependent ubiquitin-mediated degradation of Tim (Ubq, ubiquitination) and thus inhibition of the action of the Per-Tim heterodimer. **B.** In mammals, cryptochromes are integral parts of the negative feedback loop. The Cry protein interacts with Per to repress the activity of the transcription factors Clk and Bmal1 and thus to repress transcription. Cryptochromes may also be involved in the photo-entrainment of the mammalian circadian clock, but it is not yet clear whether this involves cryptochrome (Lin C. & Todo T., 2005).

1.3.8 Function of Plant Cryptochromes

Cryptochromes have been identified in many photosynthetic organisms living in terrestrial and marine environments. The first member was identified in *A. thaliana* (Ahmad M. & Cashmore A.R., 1993) and in the same year the white mustard (*Sinapis alba*) PHR gene (*SaPHR*), which was initially thought to encode a DNA photolyase (Batschauer A., 1993), was later found to be a cryptochrome (Malhotra K. *et al.*, 1995). Subsequently crys have been found in other dicots such as tomato (Ninu L. *et al.*, 1999), monocots (rice, barley, etc...), ferns (*Adiantum capillus-veneris*), mosses (*Physcomitrella patens*), and green algae (*Chlamydomonas reinhardtii*) (Lin C., 2000).

As previously described, most plant cryptochromes contain a C-terminal extension in addition to the N-terminal PHR domain. However, the C-terminal domains of different plant crys vary significantly in length, from ~380 amino acids in *Chlamydomonas*, ~ 190 amino acids and ~ 120 amino acids in *Arabidopsis* CRY1 and CRY2, respectively, to almost no C-terminal extension in the SaPHR and cryptochrome 5 of *Adiantum capillus-veneris* (AcCRY5). The main feature of the plant cry C-terminal extension is the DAS domain consisting of three recognizable motifs: (a) DQXVP in the amino end of the CCT, (b) a region containing a short stretch of acid residues (E or D), and (c) STAES and GGXVP at the carboxyl end separated by a short non-conserved spacer. The presence of the DAS domain in crys from moss to angiosperm suggests that the evolutionary history of crys in plants is likely to span over 400 million years, dating back to before the widespread dispersal of vascular plants on the Earth (Kenrich P. & Crane P.R., 1997). Genetic studies indicate that the DAS domain is important for cellular localization, intermolecular interaction, and physiological functions of cryptochromes (see later). Plant cryptochromes have a clear role as blue light photoreceptors. They act concurrently

with phytochromes to mediate photomorphogenic responses such as inhibition of stem elongation, stimulation of leaf expansion, control of photoperiodic flowering, entrainment of the circadian clock, and regulation of gene expression. The best characterized genes are *Arabidopsis CRY1* and *CRY2*.

1.3.8.1 Arabidopsis *cryptochromes*

Arabidopsis CRY1 and CRY2 are both nuclear proteins but they exhibit different light-dependent localizations. CRY2 is more or less constitutively localized in the nucleus, whereas it was found that a GUS:CCT1 (CRY1 C-terminus) fusion protein was mostly located in the nucleus in root hair cells of dark-grown transgenic plants, but was mostly cytosolic in light-grown transgenic plants (Yang H.Q. *et al.*, 2001). A bipartite nuclear localization signal was found within the DAS domain of CRY2, although no apparent bipartite NLS is found in CRY1. Notwithstanding, the C-terminal extension has been shown to be sufficient for nuclear/cytoplasmic trafficking of CRY1. The highest *CRY2* expression is observed in the root, where *CRY1* mRNA is absent. On the contrary, *CRY1* is highly expressed in the aerial parts of the plant (Chen M. *et al.*, 2004).

The *Arabidopsis* CRY2 protein undergoes rapid blue light-induced degradation (Guo H. *et al.*, 1999). Both the PHR domain and the C-terminal extension appear to be important for the blue light-induced degradation of CRY2. It is not clear whether ubiquitination is involved in CRY2 degradation, although application of proteasome inhibitors to *Arabidopsis* seedlings suppresses blue light-dependent CRY2 degradation *in vivo* (Lin C. & Shalitin D., 2003). It is important to point out that *Drosophila* cryptochrome has also been reported to undergo a light-induced and proteasome-dependent degradation mediated by a light-dependent conformational change in the molecule (Lin F.J. *et al.*, 2001). Moreover Shalitin D.

et al. (2002) have shown that CRY2 undergoes blue light-dependent phosphorylation. Phosphorylation of the CCT domain has been observed only under blue light, whereas in the dark the protein is mostly present in the unphosphorylated form. Whether this phosphorylation is required for signaling or simply modulates the activity or the level of CRY2 has not been established. In addition, it has been shown that CRY1 binds ATP (Bouly J.P. *et al.*, 2003), and the site of this binding has been determined in the crystal structure reported by Brautigam C.A. *et al.* (2004). It is not known whether animal cryptochromes also bind to ATP, although it has been shown that mouse crys are phosphorylated (Eide E.J. *et al.*, 2002).

1.3.8.2 Signaling Mechanism

Characterization of cryptochromes in *Arabidopsis* and other species has revealed at least two mechanisms by which crys may affect nuclear gene expression changes in response to light.

First, a cryptochrome molecule may interact with proteins associated with the transcriptional machinery to affect transcription directly. A CRY2 C-terminal extension fused to GFP has been shown to be chromatin associated (Cutler S.R. *et al.*, 2000), but it is unclear whether this interaction has a role in regulating gene expression. Furthermore, unlike the animal crys that have been shown to physically bind to promoter-binding transcriptional regulators, no such interaction has been reported for plant cryptochromes. An alternative model is that plant crys may interact with proteins exerting other cellular functions to regulate the stability, modification, or cellular trafficking of transcriptional regulators. For example, plant crys have been found to modulate protein degradation through interaction with the E3 ubiquitin ligase, Constitutive Photomorphogenesis 1 (COP1), (Yang H.Q. *et al.*, 2001), a protein required for the degradation of several factors involved in light-

regulated transcription (Jiao Y. *et al.*, 2007). It is well known that in the dark COP1 is associated with transcription factors that promote photomorphogenesis such as HY5 (Osterlund M.T. *et al.*, 2000), and with the COP10/CSN complex, which promotes degradation of these factors by targeting them to the proteasome (Li L. & Deng X.W., 2003). The interaction of *Arabidopsis* crys with COP1 in the light leads to inactivate the latter, resulting in the release of repression of gene expression in response to light. Consistent with this model, it has also been found that *Arabidopsis* crys mediate suppression by blue light of the proteasome-dependent degradation of an important floral regulator, CONSTANS (Valverde F. *et al.*, 2004). Other cry signaling intermediates have also been described. For example, HFR1 (long Hypocotyl in Far-Red 1) is a basic helix-loop-helix transcription factor required for both phytochrome A-mediated far-red and cryptochrome 1-mediated blue light signaling (Duek P.D. & Fankhauser C., 2003). HFR1 physically interacts with COP1, which degrades the transcription factor to desensitize light signaling (Yang J. *et al.*, 2005). Another component of cryptochrome signaling is the negative regulator of photomorphogenesis SUB1 (Short Under Blue Light 1), which is a calcium binding protein. It has been suggested that SUB1 plays an important role in photomorphogenesis responses resulting from light-induced changes in ion homeostasis (Guo H. *et al.*, 2001). Cryptochromes suppress the activity of SUB1 to induce light responses, with the consequent accumulation of the HY5 protein. Genetic studies have shown that SUB1 also acts as a modulator of phytochrome A signaling (Guo H. *et al.*, 2001). Interestingly, a novel positive regulator involved in cryptochrome action has also been identified in plants: the Ser/Thr Protein Phosphate PP7. This protein is suggested to act in concert with cry in the nucleus to activate blue light signaling (Møller S.G. *et al.*, 2003).

The functional interaction observed between cryptochrome and phytochrome (Más P. *et al.*, 2000) is also very intriguing, especially because several light responses are mediated by the coordinated action of these photoreceptors (e.g., seedling establishment, entrainment of the circadian clock, and flowering), although the molecular mechanisms are not known.

The catalytic mechanism of cryptochrome action has not been fully elucidated, but some clues can be found in the mechanism of CPD photolyases, based on their similarity. The PHR region that contains the chromophores is the most conserved part of the proteins and it has been shown to be required for the homodimerization of *Arabidopsis* CRY1, which is essential for photoreceptor activity (Sang Y. *et al.*, 2005). However the carboxy-terminal domain has also been shown to have a role in the function and regulation of both animal and plant cryptochromes. Expression of the C-terminal domains of *Arabidopsis* crys, fused to the marker enzyme GUS confers a constitutive de-etiolated response in darkness (Yang H.Q. *et al.*, 2000).

Possible models of the structural changes of plant cryptochromes in response to blue light consider phosphorylation as a key step for the regulation of binding of these molecules with several signaling partners. In plant cells, it is possible to imagine different scenarios, schematically represented in Fig. 9. The first model (a) considers the phosphorylation of the carboxy-terminal domain in response to light as the crucial step for the dissociation of the two domains that allows interaction with downstream partners. A second possible model is that phosphotransfer mediates the interaction of two cryptochromes encoded by the same gene (b) or, alternatively, that intermolecular phosphotransfer involves the interaction of different cryptochromes and leads to the formation of hetero-oligomers (c). Elucidation of the structure of

holocryptochromes, including the carboxy-terminal domain, may help to understand their mechanism of action.
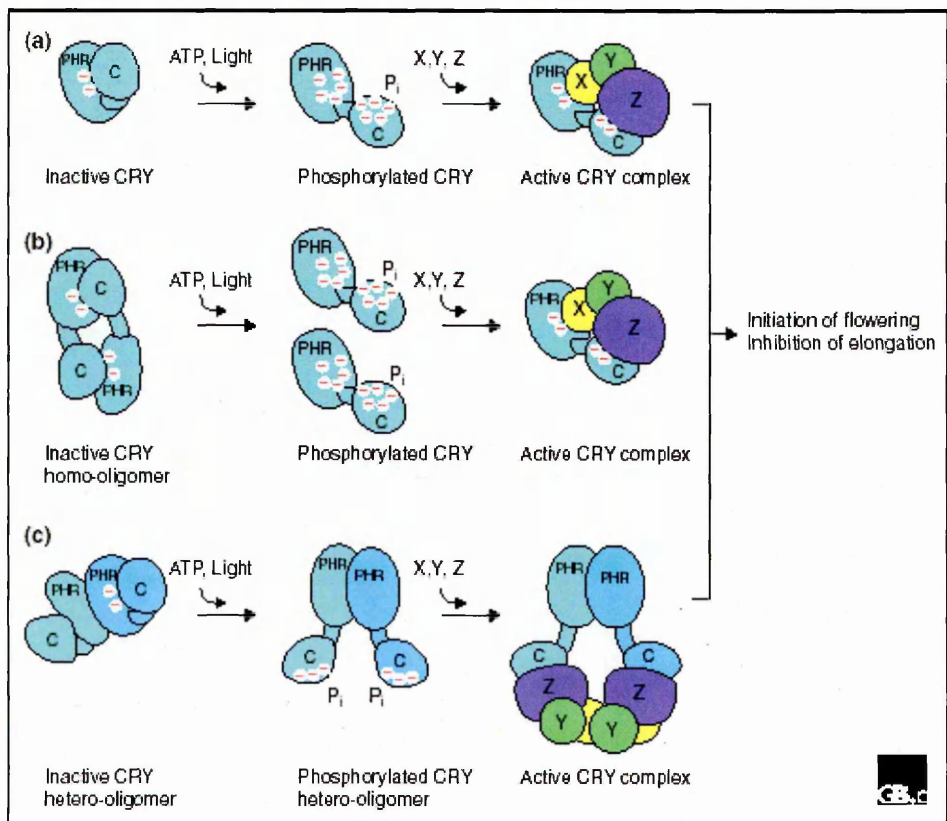
Fig. 9. Possible models of phosphorylation-dependent structural changes of plant cryptochromes. (a) Blue light induces the phosphorylation of the C-terminal domain leading to its dissociation from the N-terminal domain; (b) The CRY homo-oligomer is activated by phosphorylation; (c) Intermolecular phosphotransfer could involve the interaction between different cryptochromes (Lin C. & Todo T., 2005).

1.3.9 Other Flavin-Containing Blue Light Receptors

Flavin is the most common chromophore in nature for blue light perception. Several receptors that use flavin derivatives have been described in bacteria, algae, plants, and fungi but not in animal systems.

The phototropins (phot) were originally identified as a class of blue-light receptors involved in phototropism, the adaptive process that allows plants to grow in the direction of light to maximize light capture for photosynthesis (Huala E. *et al.*, 1997).

Phototropins are composed of two distinct domains: an amino-terminal domain for light sensing and a carboxy-terminal Ser/Thr protein kinase domain. In the sensing domain it is possible to distinguish two repeats of approximately 110 amino acids, denoted LOV1 and LOV2 (Briggs W.R. *et al.*, 2007). The LOV sensor domains are referred to as Light, Oxygen, Voltage and share sequence homology to the PAS domain superfamily. The term PAS comes from the first letter of each of the three founding members of the family: PER, ARNT, SIM. The Period (PER) protein was discovered in *Drosophila* as a result of its involvement in the regulation of circadian rhythms (Reddy P. *et al.*, 1986); ARNT is the Nuclear Translocator of the Aryl Hydrocarbon Receptor (AHR), identified in humans as an essential factor for normal signal transduction by AHR (Hoffman E.C. *et al.*, 1991) and SIM is the product of the *Drosophila Single–minded* locus identified as a regulator of midline cell lineage (Nambu J.R. *et al.*, 1991). The PAS domain indicates a region of homology of 250-300 amino acids of these three members and contains a pair of highly degenerate 50 amino acid subdomains termed the A and B repeats. The PAS domain is a signature of proteins that play roles in the detection and adaptation to environmental changes (Gu Y.Z. *et al.*, 2000).

Phototropin receptors are the only members of the PAS superfamily that contain two LOV domains. It is now well known that the two *Arabidopsis* phototropins (*PHOT1* and *PHOT2*) are also involved in the control of several processes, such as chloroplast relocation and stomatal opening (Briggs W.R. and Christie J.M., 2002; Wada M. *et al.*, 2003). Given the knowledge that phototropins and crys utilize flavins as primary chromophores, we might expect these two classes of blue light receptors to exhibit conserved photochemical properties. However, the photochemical properties are quite different, related to the fact that crys use FAD as primary chromophore, whereas phototropins use flavin mononucleotide (FMN) (Liscum E. *et al.*, 2003). Moreover the amino acid sequences of their flavin-binding domains are also distinct. These differences contribute dramatically to the production of two very different holoprotein structures that impart different photochemical properties to the two classes of chromoproteins (Sancar A., 2003).

It has been demonstrated that light activation of phototropins involves a photocycle in the LOV domains (Liscum E. *et al.*, 2003). In darkness, each phot LOV domain binds, non-covalently, one molecule of ground-state FMN. Absorption of blue light triggers a covalent binding between FMN and the Cys within the conserved GRNCRFLQ amino acid motif of the LOV domain (the cysteinyl-FMN adduct) with the subsequent activation of the protein kinase domain. This reaction is completely reversible in the dark. In addition, it is clear that the kinase activity is necessary for phototropin-mediated signal transduction and that the first event of this process involves an autophosphorylation of Ser residues (Salomon M. *et al.*, 2003). Importantly, the formation of this cysteinyl-flavin adduct in LOV2 is absolutely essential for the phototropic function of *PHOT1* in *Arabidopsis* seedlings (Christie J.M. *et al.*, 2002). Curiously, *PHOT1* function does not appear to require adduct formation in the LOV1 domain. Salomon M. *et al.* (2004) identified LOV1 as a

dimerization site, although there is no evidence that light had any effect on dimerization. To date, the exact role of LOV1 remains unknown.

At present, the only substrate known for phot kinase activity is the phot itself, thus the signal transduction pathway that links this photoreceptor to downstream physiological responses remains largely unknown. The kinase domain of *Arabidopsis PHOT2*, expressed in bacteria, can phosphorylate the artificial substrate casein *in vitro*. The substrate phosphorylation occurs constitutively but becomes light dependent upon the addition of purified LOV2 (Matsouka D. & Tokutomi S., 2005).

However, the regulation of ion channel activity by light has been proposed to play a central role in the general process of photomorphogenesis and, more specifically, in the responses mediated by phot (Spalding E.P., 2000). Baum G. *et al.* (1999) have shown a blue light-dependent transient increase in cytoplasmatic Ca^{2+} in wild-type seedlings that is dramatically attenuated in an *Arabidopsis phot1* mutant.

Furthermore, one possible phototropin substrate may be NPH3, a plasma membrane protein that possesses BTB/POZ and coiled-coil protein-protein interaction domains. It has been demonstrated that NPH3 interacts with PHOT1 in yeast two hybrid experiments and pull-down assays (Motchoulski A. & Liscum E., 1999). Another protein that seems to have a function in phototropin signaling is RPT2. This protein is similar to NPH3 because it contains a BTB/POZ domain at the N-terminus and a coiled-coil domain at the C-terminus and has been shown to interact with PHOT1 (Inada S. *et al.,* 2004). A summary of phototropism mechanism is extensively described by Christie J.M. (2007).

In *Arabidopsis*, other proteins that contain a LOV domain and that display similar light-induced photochemistry as the phototropins have been recently characterized. Proteins such as ZTL (Zeitlupe), FKF1 (Flavin-binding kelch repeat

F-box1), and LKP2 (LOV kelch Protein 2) share, in addition to the LOV domain, an F-box region that might target proteins for degradation, as well as six terminal kelch repeats, which most probably mediate specific protein-protein interactions. Bacterially expressed LOV domains derived from each of these proteins exhibit photochemical properties analogous to those of the phototropin LOV domain (Imaizumi T. *et al.*, 2003; Nakasako M. *et al.*, 2005). So, ZTL, FKF1 and/or LKP2 have been proposed as novel classes of circadian photoreceptors that regulate flowering by targeting clock components for degradation in a light-dependent manner (Yanovsky M.J. & Kay S.A., 2003).

Phototropins have also been identified in *C. reinhardtii*, and a completely novel function for phototropin has been demonstrated in this organism. In this alga, the blue light receptor has a critical role in controlling the blue light-dependent progression of the sexual life cycle, including pregamete to gamete conversion, up-regulation of gene expression during gametogenesis, maintenance of mating competence and zygote germination (Huang K. & Beck C.F., 2003), and changes in chemotaxis during the initial phase of the sexual cycle (Ermilova E.V. *et al.*, 2004). It is interesting to note that this developmental role for phot is in contrast with what is found in higher plants, where the control of developmental responses is mainly associated with cry function (Huang K. & Beck C.F., 2003).

Another example of a peculiar blue light receptor is a small protein (186 amino acids) named VIVID, characterized in the ascomycete *Neurospora crassa* (Schwerdtfeger C. & Linden H., 2003). This protein contains a LOV domain and seems to use either FAD or FMN as chromophore forming a covalent flavin-cysteinyl adduct under blue light. VIVID enables *Neurospora* to perceive and respond to the daily changes in light intensity, likely by protein-protein interactions.

As a demonstration of the complexity of blue light photoperception in different organisms, a novel FAD-binding domain involved in sensory transduction has been identified in microorganisms, and is denoted BLUF, for "sensors of blue-light using FAD" (Gomelsky M. & Klug G., 2002). The N-terminal region of the AppA protein of the phototrophic proteobacterium *Rhodobacter sphaeroides* contains the BLUF domain and is involved in the blue light regulation of expression of photosynthesis gene clusters (Masuda S. & Bauer C.E., 2002). Most of these proteins are from two branches of Bacteria, Proteobacteria and Cyanobacteria. Bacterial genomes contain up to three BLUF domains per genome. No BLUF domains are encoded by the currently available genomes of archaea. A member of this family has been identified in *Euglena gracilis*, a unicellular flagellate that changes its swimming direction in response to changes in blue light intensity (Iseki M. *et al.*, 2002). The receptor responsible for this photophobic response is an adenylyl cyclase containing the flavin adenine dinucleotide chromophore, whose activity is regulated by blue light. Considering all this evidence together it is highly likely that other blue light photoreceptors with related structures will be described in the future.

The research project has focused mainly on two different aspects: the development of molecular tools in diatoms using *Phaeodactylum tricornutum* as model species and the structural and functional characterization of a gene denoted *PtCPF1* (Cryptochrome/Photolyase Family 1) that encodes a putative photoreceptor in the diatom *P. tricornutum*.

The first important goal for the development of molecular tools in *P. tricornutum* has been to determine the sequence of the whole genome providing an incredible amount of new information for comparative genomics analyses. In addition, a useful cloning system for the expression of diatom genes has been established. Based on the Gateway Technology (Walhout A.J. *et al.*, 2000), diatom Destination vectors containing useful epitopes have been constructed offering a variety of specific purposes such as gene overexpression, protein localization, and tagged proteins for immunopurification, immunolocalization and fluorescent protein detection (Siaut M. *et al.*, 2007).

In order to characterize the diatom *PtCPF1* gene, expression at both transcriptional and translational levels have been performed in time course experiments designed to study circadian rhythmicity and acute light induction responses. From this analysis *PtCPF1* was shown to be diurnally expressed and strongly induced under blue light. In order to characterize the biochemical properties of the gene product, the protein was expressed and purified in *E. coli*. Spectral and biochemical analyses showed that PtCPF1 is a blue-light-absorbing protein with DNA repair activity. On the other hand, localization studies in diatom cells have evidenced the constitutive nuclear localization of the protein.

Interestingly, comparative analysis of the diatom PtCPF1 protein has revealed it to be more similar to the animal cryptochromes than to plant counterparts. Since animal crys act as components of the circadian clock controlling daily physiological and behavioural rhythms and as photoreceptors that mediate entrainment of the circadian clock to light, it was important to elucidate the function of the PtCPF1 protein both in a heterologous system and in an *in vivo* system. Remarkably, transcription assays developed in mammalian cells have evidenced a repressor activity of the PtCPF1 protein within the clock machinery, mimicking the function of animal crys. Furthermore, gene expression studies of transgenic diatom lines overexpressing *PtCPF1* have indicated that the protein acts as a blue light photoreceptor because it can modulate several blue light-dependent responses.

Therefore, this research project has identified a novel protein that displays both blue light photoreceptor activity as well as DNA repair activity. In other words, this protein could be considered the missing link in the evolutionary history of the Cryptochrome/Photolyase family.

CHAPTER II - MATERIALS AND METHODS

2.1 Cell culture conditions for *Phaeodactylum tricornutum* (*Pt*)

Axenic *Phaeodactylum tricornutum* cells were obtained from the Provasoli-Guillard National Center for Culture of Marine Phytoplankton (strain CCMP632, originally isolated by Coughlan in 1956 off the coast of Blackpool, U.K.). In our laboratory this strain is known as Pt1 (De Martino A. *et al.*, 2007). Cultures were normally maintained at a temperature of 18 °C in a 12 hour photoperiod with white light at an intensity of approximately 180 $\mu mol.m^{-2}.s^{-1}$ in f/2 medium (Guillard R.R. L., 1975).

2.2 Contamination test

The presence of bacterial and fungal contaminants in diatom cultures was checked periodically by a contamination test. An aliquot (around 1 mL) of diatom cells was taken and grown in the dark in sterile glass tubes containing 10 mL of f/2 medium enriched with 1 g/L peptone. The presence of bacteria or fungi was verified after approximately one week.

2.3 Extraction of high molecular weight DNA for genome sequencing

A monoclonal culture derived from Pt1 of fusiform cells (known as Pt1 8.6 and deposited as CCMP2561 in the Provasoli-Guillard National Center for Culture of Marine Phytoplankton) was grown at 18 °C in a 12 hour photoperiod at an illumination of approximately 180 $\mu mol.m^{-2}.s^{-1}$. Around 4 liters of diatom culture in exponential phase were centrifuged at 1800 *g* for 15 min at 4 °C. The cell pellet (*wet weight* around 20 g) was frozen in liquid nitrogen and resuspended in 40 mL of lysis buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0, 1% SDS, 10 mM DTT, 10 $\mu g/mL$ of proteinase K) and incubated at 50 °C for 45 minutes. Three classical phenol/chloroform extractions were performed to remove proteins, and a subsequent

extraction with chloroform isoamyl alcohol (24:1) was made to eliminate completely the phenol residues. Genomic DNA was precipitated with 1/10 volume of 3 M sodium acetate pH 5.6 and two volumes of absolute ethanol at -20 °C overnight. After two washes with 70% ethanol, genomic DNA was dried at RT (room temperature) and resuspended in TE 1x (10 mM Tris pH 8.0, 1 mM EDTA). DNA concentration was determined in a spectrophotometer at 260 nm and checked on agarose gels. After RNAse treatment at 37 °C for 30 minutes, genomic DNA was purified on cesium chloride gradients at 55,000 rpm, 20 °C for 18 hours using the vertical rotor VTi 65.2 (Beckman, USA). The genomic DNA band was removed using a 9 gauge needle. To remove the ethidium bromide from the DNA, several extractions with two volumes of isoamyl alcohol saturated with water were carried out. After dialysis against 1x TE, the DNA was precipitated overnight at -20 °C, washed with 70% ethanol, and resuspended in TE. DNA concentration was determined in a spectrophotometer at 260 nm and checked on a 0.8% agarose gel.

2.4 Construction of C-terminal Gateway Destination vectors for diatoms

To construct Gateway-compatible vectors for diatoms, pKS-FcpBpAt plasmid was created for use as parent vector in subsequent cloning steps. The 440 base pair (bp) promoter region of the *FCPB* gene (FcpBp) and the 220-bp terminator sequence of the *FCPA* gene (FcpAt) were amplified from the pFCPBp-Sh ble vector (Bhaya & Grossman, 1993; Falciatore A. *et al.*, 1999) by PCR using, respectively, the primer couple 5-FcpBp-SacII: 5'-AGT<u>CCGCGG</u>AATCTCGCCTATTCATGGTG-3' (*Sac*II site underlined) and 3-FcpBp-Not: 5'-CAT<u>GCGGCCGC</u>TGGCAACCGTGAAATATGC-3' (*Not*I site underlined) for the promoter, and 5-FcpAt-Eco-Xho: 5'-GTA<u>GAATTC</u>TCGAGCTACCTCGACTTTGGCT-3' (*Eco*RI site underlined) and

3-FcpAt-Kpn: 5'-CGA<u>GGTACC</u>TGAAGACGAGCTAGTGTT-3' (*Kpn*I site underlined) for the terminator. The resulting FcpBp and FcpAt PCR products were cloned, respectively, into the *Sac*II-*Not*I sites and the *Eco*RI-*Kpn*I sites of pBluescript KS + (Stratagene) to generate pKS-FcpBpAt.

The EYFP tag was amplified from vector pEYFP-N1 (Clontech) using as forward primer EyfpRVSma: 5'-TCC<u>CCCGGGG</u>GAGATATCC*ATG*GTGAGCAAGGGCGAG-3' (*Sma*I site underlined, start codon in italic) and as reverse primer EyfpBamRI: 5'-ATTGGATCCGC<u>GAATTC</u>*TTA*CTTGTACAGCTCGTCC- 3' (*Eco*RI site underlined; stop codon in italic). The resulting 734 bp PCR product was cloned in pKS-FcpBpAt into the *Sma*I-*Eco*RI sites to obtain the intermediate vector denoted pKS-FcpBpAt-C-eyfp. It should be noted that in the selected clone pKS-FcpBpAt-C-eyfp the *Sma*I site was lost during transformation in *E. coli* cells, as verified by sequencing.

In order to generate the final C-terminal EYFP Gateway vector (pDEST-C-EYFP), the pKS-FcpBpAt-C-eyfp vector was digested by *Eco*RV to insert the Gateway cloning cassette, the Reading Frame Cassette A (RfA- Invitrogen, Carlsbad, USA), following the Invitrogen protocol.

To build the Gateway vector containing the HA tag (HemAglutinin Influenza virus epitope), 3 repeats of HA (3HA) were amplified from pKS-3HA vector (Vaistij F.E. *et al.*, 2000) using, as forward primer, HARVSma: 5'-TCC<u>CCCGGGG</u>GAGATATCCCGATACCCCTACGACG-3' (*Sma*I site underlined) and as reverse primer HABamRI: 5'-ATTGGATCCGC<u>GAATTC</u>*TTA*AGCGGCGTAGTCGGGC-3' (*Eco*RI underlined; stop codon in italic). The obtained 119-bp PCR product was cloned in pKS-FcpBpAt into *Sma*I-*Eco*RI sites to obtain the intermediate vector named pKS-FcpBpAt-C-

3HA. A second cloning step was to digest the pKS-FcpBpAt-3HA by *Eco*RV to insert the blunt fragment RfA in order to generate the final C-terminal HA Gateway vector (pDEST-C-HA).

It is important to underline that the RfA cassette contains the F plasmid-encoded *ccdB* gene, that inhibits growth of *E. coli* (Miki T. *et al.*, 1992; Bernard P. & Couturier M., 1992) so it is necessary to use the *ccdB*-resistant *E. coli* strain DB3.1 (chemically competent cells – Invitrogen) for the propagation and maintenance of pDEST vectors.

2.5 Identification and cloning of *PtCPF1* gene

In order to identify members of the Cryptochrome/Photolyase gene family in diatoms, a bioinformatic analysis was first performed searching by BLAST (Basic Local Alignment Search Tool) for cryptochrome-like sequences in a *Phaeodactylum tricornutum* EST collection (Expressed Sequence Tag) generated by our laboratory (Scala S. *et al.*, 2002). Using specific primers on diatom cDNA, a 200 bp fragment was amplified and subsequently cloned in TOPO vector (Invitrogen). This fragment was used as probe to screen a cDNA library using standard procedures (Sambrook J. *et al.*, 1989), which allowed the isolation of a clone containing the start codon of the mRNA but lacking the C-terminal extremity.

The full length mRNA was obtained by 3'RACE (3' Rapid Amplification of cDNA Ends) using two forward internal primers (C2FW: 5'-CCAAGTATATCTACGAACCT-3', C3FW:5'-GGTGATTGTCGGTGAAAACT-3') and the universal oligonucleotide complementary to the 3' extremity of mRNAs (AUAP primer, Abridged Universal Amplification Primer). Instructions of the kit from Life Technologies were followed.

To amplify the full length mRNA from diatom cDNA and the gene from genomic DNA the external primers 5'CFW: 5'-GGATCC*ATG*GCTAAATCGGAAGAG -3' (*Bam*HI site underlined; start codon in italics) and 3'CRW: 5'- GAATTCTG*TTA*GTTGCGACG -3' (*Eco*RI site underlined; stop codon in italics) were designed. The nucleotide sequence was confirmed by sequence analysis on two independent amplifications.

2.6 Phylogenetic analysis

To investigate the evolutionary relationships of the Cryptochrome/Photolyase Family members, representative sequences were selected and a multiple alignment was constructed with the alignment software CLUSTAL W (Thompson J.D. *et al.*, 1994). The alignment of a conserved core of around 400 amino acids was used for the phylogenetic analysis by the Neighbour Joining (NJ) method (Saitou N. & Nei M., 1987). The genetic distance between every pair of aligned sequences was calculated using the JTT model (Jones D.T. *et al.*, 1992) for the amino acid substitutions. The statistical significance of the tree topology was evaluated by a bootstrap analysis (Felsenstein J., 1985) with one thousand iterative tree re-constructions.

2.7 Construction of pGEX-2TK-PtCPF1 vector

In order to express in *E. coli* the diatom PtCPF1 protein fused to GST (Glutathione S-transferase), the 5'CFW and 3'CRW oligonucleotides described previously were used to amplify the PtCPF1 coding region and to clone it in pGEX-2TK vector (Amersham Pharmacia Biotech) into *Bam*HI-*Eco*RI sites, creating pGEX-2TK-PtCPF1. Sequence analysis confirmed the nucleotide region corresponding to the amino terminal fusion.

2.8 Expression in *E. coli* and purification of GST fusion

In order to express and purify the PtCPF1 protein it was necessary to use the photolyase-deficient SY2 *E. coli* strain as host (provided by Prof. Takeshi Todo). pGEX-2TK-PtCPF1 vector was transformed in SY2 strain and different induction conditions were tested in order to generate as much protein in the soluble fraction as possible. Bacterial cells were grown at 37 °C until 0.5 OD before performing the induction at 16 °C for 21 hours using 0.1 mM IPTG. The cell pellet was centrifuged at 1200 g for 15 min at 4 °C and immediately frozen in liquid nitrogen. A freeze and thaw procedure was applied three times to lyse the cells before resuspending the broken cells in 1X PBS containing a complete protease inhibitor cocktail (Roche). The bacteria were then sonicated using three cycles of 30 seconds, with 20 seconds pause in between and the soluble fraction was separated from the pellet by centrifuging at 20000 g at 4 °C for 1 hour. The majority of recombinant GST:PtCPF1 protein was found to be insoluble, although western blotting using the α-PtCPF1 antibody revealed the presence of the diatom protein in the soluble fraction. The purification procedure was performed essentially according to the instructions of the GST Gene Fusion System (Amersham Biosciences). More precisely, the binding to glutathione resin was perfomed in batch, overnight at 4 °C in 1X PBS (protease inhibitors added). After three washes with a large volume of 1X PBS, the resin was transferred to a column to perform the elution step. One bed volume of elution buffer (50 mM Tris-HCl pH 8.0, 10 mM glutathione) was used for each elution which was collected for 15 min at room temperature. This procedure was repeated until 12-15 elution fractions were collected that were subsequently pooled and dialyzed against 50 mM Tris-HCl pH 7.5 to eliminate the glutathione. Finally the recombinant protein was concentrated using Centricon filters (cut off 30 kDa, Amicon) and immediately stored at -80 °C in the final buffer 50 mM Tris-HCl

pH 7.5, 20% glycerol. All steps were checked by SDS-PAGE analysis and the protein concentration was estimated by loading fixed amounts of BSA (bovine serum albumin) protein as reference on acrylammide gel.

2.9 Spectroscopic analysis

The absorption and fluorescence spectra of PtCPF1 were recorded with a BECKMAN DU-640 spectrophotometer and a Shimadzu RF-5300PC spectrofluorometer, respectively, in the laboratory of Prof. Takeshi Todo. To identify the flavin adenine dinucleotide cofactor, the partially purified GST:PtCPF1 protein, together with the GST protein used as negative control, were heated at 95 °C for 5 min in buffer containing 50 mM Tris-HCl pH 7.5. The precipitated proteins were removed by centrifugation and the absorption spectrum was recorded.

2.10 Determination of nucleic acid content

In order to determine the nature of nucleic acids bound to the recombinant purified PtCPF1 protein, the sample was subjected to phenol/chloroform extractions. Nucleic acids, recovered by ethanol precipitation, were treated with either RNase (DNase-free, Roche) or DNase I Amplification Grade (RNase-free, Invitrogen) following the respective instructions, and analyzed by electrophoresis on 2% agarose gels. Diatom ribosomal RNA and a 200 bp fragment of plasmid DNA were used as controls to check the specificity of the enzymes.

2.11 Detection of DNA photolyase activity *in vitro*

A gel shift binding experiment and a DNA repair assay were performed using as substrate a double stranded DNA probe of 49 bp containing a single UV photoproduct (either a CPD, a (6-4)photoproduct or a DEWAR isomer) at the *Mse*I

restriction site (prepared as described in Hitomi K. *et al.*, 1997). The substrate sequence is as follows:

d(AGCTACCATGCCTGCACGAA<u>TT</u>AAGCAATTCGTAATCATGGTCATAGC T), and the thymine dimer of the damaged strand is underlined. The oligonucleotide was labelled with [γ-^{32}P]ATP (3000 Ci/mmol) by T4 polynucleotide kinase and was annealed with the complementary strand by heating at 75 °C for 10 min and cooling to room temperature for 2-3 hours. The labelled duplex DNA, containing a single photoproduct, was purified by agarose gel electrophoresis and used as substrate for both assays.

For the gel shift binding assay, 3 μg of purified recombinant proteins were incubated with 1 nM of ^{32}P-labelled DNA substrate and subsequently analyzed by electrophoresis on a non-denaturing acrylamide gel. The *in vitro* repair assay was done using 37 μg of purified GST fusion proteins mixed with 10 nM of ^{32}P-labelled DNA substrate that were illuminated for 2 hours by daylight fluorescent lamps in 100 mM Tris-HCl pH 8.0, 1 mM DTT buffer. After phenol/chloroform extractions, the DNA substrate was recovered by ethanol precipitation, subjected to *Mse*I digestion and finally analyzed by electrophoresis on a denaturating acrylamide gel.

2.12 Detection of DNA photolyase activity *in vivo*

The SY32 *E. coli* strain has been engineered to test for (6-4)photolyase activity in a functional complementation assay. These bacterial cells are deficient in DNA repair mechanisms, but they carry a plasmid pRT2 (tetracycline resistant) encoding the *E. coli* CPD photolyase gene (*phr*). The complete genotype is *recA*[-] *uvrA*[-] *phr*[+] (Todo T. *et al.*, 1996). Vectors pGEX-4T-2-z(6-4)PHR, encoding *zebrafish* (6-4)photolyase, and pGEX-4T-2-zCRY2a, encoding *zebrafish*

cryptochrome 2a, were used, respectively, as positive and negative controls and were transformed in SY32 cells, in parallel with pGEX-2TK-PtCPF1 vector.

Bacterial cells were grown in Luria-Bertani (LB) medium (Sambrook J. *et al.*, 1989), containing 100 μg/mL ampicillin and 10 μg/mL tetracycline, at 37 °C until 0.5 OD. Cells were then transformed with pGEX-z(6-4)phr and pGEX-zCRY2a vectors and grown at 26 °C for 21 hours with 0.1 mM IPTG, whereas cells transformed with pGEX-2TK-PtCRY were grown at 16 °C for 21 hours in the presence of 0.1 mM IPTG to induce more soluble protein. Bacterial cell pellets were resuspended in 1X PBS to a concentration of 10,000 cells/mL. Cells were then exposed to a UV-C lamp for different times (0, 3, 4, 5, 6, 8, 10 seconds). The subsequent photoreactivation reaction was performed by irradiating the cells with a white fluorescent lamp for 1 h using a 5mm thick glass as a screen to eliminate UV wavelengths contained in the white light. Bacterial cells were spread on ampicillin-tetracycline LB plates and grown at 37 °C overnight in the dark. The next day survival rate was calculated by counting the numbers of colonies.

2.13 Construction of pEYFP:PtCPF1 vector

In order to study the cellular distribution of PtCPF1 in Pt, the pEYFP:PtCPF1 vector was constructed. The EYFP sequence (enhanced yellow-green variant of *Aequorea victoria* green fluorescent protein) was amplified from the pEYFP-N1 vector (Clontech) using the following primers: N5-YFP-Not 5'-ATC<u>GCGGCCGC</u>*ATG*GTGAGCAAGGGCG-3' (*Not*I site underlined; start codon italicized) and N3-YFP-Bam 5'-TAC<u>GGATCCC</u>TTGTACAGCTCGTCCATG-3' (*Bam*HI site underlined). The resulting 720 bp PCR product was cloned in the pKS-FcpBpAt plasmid, previously described, into the *Not*I-*Bam*HI sites to generate the pKS-FcpBpAt-N-EYFP vector. The nucleotide region encoding PtCPF1 was

digested from pGEX-2TK-PtCPF1 (previously described) using *Bam*HI-*Eco*RI and cloned into the *Bam*HI-*Eco*RI sites of pKS-FcpBpAt-N-EYFP in order to generate pEYFP:PtCPF1.

2.14 Transformation of *P. tricornutum* and selection of resistant clones

The protocol to transform Pt has been previously described (Falciatore A. *et al.*, 1999). Approximately 5 x 10^7 cells are spread on agar plates (1% w/v) containing 50% f/2 medium. Plasmid DNA is coated onto M17 tungsten particles of 1.1 μm diameter (BioRad) as described in the manual of the Biolistic PDS-1000/He Particle Delivery System (BioRad) used to bombard the cells.

Agar plates containing diatom cells are positioned at level 2 in the chamber (around 6 cm from the stopping screen) to be bombarded using a pressure of 1,550 psi. After bombardment the diatom cells are maintained for 48 hours in the diatom culture room and are then spread on plates containing 50% f/2 medium and 100 μg/mL phleomycin (InvivoGen, San Diego, CA, USA). The pFCPFp-Sh ble vector (Falciatore A. *et al.*, 1999) was used to confer resistance to phleomycin. Resistant colonies were obtained after three weeks of incubation at 18 °C in a normal 12 hour light-dark photoperiod. Individual resistant colonies were restreaked on 50% f/2 medium supplemented with 80 μg/mL zeocin (InvivoGen) for further analyses.

2.15 Co-transformation of *P. tricornutum* and PCR screening

For overexpression of PtCPF1 another Pt accession was used, denoted Pt8 (CCMP2560) (De Martino A. *et al.*, 2007), because it was more likely to reveal phototactic phenotypes.

Exponential phase Pt8 cells were bombarded with pFCPFp-Sh ble (expressing phleomycin resistance) and pKS-PtCPF1 vectors. Putative

cotransformants, grown under selection, were analyzed for the presence of the gene of interest (*PtCPF1*) by PCR screening using the following protocol. A diatom cell pellet (corresponding to ~10 mL of culture) was resuspended in 50 μL of lysis buffer (1% NP40, 10 mM Tris-HCl pH 7.5, 0.14 M NaCl, 5 mM KCl), placed on ice for 2 min to induce cell lysis, and centrifuged at 20,000 g for 5 min. The cell pellet was then resuspended in 30 μL of distilled water (DW) and 13 μL were used for each PCR reaction in 50 μl final volume containing as final concentrations 1x PCR buffer (Roche), 0.2 mM dNTP (deoxyribonucleotide triphosphate), 0.2 μM of each primer, 0.2 units Taq polymerase (Roche). To check for the presence of the *PtCPF1* gene in transgenic diatoms, the following PCR conditions were used: 95 °C/ 5 min; step 2 at 94 °C/ 1 min; 55 °C/ 1 min; 72 °C/ 1 min returning to step 2 for 34 cycles; 72 °C/ 10 min. PCR products were analyzed on 1% agarose gels.

2.16 Microscope analysis

Cells were observed and photographed using a Zeiss LSM 510 meta laser-scanning confocal fluorescent microscope equipped with a Zeiss plan apochromat 100x/1.40 oil objective, in collaboration with Dr. Fabio Formiggini at the CEINGE Institute (Naples). In order to stain the genomic DNA in *P. tricornutum* cells expressing the fluorescent protein EYFP:PtCPF1, exponentially growing cells were incubated at room temperature for 1 hour in the dark with 200 μg/mL of 4',6'-diamidino-2-phenylindole hydrochloride (DAPI) (Molecular Probes). EYFP and chlorophyll were excited at 514 nm and 543 nm, respectively, whereas DAPI was excited at 405 nm. The emitted fluorescence was detected using filters with a bandwidth of 530-595 nm for EYFP, 420-480 nm for DAPI, and > 560 nm for chorophyll. Images were captured and processed with the Zeiss LSM confocal software.

2.17 Expression in *E. coli* and purification of inclusion bodies

In an attempt to produce an antibody against diatom protein PtCPF1, the pGEX-2TK-PtCPF1 vector was transformed in *E. coli* strain BL21. Cells were grown until OD 1.0 at 37 °C, and then induced at 25 °C for 3 hours using 0.1 mM IPTG (isopropyl-β-D-thiogalactopyranoside). Because the diatom protein was not present in the soluble fraction it was necessary to proceed with purification from inclusion bodies. A bacterial pellet, previously frozen in liquid nitrogen, was thawed in cold water and resuspended in lysis buffer (50 mM Tris-HCl pH 7.5, 5 mM EDTA, 5 mM DTT, 100 mM NaCl) containing 10 mg/mL of lysozyme. After 20 min of lysis at room temperature the pellet was sonicated on ice, treated with 1% Triton X-100 for 30 min at 4 °C and washed three times with a large volume of wash buffer (50 mM Tris-HCl pH 7.5, 5 mM EDTA, 5 mM DTT, 100 mM NaCl, 2% Triton X-100) containing 2 M urea in the second wash. At this point the cell pellet was homogenized in urea buffer (8 M urea, 50 mM Tris-HCl pH 7.5, 5 mM EDTA, 5 mM DTT) and resuspended thoroughly for 1 h at 4 °C to solubilize inclusion bodies. The concentration of GST:PtCPF1 purified protein was measured by BioRad Protein Assay and checked by SDS-PAGE analysis. All buffers used for the purification of inclusion bodies contained 1 mM PMSF (phenylmethylsulfonyl fluoride), 1 μg/mL aprotinin, leupeptin and pepstatin as protease inhibitors. The GST:PtCPF1 protein was sent to the Biopat company (Caserta, Italy) for the generation of antibodies using rabbit as host.

2.18 Antibody Purification

In order to purify the specific antibody from the rabbit serum, 80 μg of diatom recombinant protein (purified from inclusion bodies as described) were resolved on 8% gel by SDS-PAGE and transferred to a nitrocellulose membrane.

The membrane was stained by Ponceau red for 5 min and washed 1 min with distilled water to visualize and to cut out the PtCPF1 band. The membrane was blocked in 1X phosphate buffer saline (PBS), 0.1% Tween 20, 9% milk for 2 h at room temperature before incubating it with rabbit serum in 1X PBS, 0.1% Tween 20, 5% milk overnight at 4 °C (1:5 dilution). After three washes of 20 min each at room temperature with 1X PBS, 0.1% Tween 20, the membrane was treated with 1.5 mL of stripping solution (0.2 M HCl, 0.2% gelatin, pH 2.8 final) for 2 min at room temperature to release the purified antibody. To reach pH 2.8 in the stripping solution 2 M glycine was used in a 0.2% gelatin solution with a final pH of 4.0. Subsequentely, 0.5 mL 1 M Tris base (pH 11) plus 20 µL 20% Triton X-100 were added to neutralize the solution. The purified antibody, denoted α-PtCPF1, was quantified by BioRad Protein Assay and stored at 4 °C.

2.19 Cell culture conditions for time course experiments in *P. tricornutum*

To study the expression of the diatom *PtCPF1* gene, time course experiments were performed, using diatom cells collected every three hours for two days. Cells were either grown in a 12 h photoperiod with white light at approximately 180 $\mu mol.m^{-2}.s^{-1}$, in continuous light, or in continuous dark. For acute light induction time courses the cells were first adapted in the dark for 48 hours and then exposed either to continuous white light (175 $\mu mol.m^{-2}.s^{-1}$), blue light or red light (25 $\mu mol.m^{-2}.s^{-1}$), collecting samples at different times (1, 3, 5, 8 and 12 hours) after exposure. For each time point we collected 10^8 cells grown in exponential phase at a concentration of 2 x 10^6 cells.mL^{-1} and, after washing in 1X PBS, the samples were immediately frozen in liquid nitrogen and then stored at -80 °C.

In order to characterize *PtCPF1* overexpressing lines, a gene expression analysis was performed. Wild-type cells together with two selected *PtCPF1*

overexpressing lines (denoted c4 and c15) were grown with an air bubbling system under a normal light-dark photoperiod until exponential phase. Cells were adapted in darkness for 60 hours and then exposed either to a pulse of blue (0.2; 3.3; 25 μmol.m^{-2}.s^{-1}) or red light (3.3; 25 μmol.m^{-2}.s^{-1}) for 5 minutes. Cultures were subsequently readapted in the dark and aliquots containing 10^8 cells were collected after 15 and 30 minutes. Blue light was generated using lamps with a broad band (380-450 nm), whereas a band width of 620-720 nm with peak intensity of 670 nm was used as red light.

2.20 RNA extraction and cDNA synthesis

RNA extraction was performed using the TriPure Isolation Reagent (Roche Applied Science, IN, USA) according to the manufacturer's instructions. We used 1.5 mL of reagent for approximately 10^8 diatom cells. RNA concentration was determined by spectrophotometry at 260 nm and estimated by agarose gel electrophoresis. Genomic DNA was removed from the RNA sample by DNase treatment by incubating 1 μg of RNA with Amplification Grade DNase I (Invitrogen) and following the suggested protocol. cDNA was generated from the treated RNA by RT-PCR (Reverse Translation-Polymerase Chain Reaction) using random hexamer primers following the protocol of the SuperScript^{TM} First-Strand cDNA Synthesis System (Invitrogen).

2.21 Quantitative Real Time-PCR (qRT-PCR) analyses

For the circadian analysis and acute light response studies in wild-type diatom cells, the conditions used in qRT-PCR reactions were 1 μL of template cDNA, corresponding to around 16 ng, 200 nM forward and reverse primers and FastStart SYBR Green Master mix (Roche). The PCR reactions were performed in

an Opticon Chromo4 MJ Research Thermal Cycler (BioRad, Hercules, CA, USA) in Low-Profile 0.2 mL PCR 8-Tube white Strips (BioRad). The PCR conditions comprised 10 min polymerase activation at 95 °C and 40 cycles at 95 °C for 15 sec and 60 °C for 60 sec. The reaction was ended by 5 minutes of final elongation at 72 °C. Amplicon dissociation curves, i.e., melting curves, were recorded after cycle 40 by heating from 60 °C to 95 °C with a ramp speed of 0.5 °C every second. The results obtained in the Chromo4 Sequence Detector were exported as tab-delimited text files and imported into Microsoft Excel for further analyses. Histone *H4* and 30S ribosomal protein subunit (*RPS*) genes were used as internal controls, because they were considered almost stable as described in Siaut, M. *et al.* (2007). Relative mRNA levels were calculated using the $2^{-\Delta\Delta C_T}$ method where $\Delta\Delta C_T = (C_{T,target} - C_{T,control\ gene})_{Time\ x} - (C_{T,target} - C_{T,control\ gene})_{Time\ 0}$ (Livak K.J. & Schmittgen T.D., 2001). For the time course experiments, the reference point used (time 0) was the average of all the 16 time points collected during a diel cycle, whereas for the light induction experiments, time 0 was the point taken after 48 hours (or 60 hours) of dark and before light exposure. All primers utilized in the qRT-PCR experiments were designed using the Primer3 software program (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi), selecting a primer length of 20-23 nucleotides, a melting temperature of 62 +/- 2 °C and a PCR resulting fragment of 150-180 bp. All primers were tested by RT-PCR to verify single amplification products of the expected size. The primer pair efficiency was determined according to Pfaff M.W., 2001, and the primer pairs with an efficiency > 1.8 were selected for this study. The sequences of each primer are indicated in Table 2.1.

Table 2.1

Gene name	primer name	PCR product (bp)	Primer sequence
H4	Q-H4.FW	151	5'- AGGTCCTTCGCGACAATATC -3'
H4	Q-H4.Rev		5'- ACGGAATCACGAATGACGTT -3'
CPF1	CPF1.FW	176	5'- CCAATTGTTGACCACAAGTTGG -3'
CPF1	CPF1.Rev		5'- CGATTCTTTGCACTTTCTGTTAG -3'
RPS	RPS.FW	166	5'- CGAAGTCAACCAGGAAACCAA -3'
RPS	RPS.Rev		5'- GTGCAAGAGACCGGACATACC -3'
FCPB	FcpB.FW	177	5'- GCCGATATCCCCAATGGATTT -3'
FCPB	FcpB.Rev		5'- CTTGGTCGAAGGAGTCCCATC -3'
PSY	PSY.Fw	150	5'- CACGAAGAGGTGTATTCATGCTG -3'
PSY	PSY.Rev		5'- ACAGCTTTTCCCACTTATCCACA -3'
PDS	PDS.FW	168	5'- TCTCATGAACCAGAAAATGCTCA -3'
PDS	PDS.Rev		5'- AAGACTTCTTCGTTGATGCGTTC -3'
CPF25	CPF25.FW	169	5'- CCAACATGCCCGATACCTTT-3'
CPF25	CPF25.Rev		5'- TCCGTGTACCCCAGCTCTTT-3'
CYC	CYC.FW	157	5'- AAACAGCAACATTCCCAGCAAG -3'
CYC	CYC.Rev		5'- CGCACGCTTCAACCACATAC -3'

Table 2.1. Primer sequences utilized in the qRT-PCR experiments. The length of amplified PCR products is indicated.

2.22 Western blotting

Cell pellets from around 30 mL of diatom cultures in exponential phase were resuspended for 30 minutes in lysis buffer (50 mM Tris-HCl pH 6.8, 2% SDS) at room temperature. Cells were then centrifuged at 4°C for 30 minutes and protein extracts quantified following instructions from the BCA^{TM} Protein Assay Kit (Pierce). Diatom proteins were analyzed by SDS-PAGE, transferred to PVDF membranes (polyvinylidene fluoride, Millipore) and visualized with Ponceau red. Membranes were incubated for the blocking reaction in 5% milk, 0.1% Tween 20, PBS 1X for 2 hours at RT, and overnight at 4 °C with purified α-PtCPF1 antibody using 1:200 dilution in 5% milk, 0.1% Tween 20, 1X PBS. After 1 h incubation at room temperature with the secondary antibody (1:10,000 dilution) in 5% milk, 0.1% Tween 20, 1X PBS, the signal was visualized using the enhanced chemiluminescence kit (ECL kit, Amersham Bioscience).

2.23 Construction of vectors for expression of PtCPF1 in diatoms and in mammalian cells

In order to generate overexpressing *P. tricornutum* lines, the backbone vector pUC-PtCPF1 was constructed. The full length cDNA of PtCPF1 was recovered from pGEX-2TK-PtCPF1 by *Bam*HI-*Eco*RI digestion and cloned into the *Bam*HI-*Eco*RI sites of pUC19 (Amersham).

A second cloning step was to recover the full length cDNA from pUC-PtCPF1 and to clone it downstream of the *FCPB* promoter region in plasmid pKS-FcpBpAt previously used for creating the final vector for overexpression in diatom cells denoted pKS-FcpBpAt-PtCPF1.

The full length PtCPF1 cDNA, recovered from pGEX-2TK-PtCPF1 vector, was also cloned into *Bam*HI-*Eco*RI sites of pENTRY^{TM} 3C vector (Invitrogen) to

obtain the pENTRY3C-PtCPF1 plasmid that could be used for Gateway cloning. Through the LR (Left-Right) recombination reaction the PtCPF1 cDNA was thus transferred into the pDEST™ 12.2 (Invitrogen) vector and the resulting pDEST12.2-PtCPF1 vector was used to express the diatom protein in COS7 (monkey kidney cells) and BRF41(zebrafish fibroblast cells) cells for the transcription repressor assays.

2.24 Determination of transcriptional repressor activity

Luciferase reporter gene assays in COS7 and BRF41 cells were performed in collaboration with the laboratory of Prof. Takeshi Todo. The reporter construct (described in Kobayashi Y. *et al.*, 2000) was made using a 3,700 bp segment of the 5' flanking region of the *zcry3* gene cloned in the pGL3-Basic vector (Promega), generating plasmid pzcry3-luc. To express the diatom protein in mammalian cells, the pDEST12.2-PtCPF1 vector, previously described, was used. Cells (1.5 x 10^5) were seeded in 12-well plates and transfected the next day with Lipofectamine-Plus (Invitrogen). For transfection in COS7 cells the following vectors were used: pcDNA-mCRY1, pcDNA-mCLOCK, pcDNA-mBMAL1, pDEST12.2-PtCPF1. For transfection in BRF41 cells, pcDNA-zCRY1a, pcDNA-zCLOCK1, pcDNA-zBMAL3, and pDEST12.2-PtCPF1 were used. In each transfection experiment the reporter plasmid pzcry3-luc and the pRL vector (Promega) were added. The total DNA per well was adjusted to 1 μg by adding pcDNA3.1 vector (Invitrogen) as carrier. Forty-eight hours after transfection cells were harvested, and their firefly and Renilla luciferase activities determined by luminometry. The reporter luciferase activity was normalized for each sample by determining the firefly:Renilla luciferase activity ratios. All experiments were repeated three times.

2.25 Experimental design for microarray experiment

Three independent cultures of wild-type and the *PtCPF1* overexpressing line (c15) were grown with an air bubbling system under a normal light-dark photoperiod to exponential phase. Cells were then dark-adapted for 60 hours and subsequently exposed to a pulse of blue light (3.3 $\mu mol.m^{-2}.s^{-1}$) for 5 minutes. Cultures were then replaced in the dark and aliquots containing 10^8 cells were collected after 15 minutes. Blue light was provided using lamps with a broad band (380-450 nm) light source.

2.26 RNA extraction for microarray experiment

RNA extraction was performed as described in section 2.20, but an additional purification step was required. Fixed amounts of total RNA (30 μg) were loaded on columns following the RNA Cleanup Protocol from the RNeasy Plant Mini kit (Qiagen). On-column DNase digestion was performed as suggested in the optional step. RNA samples were eluted using nuclease-free water and RNA concentration was determined by spectrophotometry at 260 nm and estimated by agarose gel electrophoresis.

2.27 Microarray design and hybridization

A whole-genome expression array was designed and generated by the RZPD German Resource Center for Genome Research (Berlin, Germany). On average, five different primers of 60 nucleotides in length were designed for each gene. In total 43,860 oligos were spotted on the array, derived from the 10,364 predicted genes in the *P. tricornutum* genome. 4 X 44k custom arrays were generated following eArray instructions from Agilent, available on the internet (http://earray.chem.agilent.com/earray/login.do). Agilent default negative control

features were used to check the hybridization performance and quality control. Quality control of RNA samples was performed using a 2100 Bioanalyzer (Agilent Technologies). 0.1 µg of total RNA from each replicate sample was labelled using Cyanine 3-CTP following the instructions for "One color Microarray-Based Gene Expression Analysis". The protocol (from Agilent Technologies) is schematically represented below.

Template Total or poly A+ RNA with Spike-In
↓
cDNA synthesis
↓
cRNA synthesis and amplification
↓
cRNA purification
↓
Preparation of hybridization sample
↓
17-hour hybridization (65°C)
↓
Wash
↓
Scan
↓
Feature extraction
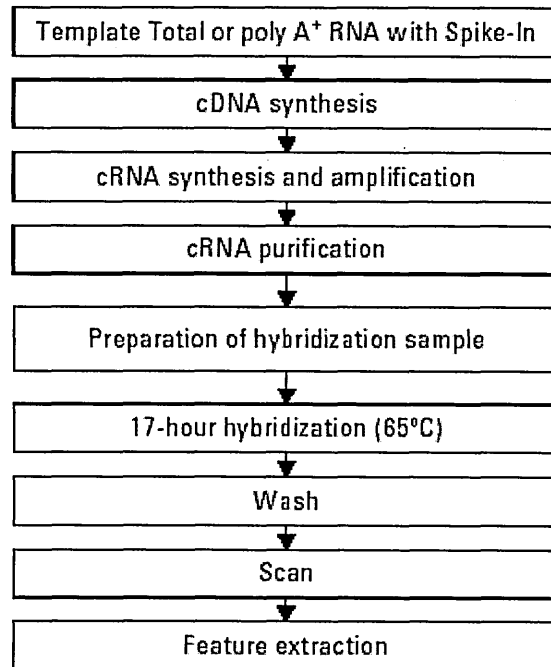
Scanning, feature extraction and normalization were performed using Agilent's Scan Control software and Feature Extraction software.

CHAPTER III - RESULTS

3.1 Development of molecular tools for *Phaeodactylum tricornutum*

3.1.1 Extraction of DNA for Sequencing of the *P. tricornutum* genome

The ecological relevance and evolutionary importance of diatoms motivated the whole genome sequencing of the centric diatom *Thalassiosira pseudonana* (*Tp*) by the US Department of Energy (DOE) (Armbrust E.V. *et al.*, 2004). The laboratory in which this thesis was developed, in collaboration with several other research groups, was actively involved in this international project. Genome assembly and annotation allowed identifying key components of several interesting or peculiar molecular pathways of diatoms (Montsant A. *et al.*, 2007).

Following this first project, a second genome sequencing project was funded by DOE for the pennate diatom *P. tricornutum* (*Pt*), and our laboratory was chosen to coordinate the project. Strain Pt1 (denoted CCMP632 in the Provasoli-Guillard National Center for Culture of Marine Phytoplankton) was chosen for the sequencing project.

In order to obtain homogeneous starting material, it was first necessary to extract genomic DNA from a clonal culture produced by repeated mitotic divisions of a single diploid founder cell. This culture was isolated by Alessandra De Martino in May 2003 and is now denoted clone CCAP 1055/1 in the Culture Collection of Algae and Protozoa (CCAP). Diatom cells were grown at 18 °C under standard conditions in a 12 hour photoperiod at approximately 180 $\mu mol.m^{-2}.s^{-1}$. Before collecting the cells, the culture was checked for contamination using standard methodologies (see Materials and Methods).

Genomic DNA was extracted following a standard protocol, but to ensure high purity it was necessary to perform ultracentrifugation in a cesium chloride gradient. The procedure is schematically represented in Fig. 10. From 4 liters of

monoclonal culture in exponential phase (*wet* weight around 20 g), around 200 µg of genomic DNA was obtained at a concentration of 2.0 µg/µL, as shown in Fig. 10. In order to confirm the high molecular weight nature of the DNA, three different molecular weight markers were loaded with it on a 0.8% agarose gel: lambda *Hind*III, lambda *Hind*III-*Eco*RI and 1kb ladder (Fermentas Life Sciences), in which the highest bands are 23 kb, 21 kb and 10 kb, respectively. Lane 3 of Fig. 10B shows 1 µL of diatom genomic DNA, whereas lanes 4 and 5 contain, respectively, 1/10 and 1/100 dilutions and the migration on the gel indeed confirmed the high molecular weight nature of the DNA. The material was therefore sent to the Joint Genome Institute (JGI, Walnut Creek, CA, USA), who generated three different libraries, of 3 kb, 8 kb and 35 kb. In order to sequence the diatom DNA, a whole-genome shotgun approach was used and a completed genome sequence is now available at: http://genome.jgi-psf.org/Phatr2/Phatr2.home.html. The nuclear genome sequence is 27 mega base pairs organized in approximately 19 chromosomes. However, an accurate study is soon to be published and will provide the basis for comparative genomics studies of diatoms with other eukaryotes and a foundation for interpreting the ecological success of these organisms.
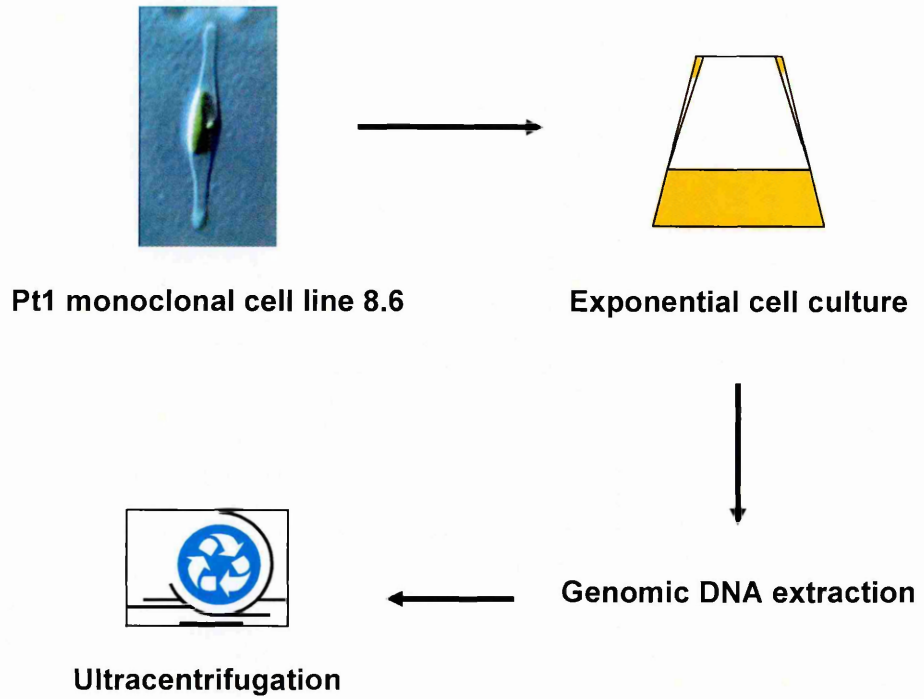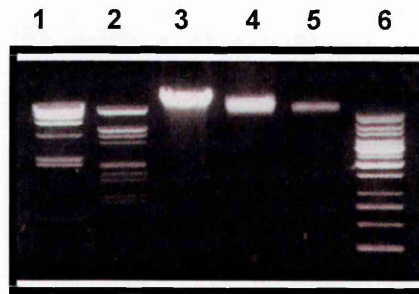
Fig. 10. A. Schematic protocol of the purification of *P. tricornutum* (*Pt1 strain*) genomic DNA. **B.** Lane 1: lambda *Hind*III marker (upper band 23 kb); Lane2: lambda *Hind*III-*Eco*RI marker (upper band 21 kb); Lane 3: 1µL of *P. tricornutum* genomic DNA; Lane 4: 1/10 dilution of *Pt* genomic DNA; Lane 5: 1/100 dilution of *Pt* genomic DNA; Lane 6: 1kb marker (upper band 10 Kb).

3.1.2 Gateway cloning technology for *P. tricornutum*

The Gateway Technology is a universal cloning method designed by Invitrogen (Carlsbad, USA) (Walhout A.J. *et al.*, 2000). This technology is based on the conservative and site-specific recombination machinery used by phage λ to integrate and excise into and out of the *E. coli* chromosome. Recombination components have been improved for optimum efficiency and specificity of the recombination process in order to provide a rapid and highly efficient way to move a gene of interest into multiple vector systems. To date, Invitrogen offers state-of the-art Gateway® destination vectors for expression of native or tagged proteins in *E. coli*, insect, yeast or mammalian cells, providing extensive possibilities for functional studies (see scheme in Fig. 11), and a common objective of our laboratory has been to generate Gateway-based diatom vectors.

The first step of the procedure consists in the cloning of a gene of interest into the vector denoted pENTRY such that the lambda recombination sites are located to the left (5') and to the right (3') of the gene of interest (attL1-attL2, respectively), replacing the position of the toxic *ccdB* gene. Moreover, in this first step it is extremely important to clone the transcript of interest in the appropriate pENTRY vector such that the gene of interest will be in frame in the derived fusion proteins. Once a gene of interest has been inserted into a pENTRY plasmid by the LR (Left-Right) recombination reaction it can be transferred into different Destination vectors (pDESTs).

In this reaction the bacteriophage λ Integrase (Int), together with the Excisionase (Xis) and the *E. coli* Integration Host Factor (IHF) proteins, replace the pDEST cassette carrying the *ccdB* gene and the *cat* gene (for chloramphenicol resistance) with the sequence present between the attL1-attL2 sites of the pENTRY vector, as schematically presented in Fig. 11. The non-recombined vectors harboring

the toxic *ccdB* gene are eliminated after transformation in *E. coli* TOP10 strain and selection on the appropiate antibiotic (ampicilllin).

Based on this technology, we decided to establish a cloning system for the expression of diatom genes by constructing diatom Destination vectors with useful epitopes (Siaut M. *et al*, 2007). As a strong promoter the upstream region of the *FCPB* gene (Fucoxanthin Chlorophyll Binding Protein B, FcpBp) was chosen and the terminator was derived from the *FCPA* gene (Fucoxanthin Chlorophyll Binding Protein A, FcpAt). In total, seven Destination vectors were designed for a variety of specific purposes including gene overexpression, protein localization, and production of tagged proteins for immunopurification, immunolocalization and protein detection. The first vector was made to overexpress a gene product in diatom cells in order to get information about its function (Fig. 12A). Four vectors were constructed for the analysis of subcellular localization, thanks to the feasibility to fuse a gene of interest at the amino- and carboxy-terminal extremity with sequences encoding enhanced yellow fluorescent protein (EYFP) or enhanced cyan fluorescent protein (ECFP), two variants of the *Aequorea victoria* green fluorescent protein (Ormo M. *et al.*, 1996; Heim R. & Tsien R.Y., 1996 respectively). Two other vectors were designed to produce proteins tagged at amino- or carboxy-terminal extremities with the HA (HemAglutinin Influenza virus) epitope (Fig. 12). In order to validate this powerful cloning system all diatom Gateway-compatible Destination vectors were tested.

My contribution to this work was to create the diatom Gateway Destination vector pDEST-C-EYFP and the diatom Gateway Destination vector pDEST-C-HA. Cloning strategies have been described in details in the Materials and Methods section. It is interesting to show two examples, made by my colleagues, that confirm the power of this cloning system for the diatom scientific community. The coding

region of the protein SEC4, a member of the small GTPase family of membrane traffic regulators, was isolated from *P. tricornutum* and cloned in a pENTRY vector. Through the recombination reaction the cDNA was transferred to the pDEST-C-EYFP vector. Transgenic diatom cells overexpressing the resulting fusion protein were generated and microscope analysis showed a subcellular localization associated with a complex network of membranes. PtSEC4 appeared localized in small vesicles arranged longitudinally on the plasma membrane (Fig. 13A), in agreement with the know function of SEC4 in polarized secretion in yeast (Goud B. *et al.*, 1988).

As a second example, the coding region of a putative phytochrome gene isolated from *P. tricornutum* (denoted *PtPHY*) was transferred from a pENTRY vector to the pDEST-C-HA destination vector. The final construct was shot into diatom cells and positive transgenic cells were screened. Western blotting analysis using a monoclonal HA antibody confirmed expression of the fusion protein at the expected molecular weight (Fig. 13B).
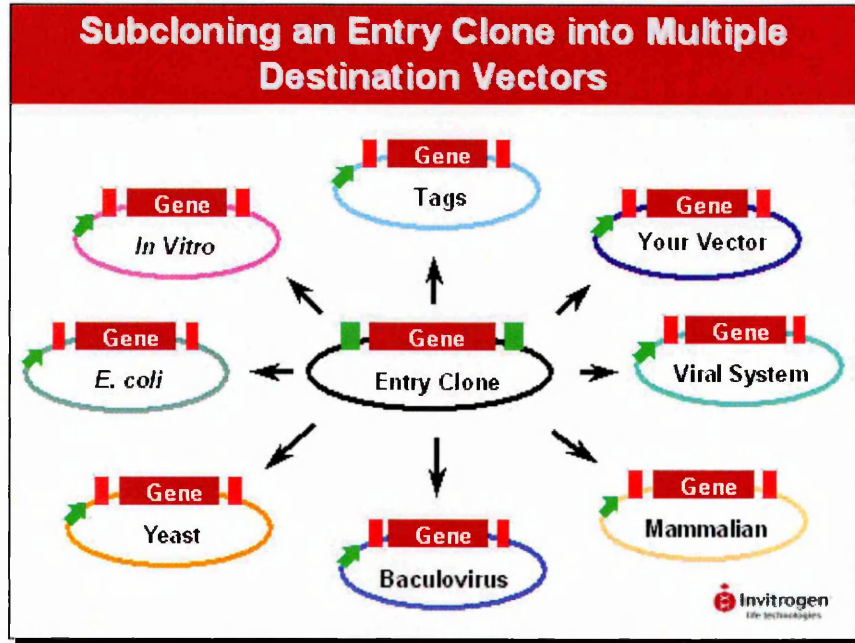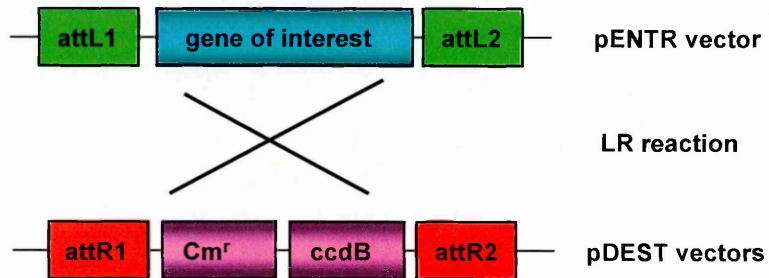
Fig. 11. A. The Gateway cloning techonology allows transfer of a gene of interest simultaneously into a range of Destination vectors. **B.** The recombination reaction (LR reaction) replaces the pDEST cassette (Cm^r, choramphenicol resistance; *ccdB* gene) with the gene cloned in pENTR vector using site specific recombination sites attL1-attL2, attR1-attR2.

A. Over-expression

| FcpBp | attR1 | Cm^r | ccdB | attR2 | FcpAt | pDEST-OX |

B. N-term tags

FcpBp	EYFP	attR1	Cm^r	ccdB	attR2	FcpAt	pDEST-N-EYFP
FcpBp	ECFP	attR1	Cm^r	ccdB	attR2	FcpAt	pDEST-N-ECFP
FcpBp	HA	attR1	Cm^r	ccdB	attR2	FcpAt	pDEST-N-HA

C. C-term tags

FcpBp	attR1	Cm^r	ccdB	attR2	EYFP	FcpAt	pDEST-C-EYFP
FcpBp	attR1	Cm^r	ccdB	attR2	ECFP	FcpAt	pDEST-C-ECFP
FcpBp	attR1	Cm^r	ccdB	attR2	HA	FcpAt	pDEST-C-HA

Fig. 12. Diagrams illustrate diatom Gateway Destination vectors for protein overexpression (**A**), amino-terminal protein tagging (**B**) and carboxy-terminal protein tagging (**C**).

A

pDEST-PtSec4:EYFP

B

pDEST-PtPHY:HA
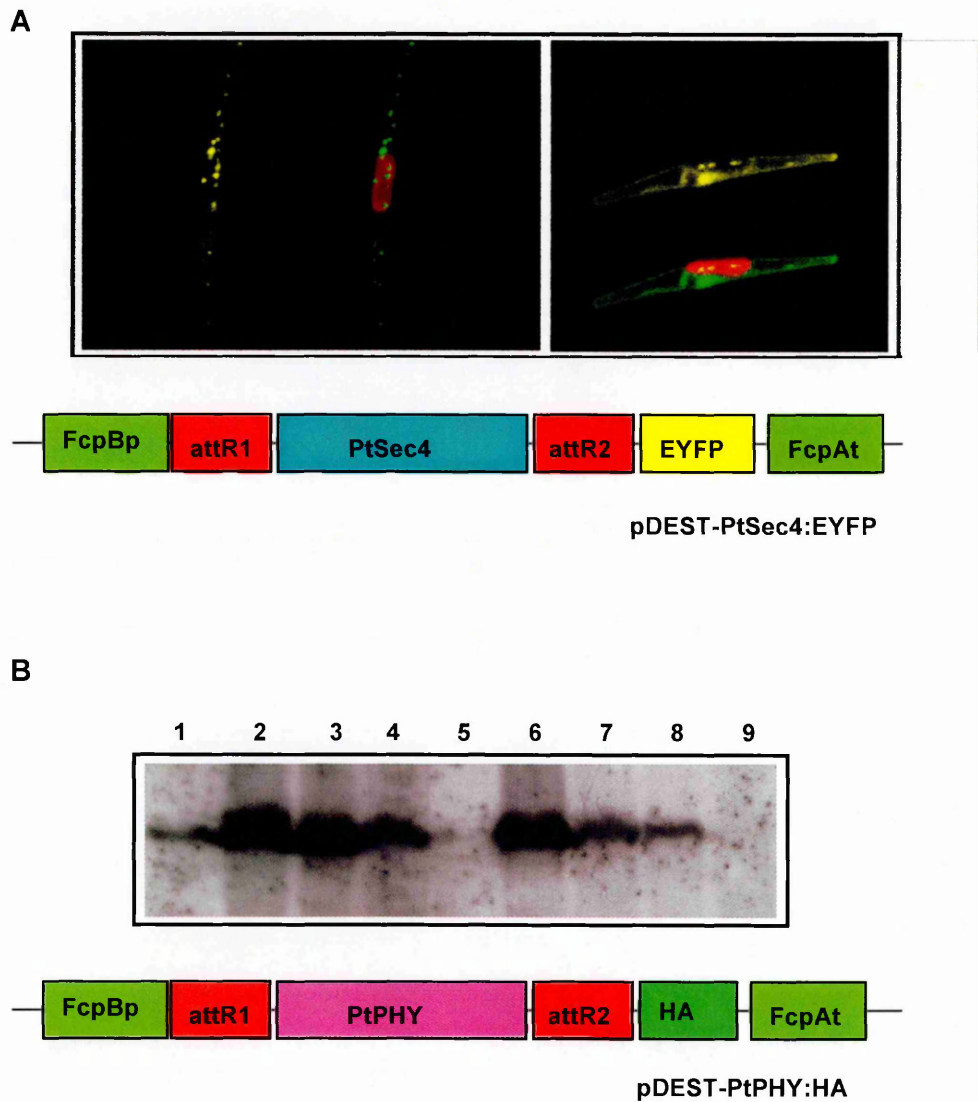
Fig. 13. **A.** Confocal microscope images of diatom transgenic cells overexpressing the PtSEC4:EYFP construct (left and right panels show the frontal and lateral vision, respectively). The EYFP signal corresponds to the subcellular localization of SEC4 protein. A diagram of the diatom gateway Destination vector is illustrated. **B.** Western blotting using α-HA antibody. Protein samples were extracted from wild type diatom cells (lane 9) and diatom transgenic cells overexpressing the PtPHY:HA construct (lanes 1-8). A diagram of the diatom gateway Destination vector is illustrated.

3.2 Identification of a putative blue light photoreceptor *PtCPF1*

3.2.1 Identification and cloning of the *PtCPF1* gene

Prior to sequencing of the *P. tricornutum* genome, the laboratory generated several EST (Expressed Sequence Tag) collections (Scala S. *et. al.*, 2002; Maheswari U. *et al.*, 2005). Searches of these databases by BLAST (Basic Local Alignment Search Tool) for photoreceptor genes revealed a fragment of 200 base pairs (bp) that showed high similarity with members of the Cryptochrome/Photolyase gene family. This fragment was amplified by PCR and subsequently used as a probe to screen a *P. tricornutum* cDNA library in an attempt to identify a full length clone. The result of this analysis was the isolation of a clone containing the ATG start codon but lacking the C-terminal extremity. To obtain the 3' region of the transcript it was necessary to use 3'RACE (Rapid Amplification of cDNA Ends, see Materials and Methods).

The complete sequence from start to stop codon consists of 1,653 bp in length and contains an open reading frame encoding a predicted protein of 550 amino acids with a calculated mass of 63 kDa. Subsequent sequencing of the corresponding genomic sequence revealed a length of 2,006 bp containing four introns of around 90 bp each. Because it was not possible from the genomic analysis to assign the gene as encoding a cryptochrome or a photolyase, we decided to name this gene *PtCPF1*, for *P. tricornutum* Cryptochrome/Photolyase Family 1.

3.2.2 Phylogenetic analysis

Following the subsequent availability of genomic sequences from two diatoms (*Pt* and *Tp*) we searched the respective gene models for Cryptochrome/Photolyase genes. We identified four sequences in *Tp* (*TpCPF1-4*)

and three putative members in *Pt* (denoted *PtCPF1*, *PtCPF25* and *PtCPF4*). To establish the phylogenetic relationships between these amino acid sequences, five representative members of each Cryptochrome/Photolyase subfamily were selected. The class II CPD photolyases were not included in this study because of their ancient divergence (Kanai S. *et al.*, 1997). On the contrary, it was important to add (6-4) photolyase sequences from different organisms, in which the amino acid composition is more similar to the animal cryptochromes, as well as class I CPD photolyases that cluster with the plant cryptochromes. A multiple alignment of the selected sequences was then made and a phylogenetic tree was constructed, by the Neighbour Joining (NJ) method (Saitou N. & Nei M., 1987), using a conserved region of around 400 amino acids (Fig. 14).

Surprisingly, none of the diatom sequences clustered with plant cryptochromes. In particular *PtCPF1* and *TpCPF1* clustered in the animal cry/(6-4)photolyase clade, whereas the other *CPF* members were more related to a novel class of cryptochromes recently identified and denoted cry-DASH (Brudler R. *et al.*, 2003) whose function has not yet been clarified. Tree topology was conserved using the alignment of full length protein sequences (data not shown).
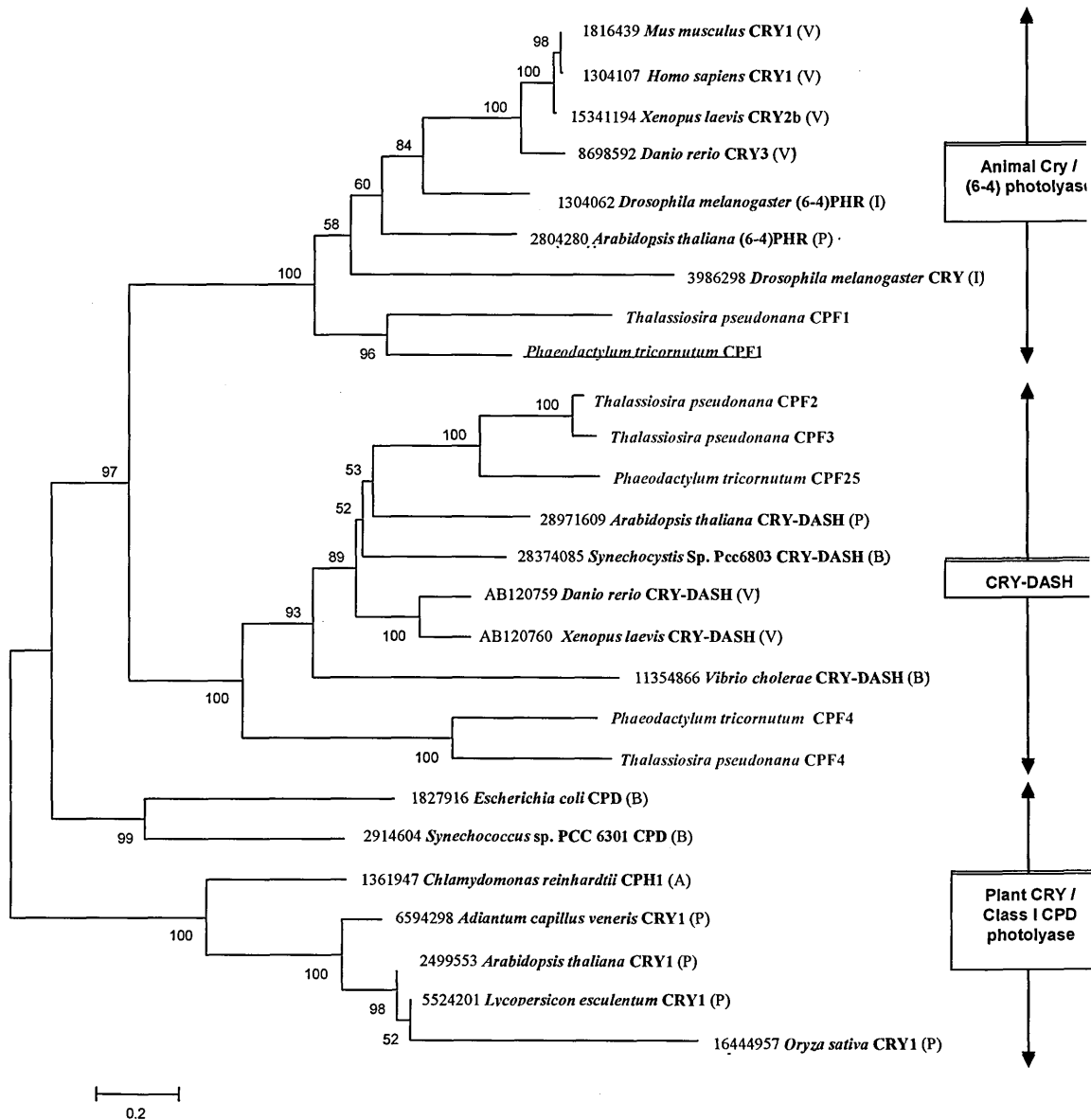
Fig. 14. An unrooted Neighbour-Joining phylogenetic tree of the Cryptochrome/Photolyase Family. Each sequence is indicated by the GI number from NCBI, except for diatom sequences. The source name is shown and a capital letter after each name indicates the taxonomic category of the source. (V) vertebrates, (P) plants, (I) insects, (B) eubacteria, (A) algae. Bootstrap probability (of 1,000 replicates) is written near the node. Diatom sequences are colored in red and the sequence characterized in this thesis is underlined.

3.2.3 Domain analysis

A domain analysis of the PtCPF1 protein sequence was performed in order to check if crucial domains and key residues were conserved (Fig. 15). Cryptochrome proteins have two domains, an N-terminal conserved domain, the "photolyase related" domain, and a carboxy-terminal "tail" that is intrinsically unstructured and varies considerably in length and primary amino acid sequences (Sancar A., 2003; Green C.B., 2004; Partch C.L. *et al.*, 2005). Both domains are important for the ability of mammalian CRY to inhibit CLOCK/BMAL driven transcription (see section 1.3.7.2). Hirayama J. *et al.* (2003) previously generated chimeras between the transcription repressing zCRY1a and the non-repressing zCRY3 proteins of zebrafish and identified three regions in zCRY1a (RD2a, RD1, and RD-2b), together with a putative nuclear localization signal (NLS) within the RD-2b region. RD2a (aa 126-196) or RD1 (aa 197-263) are required for the interaction with the CLOCK-BMAL1 heterodimer, and either RD1 or RD-2b (aa 264-293) is required for nuclear translocation of the protein. In particular, the NLS-like sequence is identified between 265-282 amino acids of the RD-2b region. Moreover, the functional nuclear localization signal identified in the RD-2b region is well conserved among repressor-type, including mCRY1 (aa 265-282). Mutations in the NLS of mCRY1 reduce the extent of its nuclear localization (Hirayama J. *et al.* 2003).

Surprisingly, study of the primary structure of the PtCPF1 protein showed that the overall similarity between the diatom protein and mammalian-type CRY includes these three functional domains (40% identity, 60% similarity) that, on the contrary, are poorly conserved in plant cryptochromes (note AtCRY1 in the alignment of Fig. 15). The functional NLS of the RD-2b region is almost conserved in PtCPF1. Moreover, Sanada K. *et al.* (2004) identified a key residue for the phosphorylation of mCRYs (Ser247 in mCRY1, corresponding to Ser265 in

mCRY2). This post-translational modification is an essential step for the circadian time-keeping mechanism. This residue is also conserved in PtCPF1 and corresponds to Ser276. In addition, the highly conserved flavin-binding residue D410 of *Drosophila* CRY, that has been shown to be essential for circadian photoreceptor function (Stanewsky R. *et al.*, 1998), is conserved in the PtCPF1 sequence and corresponds to the aspartic acid at position 418.

Much more variation in the light-dependent cry functions between *Drosophila*, mammals, and plants is attributed to the C-terminal. In particular, further studies have shown that the C-terminal extension of mCRY1 harbors an additional nuclear localization signal (aa 578-606) and a putative coiled-coil domain (aa 471-493) that drive nuclear localization via two independent mechanisms (Chaves I. *et al.*, 2006). Importantly, *Arabidopsis* (6-4)PHR is able to confer CLOCK:BMAL1 transcription-inhibitory activity when is fused to the last 100 amino acids of the mCRY1 core and its C terminus (aa 371-606) (Chaves I. *et al.*, 2006).

PtCPF1 contains a short C-terminal extension comprised of 28 amino acids, comparable to that described in *Drosophila* of 26 amino acids. The NLS present in the C-terminal domain of mCRYs, characterized also in *Xenopus* CRY2b (Zhu H. *et al.*, 2003), is absent in the diatom sequence, whereas it is important to note that a potential coiled-coil domain is predicted by the COILS program (http://www.ch.embnet.org/software/COILS_form.html) at the C-terminal extension of PtCPF1 (aa 502-524).

Finally, the typical feature of the C-terminal extension of plant CRYs, referred as the DAS domain (see section 1.3.1), is absent in PtCPF1 sequence as is the NC80 motif. The latter has recently been described in AtCRY2 as a critical motif for photoactivation (Yu X. *et al.*, 2007). In summary, these analyses show that the

diatom sequence appears more similar to the mammalian-type CRY proteins than the plant members.
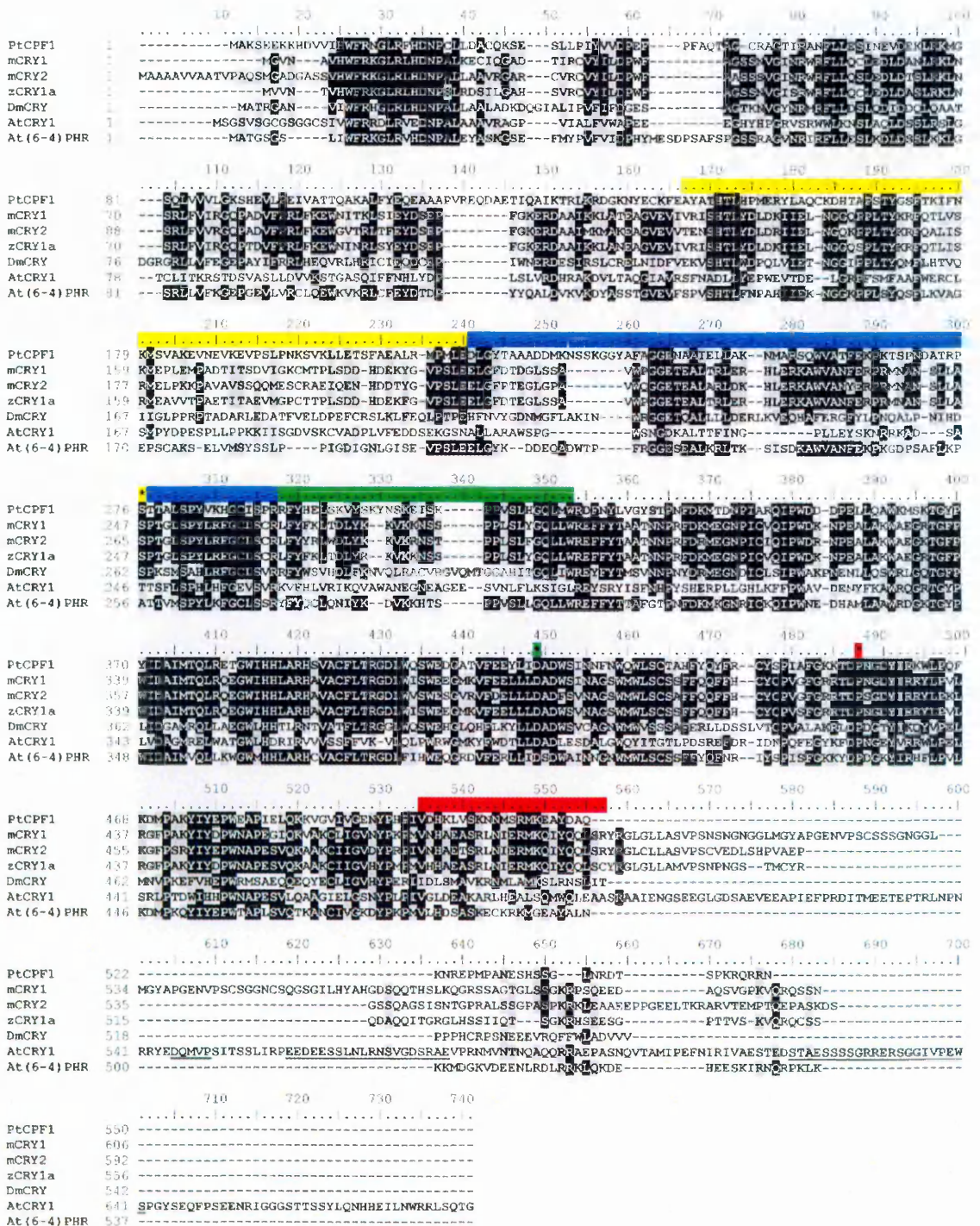
Fig. 15. Amino acid sequence alignment of PtCPF1 with representative Cryptochrome/Photolyase sequences. The sequences were aligned using ClustalW. Black-boxed and gray-boxed letters represent identical or similar residues, respectively. Yellow, cyan and green boxes indicate RD2a, RD1 and RD-2b domains, respectively. The red box shows the coiled-coil domain. The DAS domain is underlined. Yellow asterisk indicates the conserved Ser residue for the phosphorylation of mCRYs. Green asterisk indicates the conserved flavin-binding residue identified in the *Drosophila cry^b* mutant, and red asterisk indicates the last amino acid of the putative short PtCPF1 protein. For descriptions see text.

3.3 Biochemical and Functional Characterization of PtCPF1

3.3.1 Expression and Purification of GST:PtCPF1 in *E. coli*

An important objective was to develop a protocol for the purification of recombinant PtCPF1 that could be used to perform biochemical and functional studies, which could help to characterize its potential photolyase and/or photoreceptor activities. This was attempted using an *E. coli*-based expression system to overexpress the PtCPF1 protein fused to a GST (Glutathione S-Transferase) tag. Because cryptochromes are thought to be derived from photolyases and because some cryptochromes also display some photolyase activity, it was considered important to express the protein in a photolyase-deficient *E. coli* host. Such an *E. coli* strain, denoted SY2, was obtained from Prof. Takeshi Todo (Kyoto University, Japan). Initial attempts to purify the protein were hampered by the fact that the majority of recombinant GST:PtCPF1 was found to be insoluble and to be present in inclusion bodies. An expression protocol was therefore optimized in which induction was performed with 0.1 mM IPTG at 16 °C for 21 hours, in order to generate as much protein in the soluble fraction as possible. Using this protocol, it was possible to detect approximately 5% of the protein in the soluble fraction by western blotting using the α-PtCPF1 antibody that was considered sufficient to begin to purify the protein to near-homogeneity. The purification procedure, schematically represented in Fig. 16, consists of binding to glutathione resin in batch overnight at 4 °C, followed by an elution step using a column, dialysis in 50 mM Tris-HCl pH 7.5, and concentration on Centricon filters (for details see Materials and Methods section). The GST:PtCPF1 protein, purified to near homogeneity, displayed a pale yellow colour, suggesting the presence of the flavin chromophore. From 3 litres of bacterial culture it was possible to generate approximately 1

milligram of purified soluble recombinant protein. Fig. 16B shows the profile of the purified protein. Two major bands could be observed: the full length GST:PtCPF1 protein of around 90 kDa and a major degradation product that could correspond to the PtCPF1 protein of 63 kDa without the GST tag.
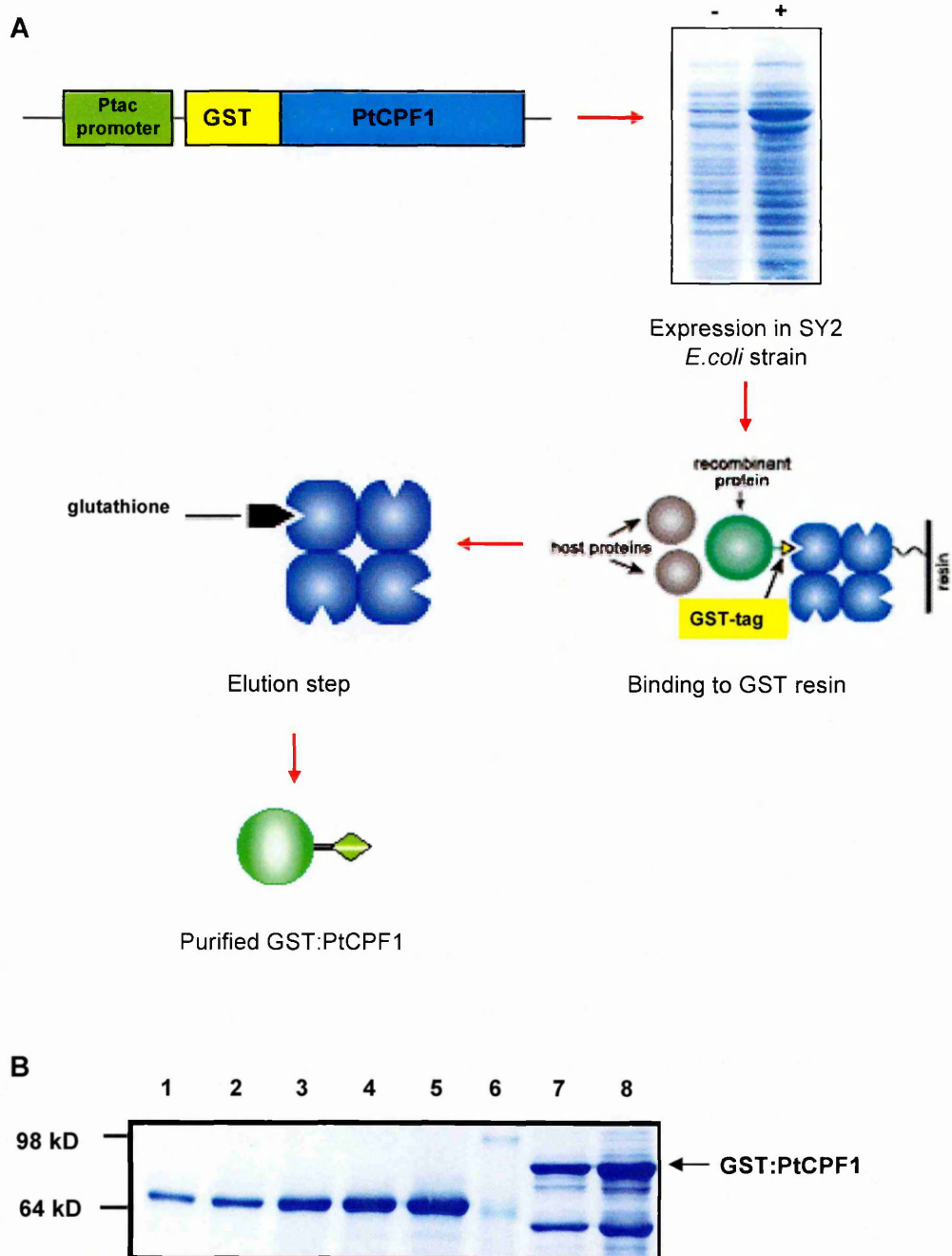
Fig. 16. A. Schematic protocol for PtCPF1 protein purification. Coomassie Blue stained SDS-PAGE gel (10% polyacrylamide) of not induced (-) and induced (+) extracts of *E. coli* transformed whit pGEX-2TK-PtCPF1. **B.** Coomassie Blue stained SDS-PAGE gel (10% polyacrylamide) of purified PtCPF1 (1 μL and 3μL, respectively, loaded in lanes 7, 8). Fixed amounts of BSA protein were loaded to estimate the purified GST:PtCPF1 (1, 2, 4, 6, 10 μg, respectively, loaded in lanes 1-5). Molecular weight marker (lane 6).

3.3.2 Absorption Spectra and Fluorescence Studies

All Cryptochrome/Photolyase Family proteins characterized to date contain FAD (Flavin Adenine Dinucleotide) as an essential chromophore. In addition these proteins often contain a second chromophore, which in the majority of organisms is (N5,N10)-methenyl-5,6,7,8-tetrahydrofolate (MTHF) and in a limited number of species is 8-hydroxy-7,8-didemethyl-5-deaza-riboflavin (8-HDF) (Cashmore A.R. *et al.*, 1999; Sancar A., 2003). The folate or deazaflavin class enzymes exhibit a major absorption peak at 375-410 nm or at 440 nm caused by each second chromophore, respectively. Moreover, both classes of enzymes and either an additional peak at 440 nm caused by fully oxidized FAD (FAD_{ox}) or several peaks at 480, 580 and 625 nm, caused by the flavin blue neutral radical, $FADH^0$.

Thanks to the optimized purification protocol, it was possible to elucidate the spectroscopic properties of the PtCPF1 protein. Absorption and fluorescence spectra of PtCPF1 were performed in collaboration with the laboratory of Prof. Takeshi Todo (Kyoto University). As shown in Fig. 17, the absorption profile does not exhibit any peak at 375-410 nm, indicating that the protein sample does not contain the second chromophore. On the contrary, by examining the profile of the denatured protein, it was possible to observe two peaks at 350 and 450 nm, indicating the presence of the FAD chromophore (absorption spectrum, profile c). Moreover it is important to notice the presence of a large peak at 260 nm of the PtCPF1 absorption spectrum that could suggest the presence of nucleic acids co-purified with the diatom protein (see later). The denatured GST protein, used as a negative control, did not show any peak at 260 nm (Fig. 17B).

To obtain further evidence that the diatom protein contained the FAD chromophore, the fluorescence spectrum was determined. Typically, the flavin molecule has an emission maximum at 502-520 nm with excitation maxima at 370

and 440 nm. Unfortunately, the classical fluorescence spectrum showed the same pattern for both proteins, the GST:PtCPF1 and the negative control, probably because only a limited amount of purified protein carried the FAD cofactor (Fig. 18).

It was therefore necessary to confirm the presence of the chromophore by releasing it from the protein. The diatom protein and the negative control were heat-denatured at neutral pH and, following removal of precipitated material by centrifugation, the absorption spectra were recorded. For the diatom protein the fluorescence emission was pH dependent, showing a 2.5 fold-higher fluorescence at pH 3.0 than at pH 7.5. This pH dependency of fluorescence emission is typical of the flavin molecule (Weber G., 1959), thus confirming the presence of FAD in PtCPF1. Moreover, the excitation at 450 nm resulted in fluorescence emission with a maximum at 530 nm, indicating that the FAD molecule is in a reduced state (Fig. 18C). By contrast, the denatured GST protein did not display any fluorescence emission.
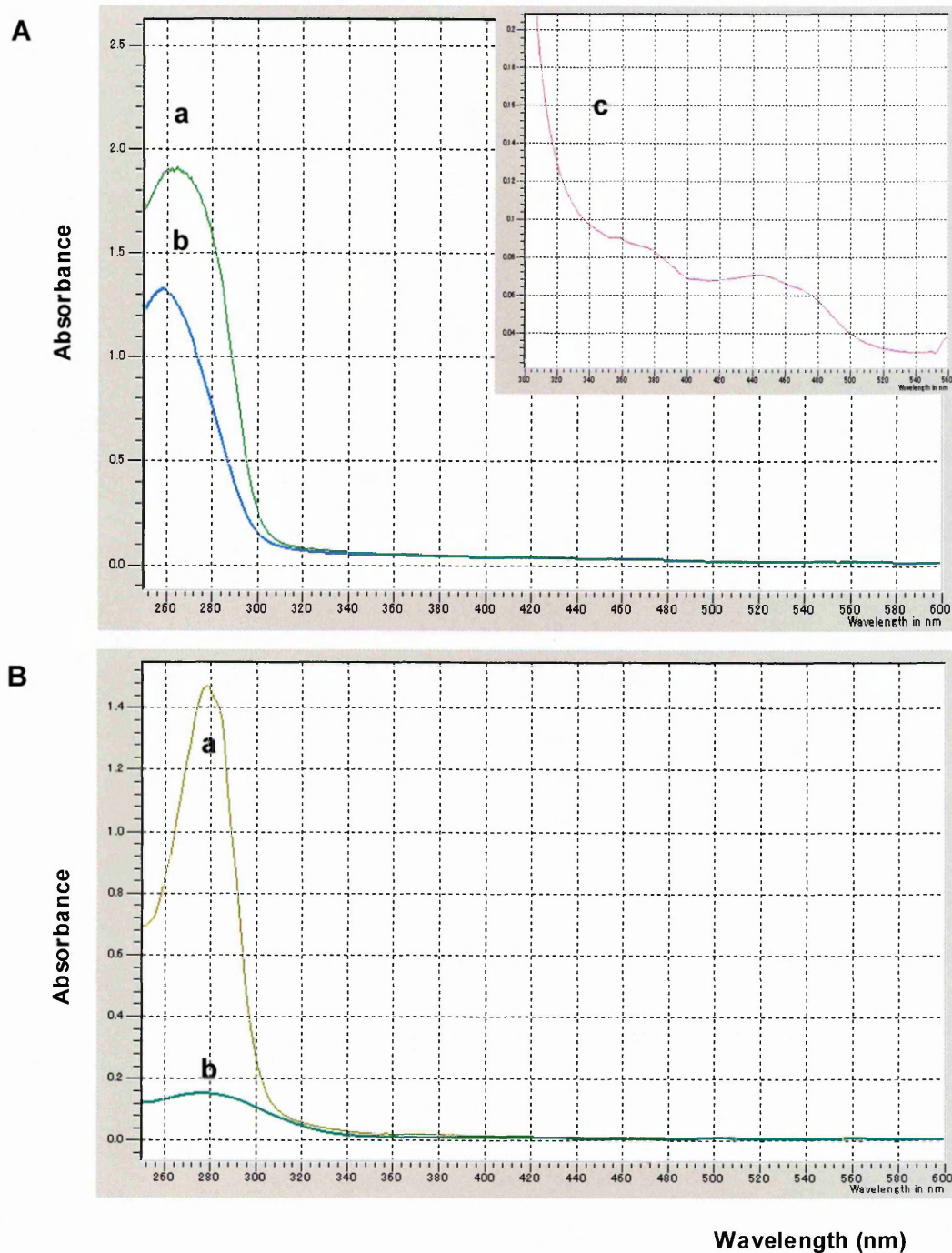
Fig. 17. A. Absorption spectra of PtCPF1 (profile a), of heat-denatured PtCPF1 (profile b). The inset (profile c) shows an expanded scale of the absorption spectrum in the 300-560 nm range of heat-denatured PtCPF1. **B.** Absorption spectra of GST (profile a), and of heat-denatured GST (profile b).
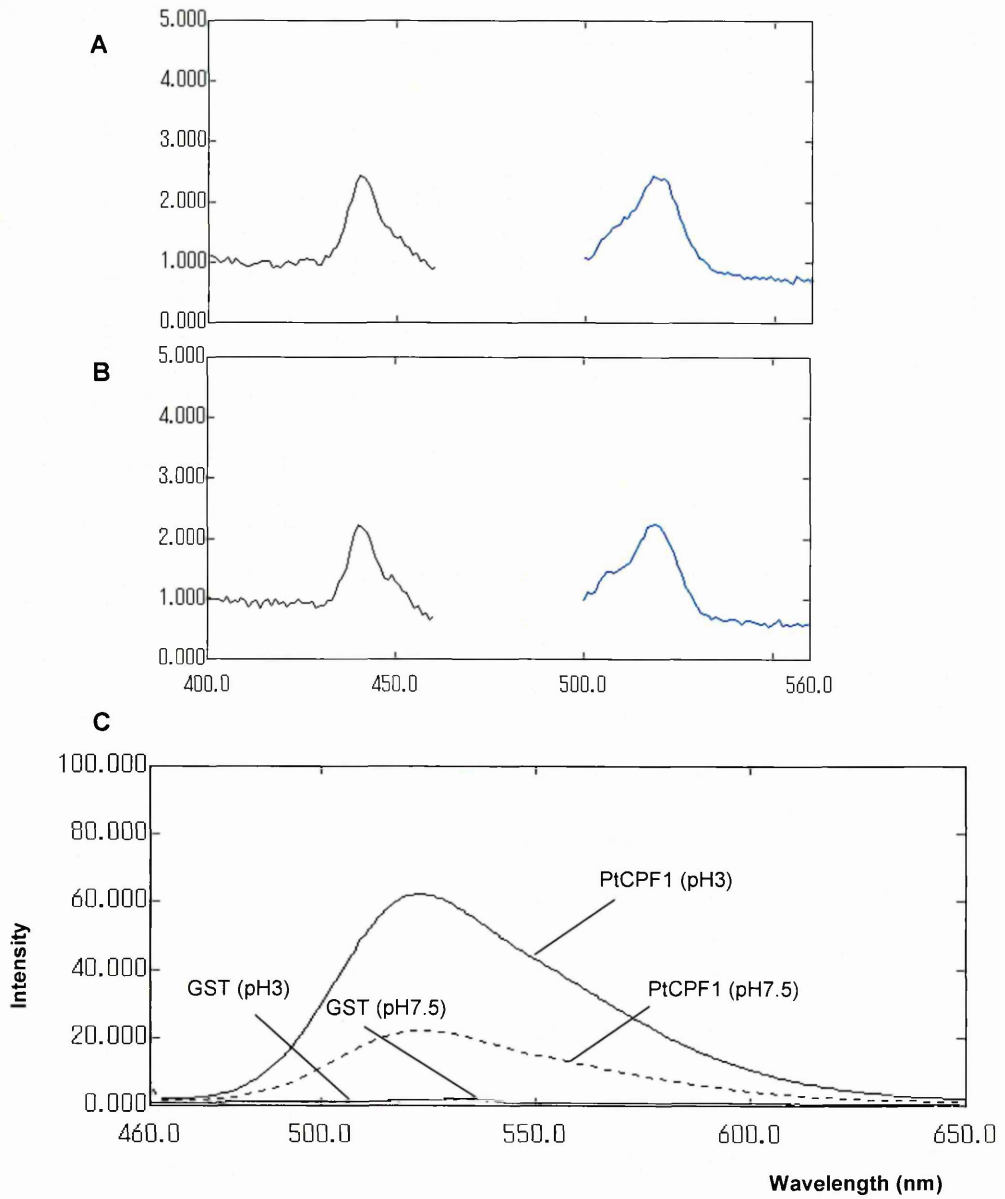
Fig. 18. Fluorescence spectra of PtCPF1 (**A**) and GST (**B**). Flavin fluorescence using 470 nm excitation for the emission spectrum and 520 nm emission for the excitation spectrum. **C.** Fluorescence spectra of heat-denatured PtCPF1 and GST proteins at neutral pH. Fluorescence emission of the diatom protein is pH dependent.

3.3.3 Determination of composition of nucleic acids extracted from the purified protein

Remarkably, as anticipated previously, the purified recombinant PtCPF1protein always exhibited a 260 nm absorbing peak in the near UV, suggesting the presence of nucleic acids in the sample preparation. Although the nucleic acids bound to the protein cannot be the native ones because the protein is expressed and purified in *E. coli*, we nonetheless considered it interesting to determine the nature of this nucleic acid.

In order to answer this question, nucleic acids were extracted from the protein sample by phenol/chloroform and recovered by ethanol precipitation. They were then treated with either RNase (DNase-free) or DNase (RNase-free) and analyzed by electrophoresis on 2% agarose gels. Diatom ribosomal RNA and a 200 bp fragment of plasmid DNA were used as controls to check the specificity of the enzymes. The result of this analysis showed that the PtCPF1-associated nucleic acid was digested by RNase but not by DNase and therefore can be identified as RNA (Fig. 19). From the molecular weight markers used as reference, it appears that the RNA sample is less than 100 nucleotides, in reasonable agreement with the estimate made from the absorption spectra. It is interesting to note that the *Vibrio cholerae* cryptochrome 1, purified from *E. coli* by amylose affinity chromatography, has also been shown to be associated with RNA (Worthington E.N. *et al*, 2003). Further analyses, such as protein purification directly from diatom cultures, will clarify the nature of this unknown RNA and its biological role.
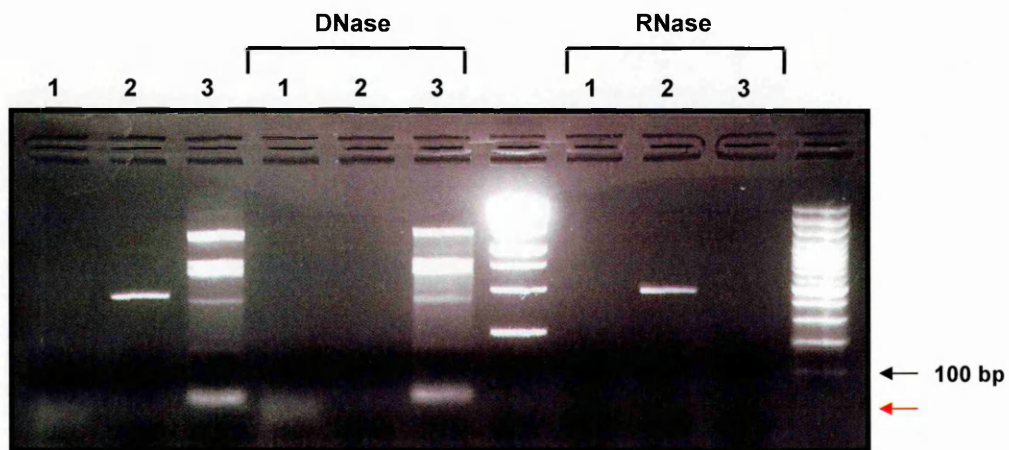
Fig. 19. Nucleic acids extracted from the purified PtCPF1 protein (1) were treated with DNase or RNase enzymes. A 200-bp DNA fragment (2) and ribosomal diatom RNA (3) were used to check the specificity of the enzymes. Nucleic acids associated with PtCPF1 protein are indicated by red arraow. Last lane shows molecular weight marker.

3.3.4 DNA Binding and Repair assay - *In vitro* assays -

Because the PtCPF1 protein is a member of the Cryptochrome/Photolyase Family, it was fundamental to test its photolyase activity using established *in vitro* experiments. In collaboration with the laboratory of Prof. Takeshi Todo, two important experiments were performed: a gel mobility shift to assess binding to DNA and an *in vitro* DNA repair assay. For the binding experiment, the purified protein was incubated with a 49 bp DNA fragment which contains a single UV photoproduct (either a CPD, a (6-4) photoproduct, or a DEWAR isomer) at the *MseI* restriction site, as is schematically shown in Fig. 20. Dewar isomer is a photoproduct derived from the irradiation of (6-4)photoproducts at 313 nm (Taylor J.-S. & Cohrs M.P., 1987). The (6-4)photolyase enzyme binds DNA contaning the Dewar photoproduct, albeit its repair is extremely slow (Zhao X. *et al.*, 1997).

The principle of this experiment is to test if the diatom protein is able to recognize UV-induced DNA damage and bind the specific DNA sequence in order to repair it. The result of this experiment shows that the diatom protein can specifically bind only the sequence containing the (6-4)photoproduct and the Dewar isomer, not the CPD photoproduct (Fig. 20B). *E. coli* CPD photolyase and *A. thaliana* (6-4) photolyase were used as controls.

The next step was to test the repair activity of the purified diatom protein. The same DNA probe of 49 bp can be digested by *MseI* only if the UV photoproduct has been efficiently repaired by the specific enzyme. In our experiments the presence of a 21-mer, due to repair of the UV-induced photoproduct, was detected only with the (6-4) photoproduct DNA probe after treatment with PtCPF1, demonstrating that the diatom protein has a specific (6-4) photolyase activity (Fig. 21B).
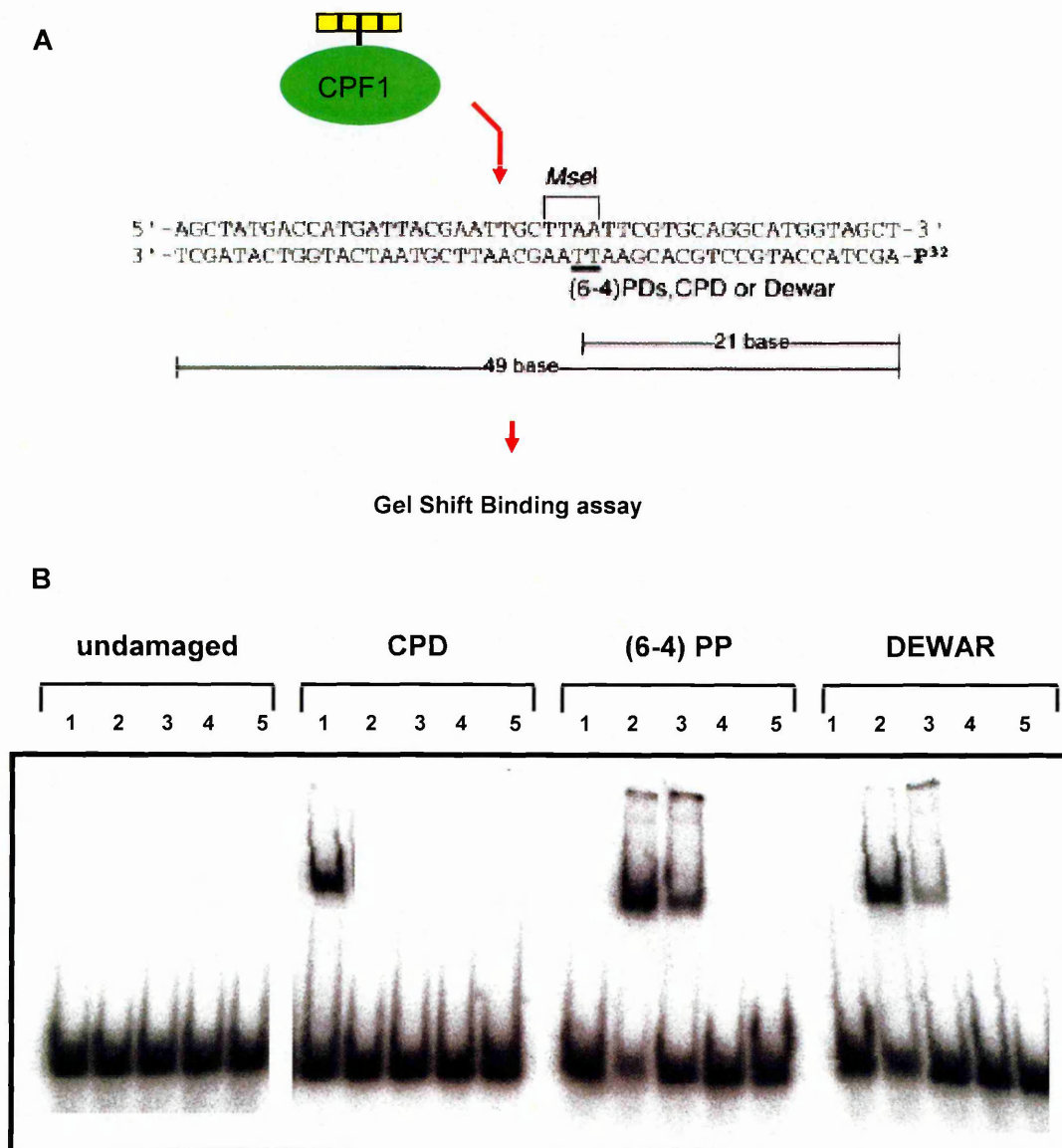
Fig. 20. A. Schematic of DNA gel shift binding experiment. PtCPF1 protein is incubated with the 49 bp DNA probe containing a single UV photoproduct at the *Mse*I restriction enzyme. **B.** Gel retardation analysis by electrophoresis in a non denaturing gel. Sample 1: *E. coli* CPD photolyase, sample 2: *A. thaliana* (6-4) photolyase, sample 3: PtCPF1, sample 4: GST, sample 5: without protein. PtCPF1 shows specific binding to (6-4) photoproducts.
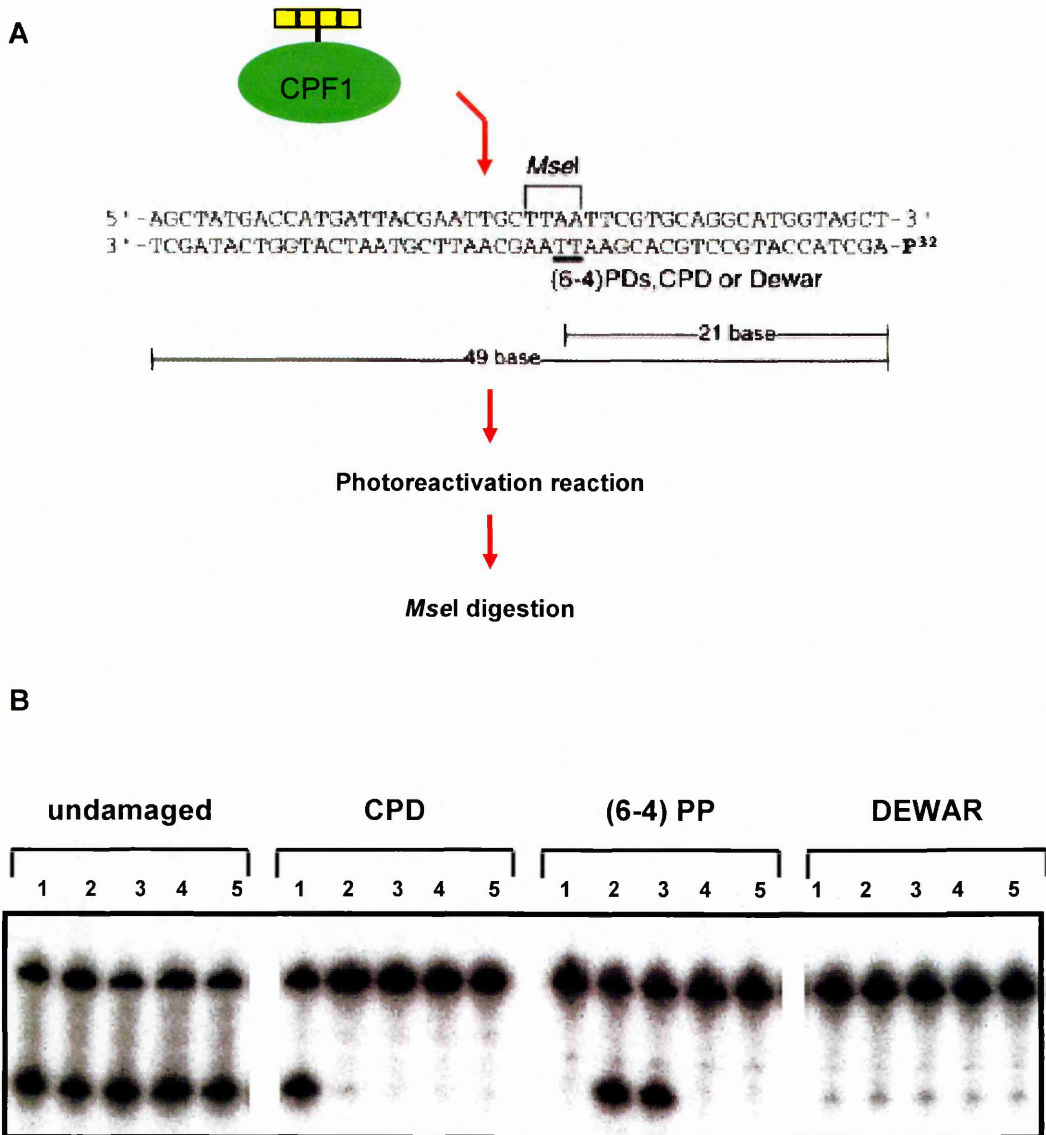
Fig. 21. A. Schematic of DNA photorepair assay. PtCPF1 protein is incubated with the 49 bp DNA probe containing a single UV photoproduct at the *Mse*I restriction site. Following photoreactivation the DNA is digested with *Mse*I. **B.** Electrophoresis in a 10% polyacrylamide gel. Sample 1: *E. coli* CPD photolyase, sample 2: *A. thaliana* (6-4) photolyase, sample 3: PtCPF1, sample 4: GST, sample 5: without protein. PtCPF1 protein shows specific (6-4)photolyase activity.

3.3.5 Photoreactivation of UV-damage in the repair-defective SY32 strain - *In vivo* assay -

To confirm the functional role of DNA repair of the PtCPF1 protein it was important to use an *in vivo* assay. *E. coli* cells normally do not photoreactivate (6-4) photoproducts (Todo T. *et al.*, 1993), but the SY32 *E. coli* strain, kindly provided by Prof. Takeshi Todo, has been engineered to test for (6-4) photolyase activity in a functional complementation assay. These bacterial cells are deficient for DNA repair mechanisms, but they carry a plasmid encoding the *E. coli* CPD photolyase gene (*phr*). The complete genotype is *recA⁻ uvrA⁻ phr⁺* (Todo T. *et al.*, 1996). In these cells, CPD photoproducts are repaired efficiently by the CPD enzyme so the extent of repair of (6-4) photoproducts determines the sensitivity of the cells to UV light.

The SY32 strain was transformed with an inducible vector encoding the diatom protein (pGEX-2TK-PtCPF1), as well as plasmids encoding zebrafish photolyase (pGEX-4T-2-z(6-4)PHR) and zebrafish cryptochrome 2a (pGEX-4T-2-zCRY2a), to be used as positive and negative controls, respectively. *E. coli* cells were induced by IPTG in order to express the proteins, then treated under UV light and subjected to a photoreactivation reaction using fluorescent white light for one hour. The next day, the survival rate of bacteria was measured. The result of this assay clearly demonstrated that the diatom protein displays (6-4) photolyase activity *in vivo* similar to a *bona fide* (6-4) photolyase from zebrafish (Fig. 22).
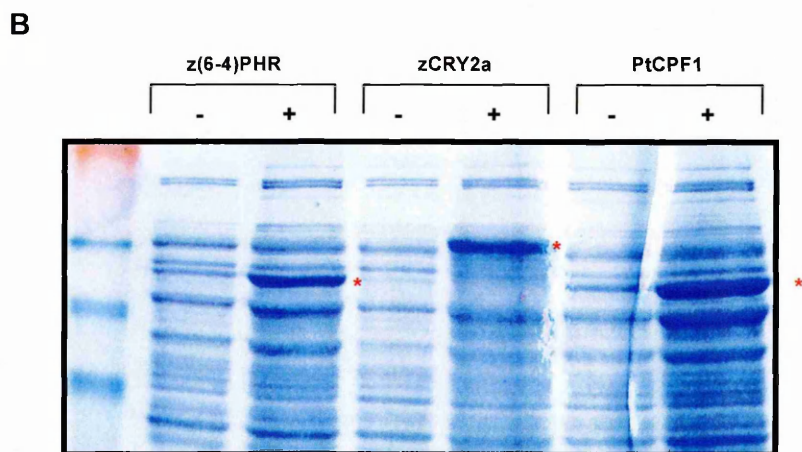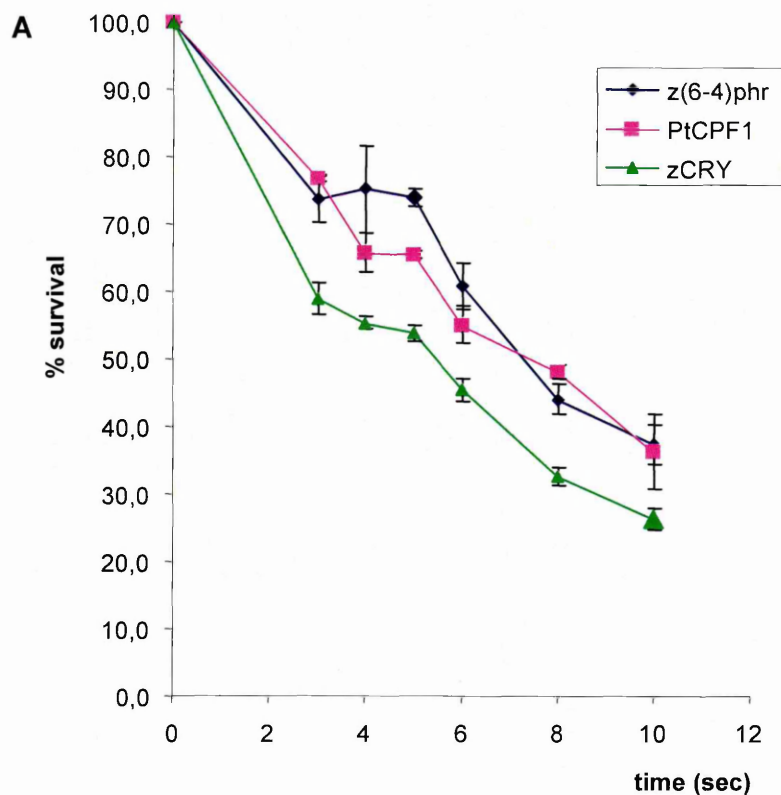
Fig. 22. A. *In vivo* DNA photolyase assay. Effect of photoreactivation on the survival of UV-irradiated *E. coli* SY32 cells carrying the pGEX-2TK-PtCPF1 vector (square), pGEX-4T-2-z(6-4)PHR vector (diamond) and pGEX-4T-2-zCRY2a (triangle). The bars indicate standard error of three independent experiments. **B.** Coomassie Blue stained SDS-PAGE gel (10% polyacrylamide). Lane 1: molecular weight marker. Not induced (-) and induced (+) extracts of *E. coli* transformed whit pGEX-4T-2-z(6-4)PHR (lanes 2-3), pGEX-4T-2-zCRY2a (lanes 4-5) and pGEX-2TK-PtCPF1 (lanes 6-7). Induced proteins are indicated by red asterisks.

3.3.6 Subcellular localization of PtCPF1 in transgenic diatoms

In order to identify the cellular localisation of PtCPF1, a construct was generated containing EYFP fused to PtCPF1 at the N-terminus, denoted pEYFP:PtCPF1 (see Materials and Methods). EYFP is an enhanced yellow variant of the *Aequorea victoria* green fluorescent protein (Ormo M. *et al.*, 1996) that shows a yellow fluorescence clearly distinguishable next to the red chlorophyll fluorescence. Transgenic diatoms overexpressing this resulting protein were generated by shooting the plasmid DNA in the cells by helium-accelerated particle bombardment at high pressure, as described in Materials and Methods. The presence of the EYFP:PtCPF1 fusion protein of the correct molecular weight in transgenic diatom lines was confirmed by western blotting experiments using both the α-PtCPF1 and the α-GFP antibodies (data not shown). Confocal fluorescence microscopy on positive EYFP clones showed that PtCPF1 is localised in the nucleus (Fig. 23). Interestingly, the yellow fluorescence signal of EYFP:PtCPF1 appears to be localized specifically in the euchromatic nuclear region, suggesting that the fusion protein is correctly localized and probably active (Fig. 23C). Further localization analyses were performed in order to understand whether the protein changed its localization under different light treatments, as is known for plant cryptochromes (Yang H.Q. *et al.*, 2000). Transgenic diatom cells were observed in the morning, during the night period and in free running dark conditions for 48 or 60 hours. It was found that the protein of interest was constitutively nuclear localized, independent of the light regime (data not shown). Moreover, the diatom Gateway Destination vector pDEST-C-EYFP was utilized to fuse the fluorescent protein to the C-terminal extremity of PtCPF1. Positive transgenic lines containing the PtCPF1:EYFP construct were subject to microscope analysis, but no signal was detected (data not shown) suggesting that this protein was not functional and was

degraded. As a control, diatom lines overexpressing the *EYFP* gene alone were generated and microscopic analysis revealed a diffuse signal, present in both the nucleus and the cytoplasm (data not shown).
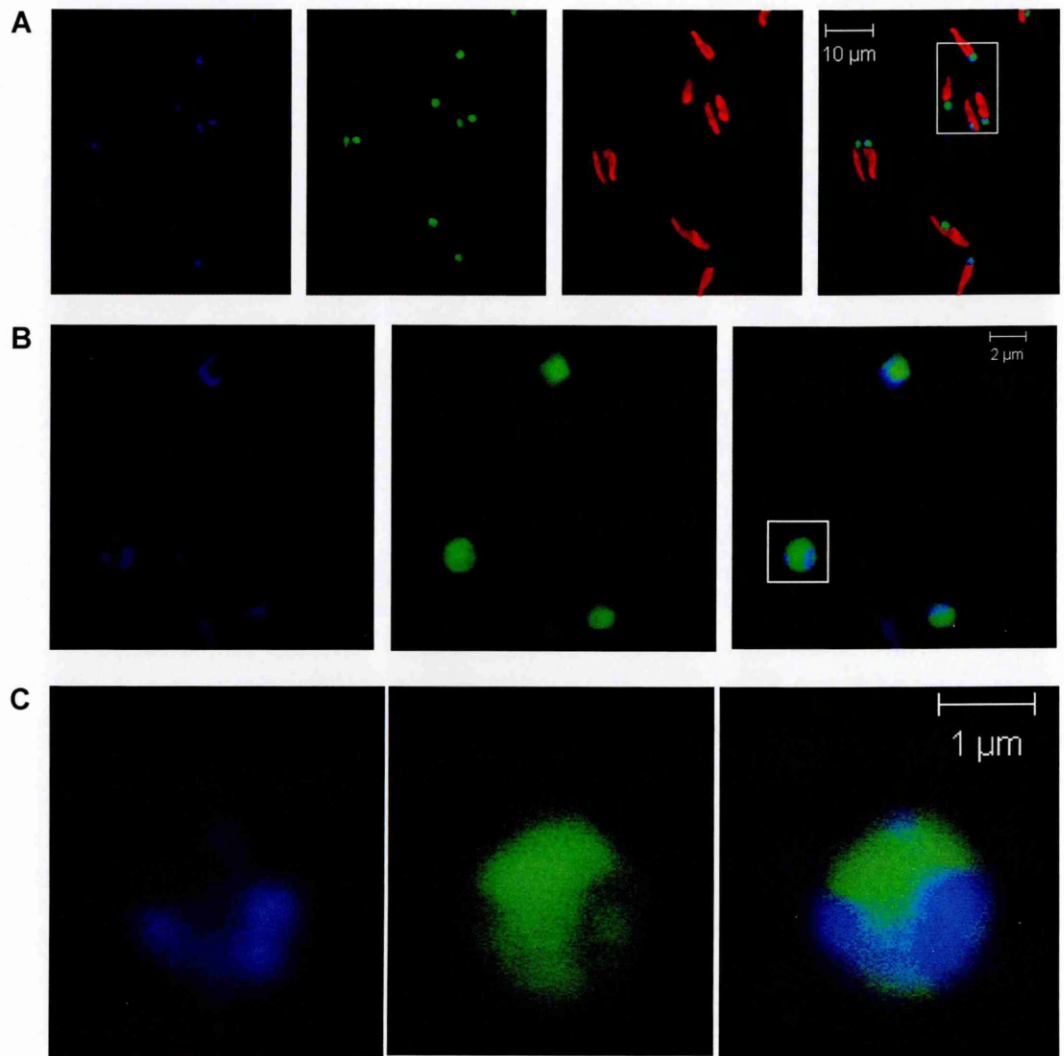
Fig. 23. Confocal microscopy of *P. tricornutum* cells expressing EYFP:PtCPF1 fusion protein. The blue signal represents DAPI staining, the green signal shows EYFP, and the merge is represented in the right panels. Scan zoom levels are 1x (**A**), 5x (**B**), 12x (**C**). EYFP:PtCPF1 is nuclear localized and present within the euchromatic space, whereas DAPI localized principally to heterochromatin.

3. 4 Expression Analysis of *PtCPF1*

3.4.1 Generation of an antibody against PtCPF1

To understand the function of a gene it is important to be able to follow the corresponding endogenous protein in its *in vivo* context. Based on this consideration effort was invested to produce a specific antibody against the diatom PtCPF1 protein.

The expression vector pGEX-2TK-PtCPF1 encoding the GST:PtCPF1 fusion protein (see section 3.3.1) was transfected in the *E. coli* strain BL21 in order to find good induction conditions for production of the diatom protein. In all conditions tested, GST:PtCPF1 was never found in the soluble fraction but always accumulated in inclusion bodies. When the intent is to produce a protein for the generation of antibody it is possible to purify *E. coli* inclusion bodies using denaturing reagents. The advantage of this is that, after purification, the protein of interest is almost pure, albeit no longer active.

For purification of inclusion bodies containing the recombinant PtCPF1 protein a lysis buffer containing 8 M urea was utilized (details described in the Materials and Methods section). The purified protein sample was then sent to the Biopat company (Caserta, Italy) to be injected in a rabbit. After three injections, necessary to stimulate the antigenic response, the rabbit serum was recovered and tested by western blotting on diatom protein extracts. A band of the expected size was revealed but with high crossreactivity (Fig. 24B, right panel). Because the serum is a mixture of proteins containing the specific immunoglobulins against the injected protein as well as a rich pool of different molecules, that may be responsible for the high background, the antibody was purified from the serum. Recombinant GST:PtCPF1 was therefore resolved by SDS-PAGE and transferred to nitrocellulose

membrane in order to isolate the antigen band. The membrane was then incubated with the rabbit serum overnight at 4 °C to permit binding of the specific antibody. A final step was to strip the diatom antibody by treating the membrane with a solution at acid pH. The purified antibody, denoted α-PtCPF1, revealed a specific band that corresponded to the PtCPF1 protein in diatom protein extracts (Fig. 24B, left panel).
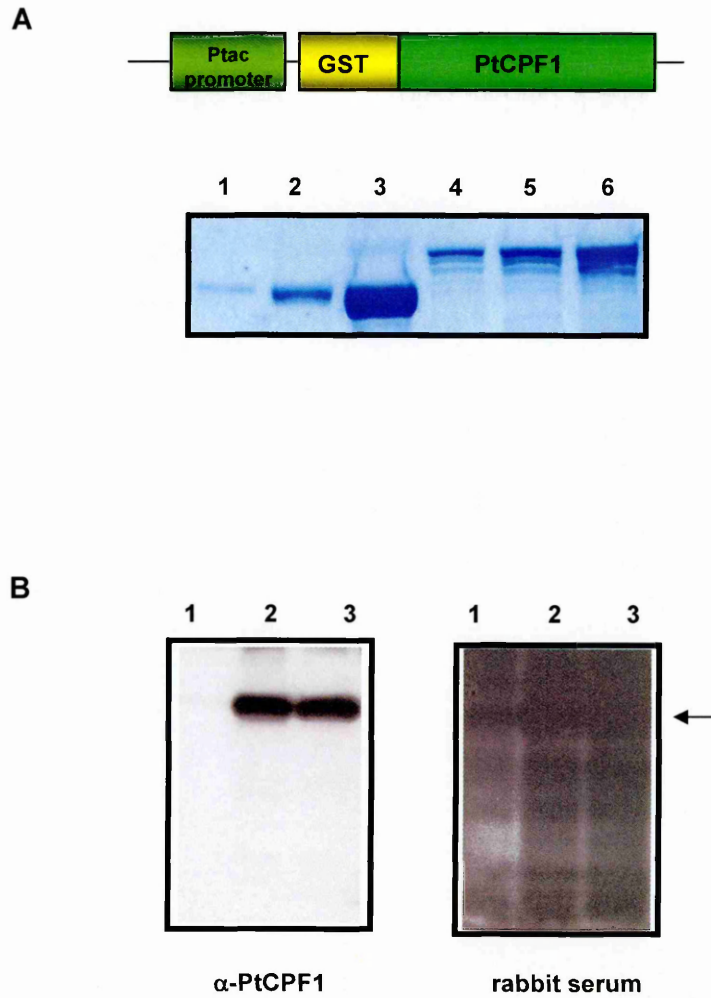
Fig. 24. A. Schematic of the pGEX-2TK-PtCPF1 vector. Coomassie Blue stained SDS-PAGE gel (10% polyacrylamide) showing 1 µg, 2 µg and 10µg of BSA protein (lanes 1 to 3 respectively) and 1.2 µg, 3.6 µg and 6 µg of purified GST:PtCPF1 protein (lanes 4 to 6 respectively). **B.** Western blotting using purified α-PtCPF1 antibody (left panel) and the rabbit serum (right panel). Same exposure time for both filters (1 minute). Protein extracts of wild type diatom cells (lane 1) and two diatom transgenic lines overexpressing *PtCPF1* (lanes 2, 3).

3.4.2 Studies of expression of *PtCPF1*: circadian analysis and acute light response

To characterize *PtCPF1* gene expression, time course experiments were performed aimed at following PtCPF1 mRNA abundance during a diurnal cycle. First of all, semi-quantitative RT-PCR analysis was performed on RNA samples extracted from diatom cells that had been collected every four hours for 36 hours in a normal diurnal cycle with a 12 hour photoperiod. For these experiments, two sets of primers annealing in the 5' and in the 3' region of the mRNA were used together with primers for the *FCPB* gene that is known to be strongly expressed during the light period and almost absent during the night, as shown in Fig. 25. The results of this analysis were rather interesting because amplification of the 5' region of *PtCPF1* resulted in a single fragment, whose concentration was essentially stable throughout the time course (Fig. 25B), whereas amplification of the 3' region resulted in the amplification of several DNA fragments, some of which appeared to oscillate during the time course (Fig. 25C). The lower band corresponded to the size of the mature transcript. The other bands appeared most likely to represent unspliced transcripts because introns are present in this region.

By sequence analysis, the presence of at least three different mRNA species was detected. As expected, the lowest band corresponded to the mature fully processed mRNA and was more abundant in the light period. The uppermost band corresponded to the immature mRNA, containing all four introns, and is predominant during the night, whereas a third transcript, that is always present, corresponds to an incompletely spliced mRNA that in the majority of cases contained the last intron, number 4. It should be noted that the last intron contains an in-frame stop codon, so we can speculate the existence of two proteins during the diurnal cycle or the existence of a mechanism of splicing regulation. These possibilities are further discussed in the Discussion section.

To further explore these observations, we repeated the time course experiment and collected cells every three hours for two days through the normal light-dark period. The RT-PCR was repeated using only the 3' primer pair, and this more accurate analysis revealed at least 5 intermediate mRNAs, confirming oscillation of the mature mRNA during the diurnal cycle (Fig. 25D).

The next experiment was to understand if PtCPF1 mRNA levels oscillated in circadian time. To clarify this point it is necessary to analyse the expression pattern of the gene of interest in cells grown in a normal diel cycle and in free-running conditions. The rhythmic expression of a circadian regulated gene should persist under constant conditions (Más P., 2005).

A pilot time course experiment was performed using diatom cell cultures adapted to a normal 12 hour photoperiod and then shifted to continuous white light or continuous darkness for two days, collecting samples every three hours. In order to perform a more quantitative study a more precise approach based on quantitative real time PCR (qRT-PCR) was used. In parallel the levels of the endogenous PtCPF1 protein was examined by western blotting, using the specific polyclonal antibody α-PtCPF1.

During a normal diurnal cycle the PtCPF1 transcript displays maximal levels at 4PM in the afternoon and a minimum at around 1AM during the night. During the night we also observe a slow increase in abundance suggesting an anticipation of the approaching light period (Fig. 26A). In free running light conditions, we did not observe these expression patterns and mRNA levels no longer appeared to be responsive to light (Fig. 26B). In free-running darkness, on the other hand, the expression pattern was similar to that seen during a normal diel cycle. During the first 24 hours mRNA levels increase during the subjective day and then decrease during the subjective dark period, and this profile, albeit less strong, was conserved

through the second subjective day (Fig. 26C). This circadian control in darkness implies that light inputs are not necessary for rhythmicity in PtCPF1 mRNA levels, suggesting regulation by a light-independent pathway.

Western blotting analysis using the α-PtCPF1 antibody shows a faint oscillation of the protein between the light and dark period. However, this oscillation is not strongly conserved under constant conditions (neither continuous light nor continuous dark), indicating the presence of only a weak circadian regulation or the existence of a post-translational regulation that is undetectable by western blotting (Fig. 27).

In order to understand whether *PtCPF1* is regulated by acute light signals and to examine light quality dependence, diatom cells were adapted for two days in the dark and then exposed to white, blue or red light. After dark adaptation for 48 hours, blue light was clearly seen to induce the expression of *PtCPF1*, with a maximum at 5 hours, whereas white light and red light had no effect on its expression (Fig. 28A). PtCPF1 protein levels were similarly modulated by blue light, and also to some extent by red light. Finally, although mRNA levels were not altered in response to white light, the PtCPF1 protein showed a clear increase in these conditions (Fig. 28B).
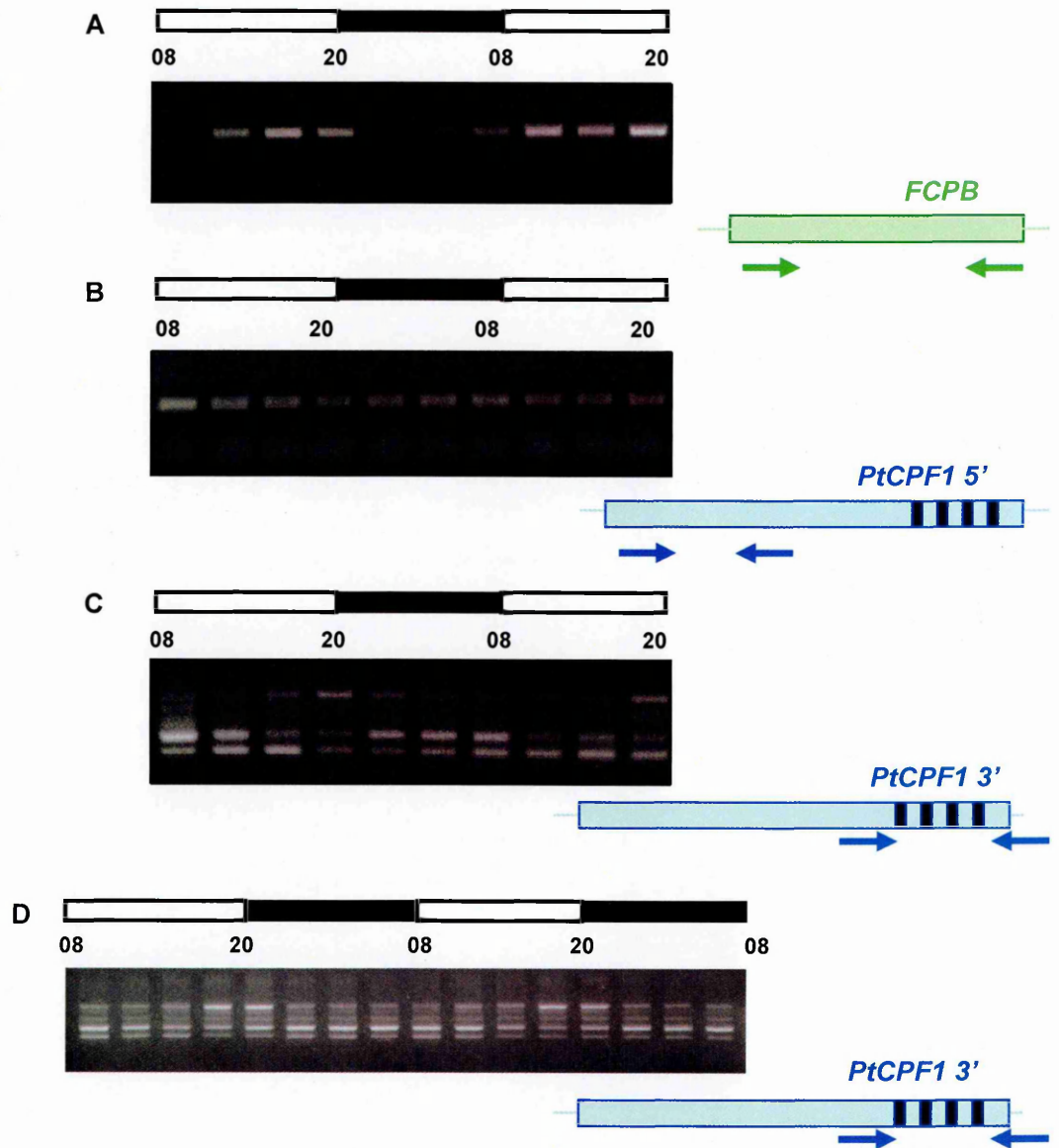
Fig. 25. Semi-quantitative RT-PCR analysis on RNA samples from diatom cells grown in a normal light-dark cycle, collected every four hours. (**A**) FcpB transcript, (**B**) 5′ region of *PtCPF1* transcript, (**C**) 3′ region of *PtCPF1* transcript, (**D**) Semi-quantitative RT-PCR analysis on RNA samples from diatom cells grown in a normal light-dark cycle, collected every three hours, 3′ region of *PtCPF1* transcript.
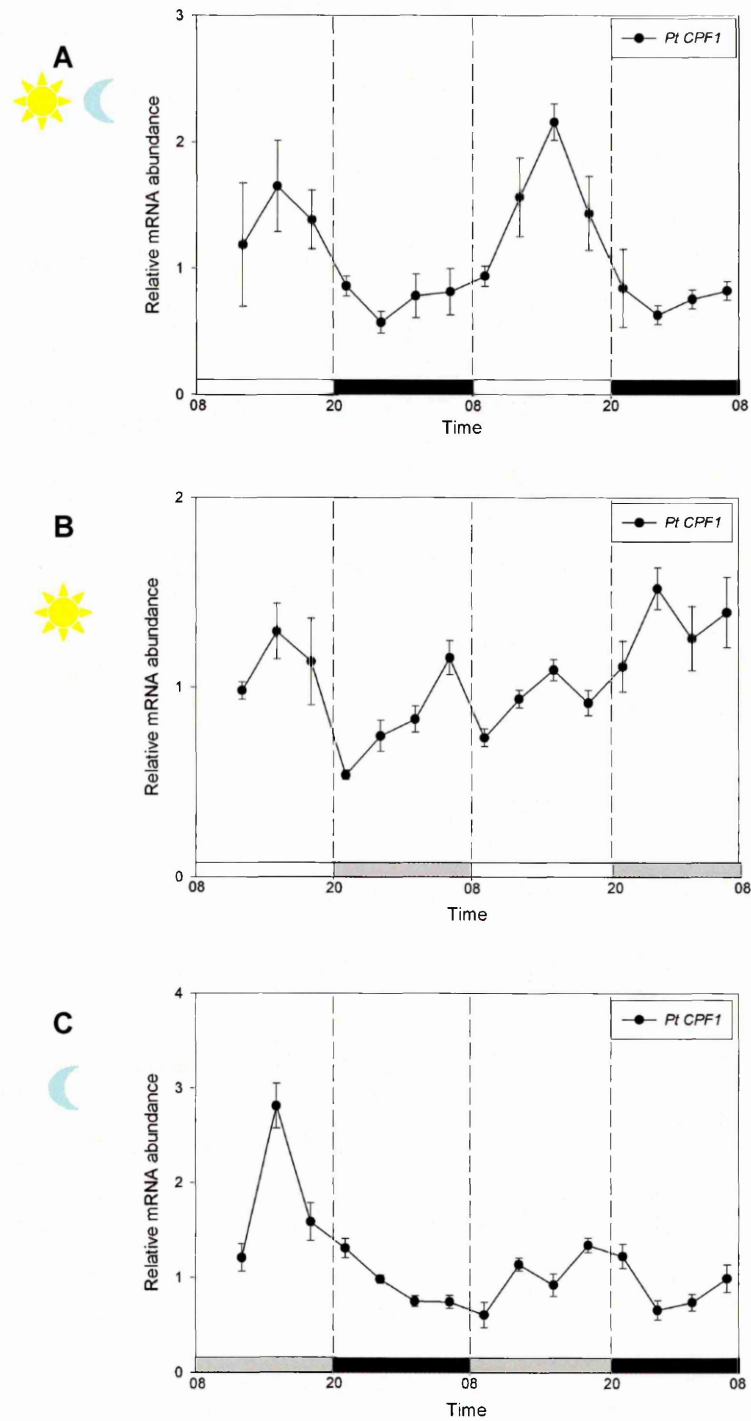
Fig. 26. Quantitative real time PCR on RNA samples extracted from diatom cells grown in a normal light-dark cycle (**A**), in continuous light (**B**), or in continuous dark (**C**), collected every three hours. The bars indicate standard errors of two independent measurements. Data are averages of triplicate measurements of 2 independent cDNAs (A, B) and of 1 cDNA (C). The error bars indicate the standard deviation.
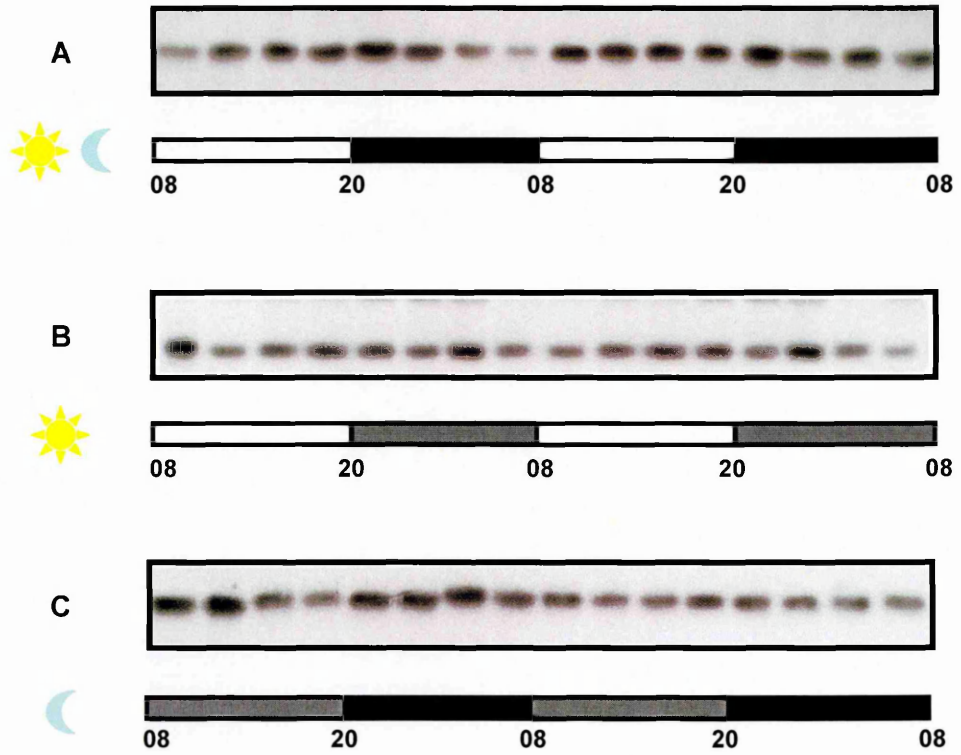
Fig. 27. Western blotting experiments using α-PtCPF1 on protein samples extracted from diatom cells grown in a normal light-dark cycle (**A**), in continuous light (**B**), or in continuous dark (**C**), collected every three hours.
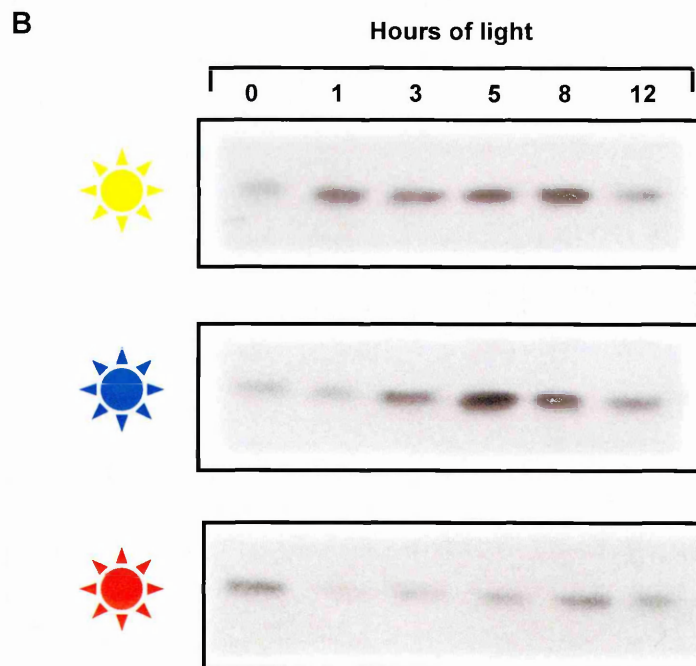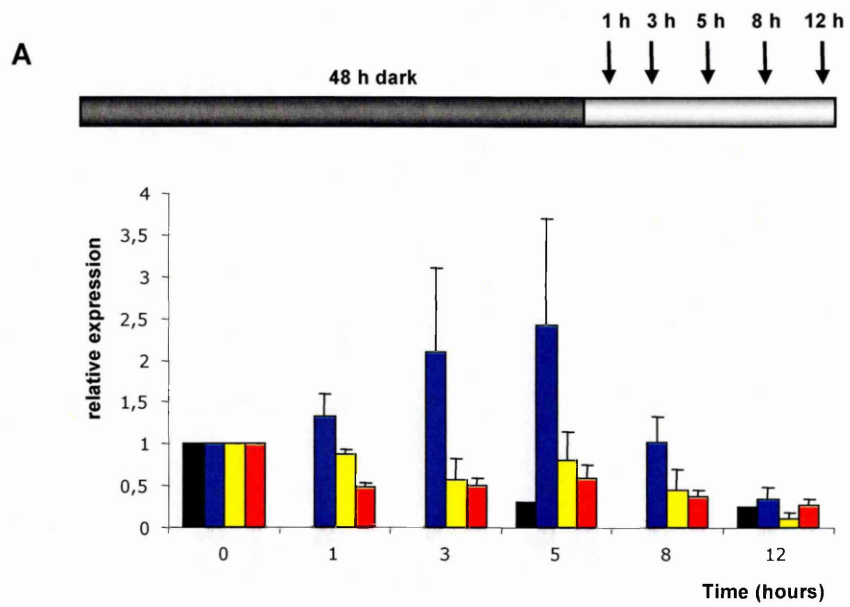
Fig. 28. Quantitative real time PCR (**A**) and western blotting experiments using α-PtCPF1 (**B**) on samples extracted from 48 hours dark-adapted diatom cells (point 0) subsequently exposed to white, blue, and red light. The bars in panel **A** indicate standard errors of two independent measurements.

3. 5 Functional characterization of *PtCPF1*

3.5.1 Detection of transcriptional repressor activity

As previously discussed in the Introduction (section 1.3.7.2), mammalian cryptochromes are components of the negative-feedback loop of the circadian clock and their critical role is the inhibition of CLOCK:BMAL1-mediated transcription (Kume K. *et al.*, 1999). To examine this function, a luciferase reporter gene assay in mammalian cells is generally used. The reporter construct utilizes the promoter region of a clock regulated gene that carries a CACGTG E box enhancer. CLOCK and BMAL1 molecules together bind this sequence in target promoters and activate expression. In the presence of the CLOCK:BMAL1 repressors luciferase expression is thereby abolished.

Based on this knowledge, we decided to perform an ambitious experiment. The idea was to test if the diatom PtCPF1 protein could play a functional role similar to the mammalian cryptochromes in a heterologous system (see scheme shown in Fig. 29). We tested PtCPF1 protein function in two different mammalian cells, the monkey cell line COS7 and the BRF41 zebrafish cell line. In both experiments the reporter construct was made by cloning a 3,700 bp fragment of the 5' flanking region of the *zcry3* gene upstream of a modified coding region of firefly luciferase, generating the pzcry3-luc vector. Moreover, for the COS7 cells the CLOCK-BMAL1 molecules from mouse (mCLOCK, mBMAL1) were used, together with the mouse cryptochrome 1 (mCRY1) as a positive control for repressor activity, whereas for the BRF41 cells the CLOCK and BMAL genes were from zebrafish (zCLOCK1 and BMAL3), with the zebrafish cryptochrome 1a as a positive control (zCRY1a).

The results of these experiments were exciting. Surprisingly, the diatom protein strongly repressed CLOCK:BMAL1-mediated transcription to the same level

as mCRY1 in COS7 cells (Fig. 29). In zebrafish cells, PtCPF1 protein also shows repressor activity, albeit weaker than that of zCRY1a used as the positive control (Fig. 30). Based on these results, it could be hypothesized that the diatom protein acts as a general repressor in BRF41 cells. In conclusion, these results suggested that we have identified a new member of the Cryptochrome/Photolyase Family, a molecule that has double functions: a specific DNA repair activity and a transcriptional repressor regulation activity.
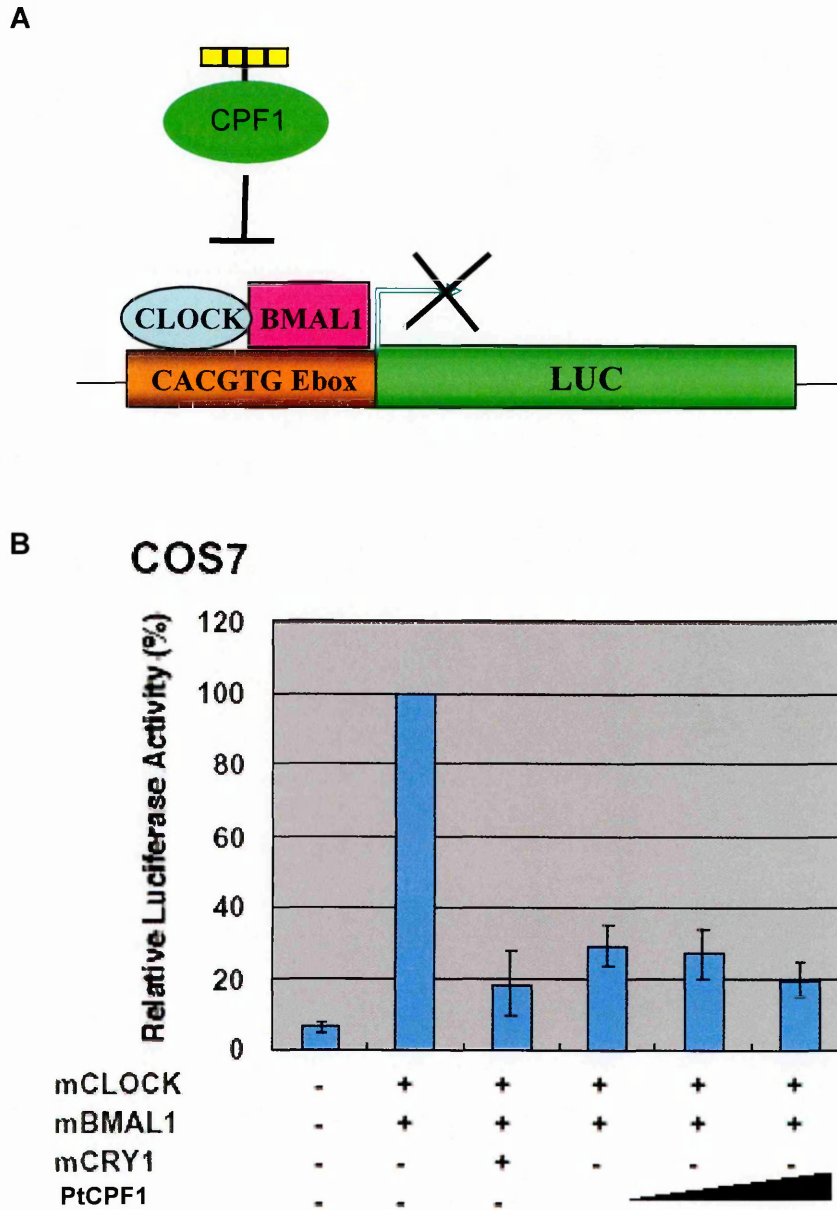
Fig. 29. A. Scheme illustrating the inhibition of CLOCK:BMAL1-mediated transcription. **B.** Transcriptional repressor activity of PtCPF1, determined by the luciferase reporter gene assay. Monkey COS7 cells were transfected with the pzcry3-luc reporter plasmid (50 ng) and the expression vectors shown (200 ng each). Transactivations of the reporter plasmid were examined. Values are mean ± S.E. of three independent experiments. In each experiment, the luciferase activity of the mCLOCK:mBMAL1-containing sample was taken as 100% (◄ = Increasing amount).
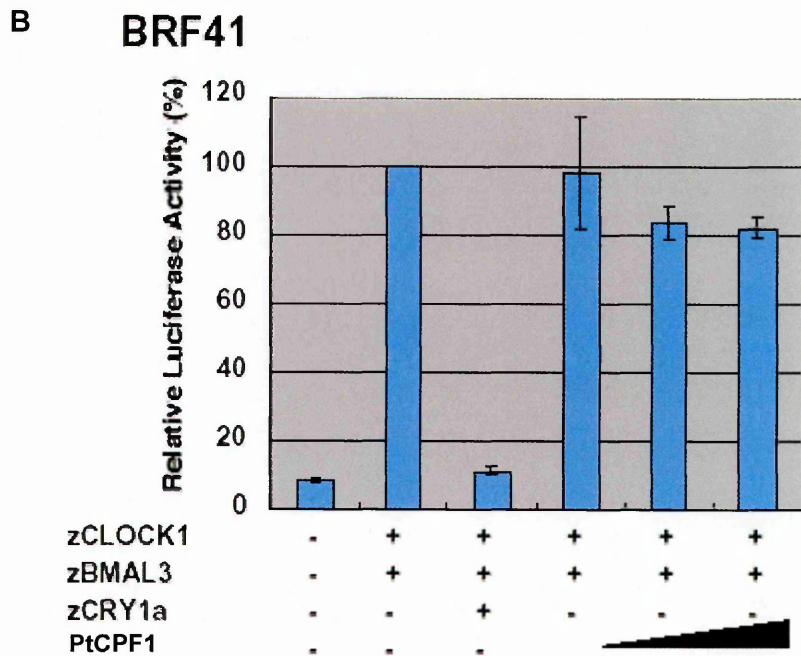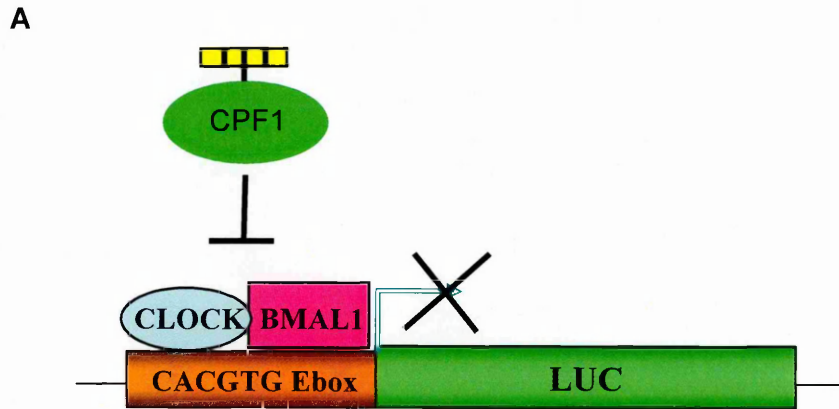
Fig. 30. **A.** Scheme illustrating the inhibition of CLOCK:BMAL1-mediated transcription. **B.** Transcriptional repressor activity of PtCPF1, determined by the luciferase reporter gene assay. Zebrafish BRF41 cells were transfected with the pzcry3-luc reporter plasmid (50 ng) and the expression vectors shown (200 ng each). Transactivations of the reporter plasmid were examined. Values are mean ± S.E. of three independent experiments. In each experiment, the luciferase activity of the zCLOCK:zBMAL3-containing sample was taken as 100% (◄ = Increasing amount).

3.5.2 Generation of diatom lines overexpressing *PtCPF1*

In order to study the role of *PtCPF1* in diatoms, transgenic lines were generated using the pKS-FcpBpAt-PtCPF1 construct (described in the Materials and Methods section) to overexpress *PtCPF1* from the strong FcpB promoter. Independent clones were selected by Western blotting (Fig. 31A) and further studies were performed on three clones denoted c2, c4 and c15 (c = overexpressing CPF1). In cultures adapted to 48 hours of dark and then exposed to 1 hour of white light (200 $\mu mol.m^{-2}.s^{-1}$), the PtCPF1 protein levels were estimated to be 33.8, 28.3 and 35.4 fold higher in c2, c4 and c15, respectively, as compared to wild-type cells (Fig. 31B). Analysis of PtCPF1 content in cells subjected to various light treatments showed that the protein levels remained relatively constant over time and light regime indicating that the PtCPF1 protein is somewhat stable (data not shown).
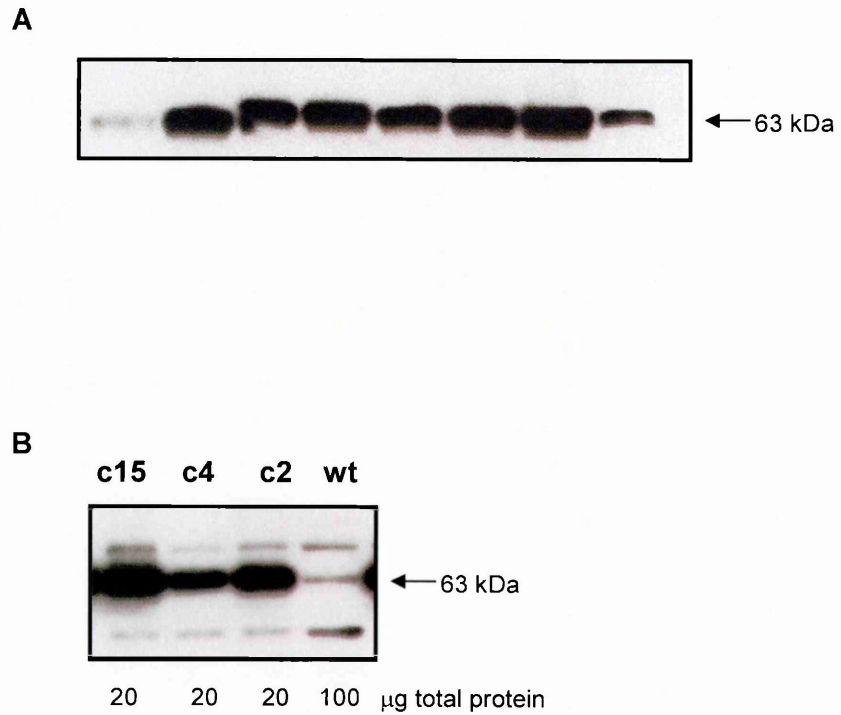
Fig. 31. A. Western blotting using α-PtCPF1 antibody on diatom cell extracts. Wild-type (lane 1) and diatom transgenic lines overepressing *PtCPF1* were quantified by cell numbers. **B.** Western blotting using α-PtCPF1 antibody on protein extracts. The content of PtCPF1 in the three selected transgenic lines was quantified (see text).

3.5.3 Transcriptional regulation of light-induced genes in *PtCPF1* overexpressing lines

Cryptochromes in animals and in plants are known to function as photoreceptors and or as regulators in the entrainment of the circadian clock by light. These proteins must play a pleiotropic role modulating several pathways, as is expected for a gene integrated in the light signalling. Based on these considerations and on our previous results in which the PtCPF1 protein was found to be able to mimic the cryptochrome animal-type photoreceptor by repressing CLOCK:BMAL1-mediated transcription (section 3.5.1), we decided to look for transcriptional regulation of target genes in diatoms using the overexpressing lines.

A crucial step was to define the light conditions in which to perform the analysis, in order to select for specific and rapid responses. We decided to focus attention on acute light responses to eliminate possible redox effects or feedback from photosynthesis. Wild-type diatom cells together with two selected overexpressing lines (c4 and c15), grown until exponential phase in a normal diel cycle, were adapted in dark for 60 hours, and were then exposed to a pulse of 5 minutes of blue light, and, immediately after, placed back in darkness. Diatom cells were collected after 15 and 30 minutes of the dark period subsequent to the light pulse. In addition, in order to design a fluence rate curve, the experiment was repeated with three increasing light intensities: 0.02, 3.3, and 25 $\mu mol.m^{-2}.s^{-1}$.

A second step was to choose genes putatively targeted by *PtCPF1* in order to check if their expression was altered in the selected overexpressing lines. It is known that the general role of a photoreceptor is to regulate several pathways. In particular, plant CRY overexpressors show a high-pigment phenotype, resulting in overproduction of anthocyanins and chlorophyll in leaves and of flavonoids and lycopene in fruits (Giliberto L. *et al.*, 2005; Lin C. & Shalitin D., 2003).

Furthermore, a parallel analysis performed in our laboratory has shown an altered pigment content of *PtCPF1* overexpressing lines and an increased photoprotection response (data not shown). We therefore decided to focus on the carotenoid biosynthesis pathway. The first two enzymes specifically committed to the carotenoid biosynthetic pathway are *phytoene synthase* (*Psy*) and *phytoene desaturase* (*Pds*) which convert two geranylgeranyl diphosphate (GGDP) molecules into phytoene and desaturate the latter into ζ-carotene, respectively. *Psy* and *Pds* are under transcriptional control in response to environmental stimuli and are considered to play key roles in the regulation of carotenogenesis in higher plants (Steinbrenner J. & Linden H., 2003; Chew A.G. & Bryant D.A., 2007).

In parallel, a putative cry-DASH member of the Cryptochrome/Photolyase Family, *PtCPF25*, was analyzed because previous experiments have shown a strong induction of this gene upon blue light exposure (data not shown). In addition, the *FcpB* gene was selected because it is well studied in *P. tricornutum*. The transcript shows a strong induction in a diel cycle (Siaut M. *et al.*, 2007) and upon one hour treatment of continuous blue light at 25 μmol.m^{-2}.s^{-1} (data not shown). Finally, since the cell cycle is regulated by light in diatoms (Vaulot D., 1986), a putative *Cyclin* gene (*CYC*), that is thought to encode a key component of the G1/S checkpoint, was tested in order to look for differences between the wild-type and overexpressing lines.

Quantitative real time PCR (qRT-PCR) experiments were performed in order to compare gene expression in wild-type cells against the two overexpressing lines. In every quantitative PCR experiment, the *RPS* (30S Ribosomal Protein Subunit) transcript was used as endogenous and constitutive reference gene (details in the Materials and Methods section).

As can be seen in Fig. 32, the *Psy* and the *Pds* transcripts are rapidly induced by the acute blue light treatment, showing the strongest increase at 15 and 30 minutes, respectively, after exposure to 3.3 $\mu mol.m^{-2}.s^{-1}$, in both wild-type and overexpressors. In the two *PtCPF1* overexpressing lines the level of induction was significantly dampened, suggesting a negative regulation mediated by PtCPF1. *Psy* transcript levels were around two-fold less induced in the c4 and c15 clones, and the same trend, albeit less strong, was observed for the *Pds* transcript (panels A, B). Surprisingly *FcpB*, that is known to be strongly upregulated upon continuous blue light treatment, was only very weakly induced by exposure to a blue light pulse and was unaffected in the *PtCPF1* overexpressing lines, suggesting that the photosynthetic redox state controls *FcpB* gene expression more than blue-light photoreceptors (panel C).

The cryptochrome DASH *PtCPF25* showed an enormous induction after the blue light pulses, and followed a fluence rate dependent response, with induction being proportional to the light intensity (note the scale on y axis). More interesting, the overxpression of *PtCPF1* resulted in a hyper-responsiveness of *PtCPF25* to blue light (panel D). Finally, *CYC* transcript levels were rapidly induced at low fluence rate (0.2 $\mu mol.m^{-2}.s^{-1}$) and repressed by increasing fluence intensity. The level of fluence-dependent repression was less strong in the overexpressing lines compared to wild-type (panel E), suggesting that *PtCPF1* could be also involved in the light-regulated cell cycle progression in *P. tricornutum*.

The remarkable conclusion of this analysis was that *PtCPF1* is likely involved in the regulation of several responses in diatoms, specifically mediated by blue light. Based on this consideration, it was interesting to look at gene expression in cells treated with a red light pulse. Wild-type and overexpressing lines were adapted in dark, as previously described, and exposed to a pulse of 5 minutes of red

light at 3.3 and 25 $\mu mol.m^{-2}.s^{-1}$. In this case, overexpression of *PtCPF1* did not result in hypersensitivity of gene expression to red light for any of the genes analyzed, as can be seen in the right inset present in every panel. These results motivated us to perform a high-throughput approach to study the effect of *PtCPF1* overexpression on the full *P. tricornutum* transcriptome.
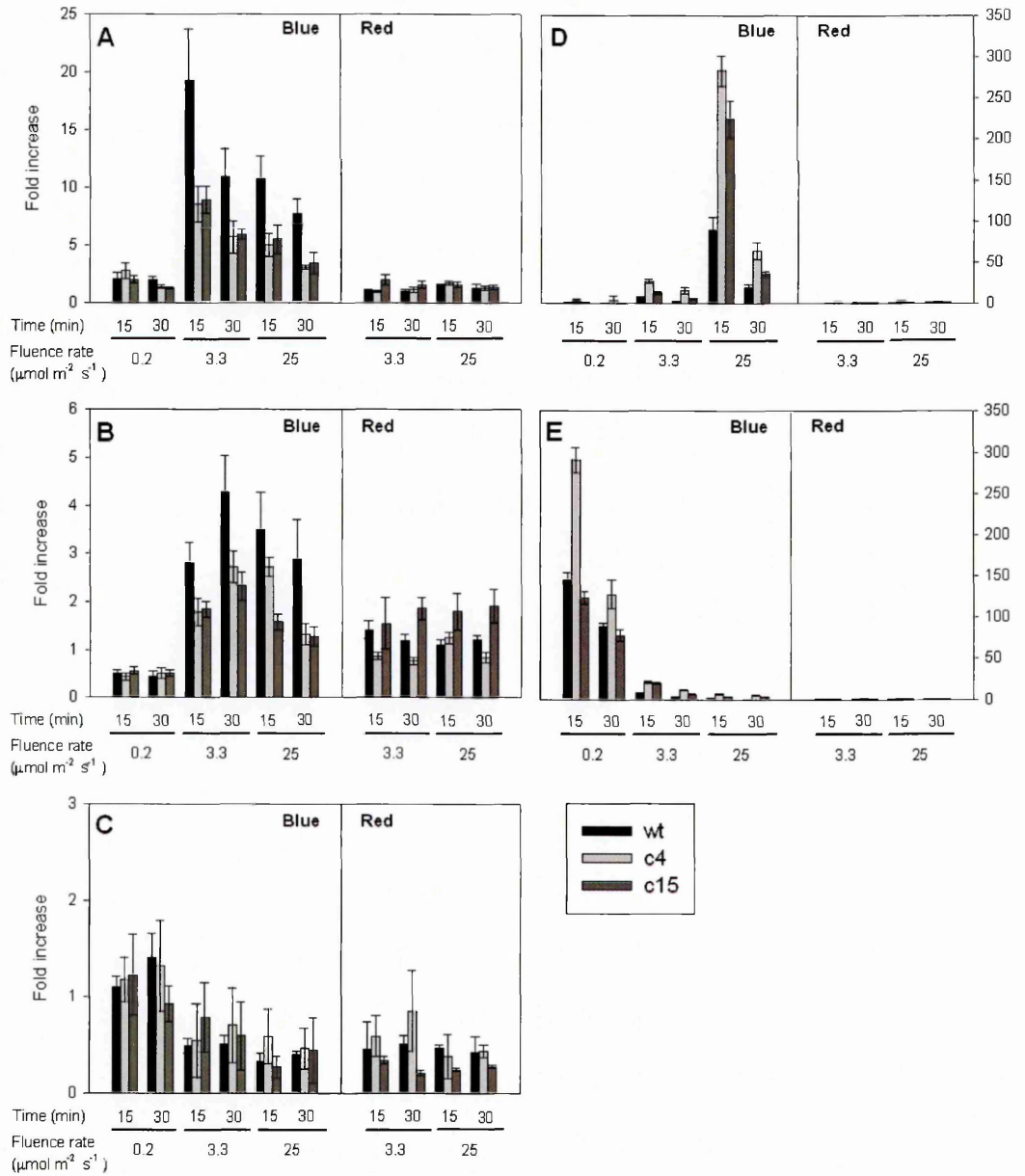
Fig. 32. The effect of blue and red light pulses on *PSY* (A), *PDS* (B), *FcpB* (C), *CPF25* (D), and *CYC* (E) transcript levels in wild type and two *PtCPF1* overexpressing lines using *RPS* as a reference. The values were normalized to the gene expression levels in dark. Data are averages of triplicate measurements of 2 independent cDNAs. The error bars indicate the standard deviation. For further descriptions see text.

3.5.4 Microarray Experiment

As a global approach to identify genes specifically modulated by *PtCPF1*, the strongest overexpressing line (c15, see section 3.5.2) was selected for microarray analysis. To identify genes responsive to an acute blue-light response and to avoid photosynthesis effects, we decided to utilize an intermediate fluence of 3.3 $\mu mol.m^{-2}.s^{-1}$ (*in medio stat virtus*) together with the shortest time (15 minutes), based on the results from qRT-PCR analysis (see section 3.5.3). Three independent diatom cultures of both wild-type and the c15 overexpressing line were exposed to the blue pulse for 5 minutes and then placed back in the dark as described in section 3.5.3. Cells were also collected after 60 hours of darkness in order to compare the gene expression profile of wild-type cells against that of the overexpressing line. This condition was denoted "dark time point". On the other hand, cells collected after the blue pulse were referred to as "light time point", although cells were placed back in darkness for 15 minutes after the light pulse.

Thanks to the availability of the annotated *P. tricornutum* genome sequence, a whole-genome expression array was designed and printed by the RZPD German Resource Center for Genome Research (Berlin, Germany). On average, five different primers of 60 nucleotides in length were designed for each gene. In total 43,860 oligos were spotted on the array that corresponded to 10,364 genes, covering thus the full *P. tricornutum* transcriptome. To minimize the inherent variability of the microarray assay (Lee M.L. *et al.*, 2000) and to ensure the reliability of the results, total RNA was extracted from three independent biological replicas for both conditions, corresponding to a total number of 12 samples (see Materials and Methods). Subsequentely, the "One-Colour Microarray-Based Gene Expression Analysis" was performed because it is generally considered to be more accurate and flexible. In this methodology only one dye (Cyanine 3-CTP) is used to label the

messenger RNA, thus avoiding problems related to the utilization of two different fluorescent dyes (dye-related differences in efficiency of labelling, different laser settings, etc...). Moreover, the one-colour microarray expression analysis allows comparing fluorescence signals from different experiments in an independent manner. In the designed experiment there are two variables: the genetic characteristic (wild-type versus c15) and the treatment (dark versus light). The one-colour microarray expression analysis allowed all possible comparisons to be made.

3.5.5 Preliminary Analysis

In order to find genes differentially expressed in the dark to light transition, dark mRNA levels were used as base line and a fold change threshold of 2.0 was selected together with an error rate of 1%. Wild-type cells displayed 3,057 genes that changed their expression level. In particular 1,466 genes were upregulated in response to the light pulse, whereas 1,591 genes were downregulated. The same comparison in the *PtCPF1* overepressing line displayed similar numbers: 2,801 genes were differentially expressed, of which 1,220 were upregulated and 1,581 were downregulated. To identify genes differentially expressed between wild-type and the *PtCPF1* overexpressing line in the dark to light transition, the two gene lists were compared and from the list of genes in common a group of 61 genes was identified that showed a significant difference in fold increase or decrease between the two genotypes. In particular, 37 genes were upregulated in the overexpressing line (Table 3.1), while 24 genes were downregulated (Table 3.2). A schematic representation of these comparisons is indicated in Fig. 33.

Moreover, comparison of the dark time point between wild-type and the overexpressing line revealed 142 differentially expressed genes, of which 78 were upregulated (Table 3.3) and 64 were downregulated (Table 3.4). Finally, direct

comparison of the light time point showed that just one gene was differentially expressed between the two genotypes. This low number was probably a consequence of the reduced statistical reliability of the three biological replicates from the light treated sample from c15 (data not shown).

To obtain an independent confirmation of the microarray results, several interesting genes that appeared to be differentially expressed from different comparisons were analyzed by qRT-PCR. Table 3.5 summarizes the selected genes together with their relative fold change in different conditions. From the "dark versus light" comparison the *Phatr_48732* gene encoding a putative hydroxylase appeared to be more upregulated in the c15 line with a difference of three fold compared to expression in wild-type cells (Fig. 34). The same tendency, albeit less pronounced, was observed for *Phatr_51703*, encoding a violaxanthin de-epoxidase enzyme, suggesting regulation by *PtCFP1* of pigment synthesis and photoprotection, already supported by physiological data in overexpressing lines (data not shown).

The *Phart_18180* gene, encoding a putative chlorophyll binding protein, displayed a strong induction of 70 fold following the blue light pulse in wild-type cells. In the c15 overexpressing line the level of induction was significantly dampened, in agreement with the microarray data.

Furthermore, analysis of the genes that showed differential expression in the dark could be useful to understand how the basal metabolism is affected in the overexpressing line. The *Phatr_51916* gene, encoding a DNA mismatch repair protein, was less expressed in the *PtCPF1* overexpressing line, whereas two heat-shock transcription factors, *Phatr_49594* and *Phatr_49596*, displayed higher transcript levels (see relative levels in Fig. 33). Again, these results were consistent with the microarray results.

To summarize, the qRT-PCR data generally confirmed the microarray results, which will encourage us to perform a more accurate analysis in the future aimed at identifying signal transduction pathways specifically modulated by *PtCPF1* in order to better elucidate the role of this protein *in vivo* in diatoms.
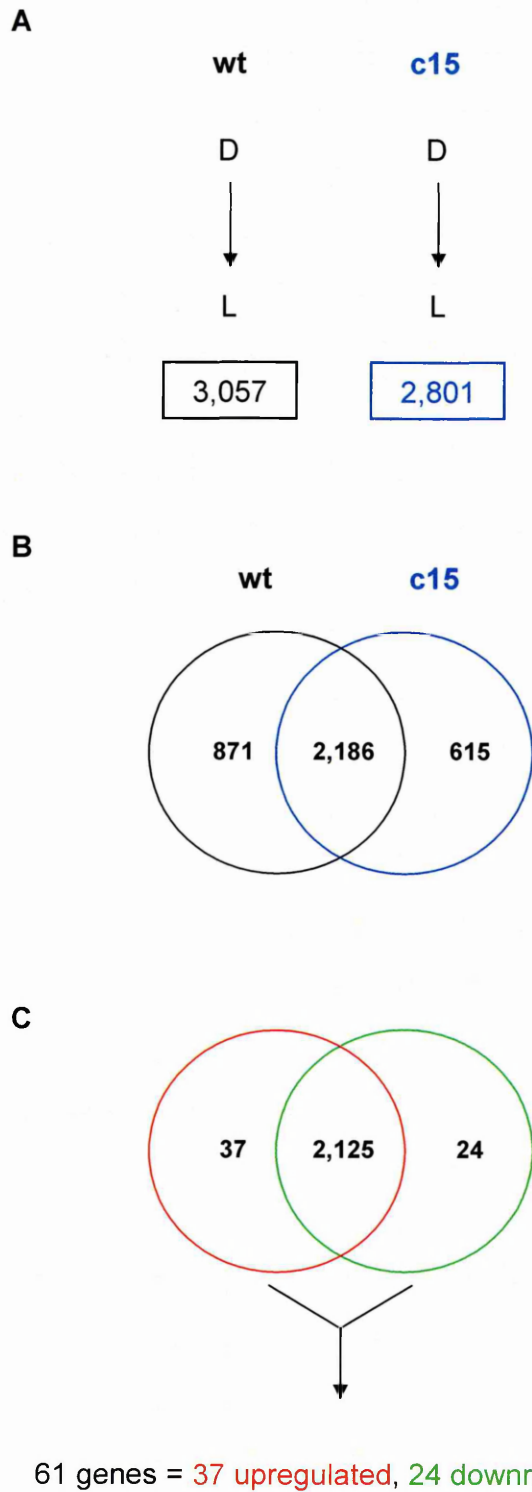
Fig. 33. A. Schematic representation showing the number of genes differentially expressed from the dark (D) to light (L) transition in wild-type cells and in the *PtCPF1* overexpressing line (c15). **B.** Venn diagram comparing the gene lists shown in panel A. **C.** Analysis of the common gene set in the Venn diagram in B. 2,125 genes showed similar expression levels, whereas 37 genes were upregulated in c15, and 24 genes were downregulated in c15. Details are described in the text and the gene lists are provided in Table 3.1 and 3.2.

Table 3.1

Target ID	Ratio in WT	Ratio in c15	Ratio c15/WT	Description
Pathr2_48732	10.78	83.65	7.76	Prolyl 4-hydroxylase
Pathr2_44340	3.69	26.38	7.15	unknown function
Pathr2_10872	32.40	203.85	6.29	unknown function
Pathr2_35311	8.35	48.91	5.86	Nucleoporin
Pathr2_39274	11.18	51.85	4.64	Potassium transporter
Pathr2_12106	5.64	21.87	3.88	Ku70-binding family protein
Pathr2_48545	5.10	19.57	3.84	unknown function
Pathr2_9255	5.45	20.27	3.72	Fatty acyl-CoA elongase
Pathr2_46588	3.98	13.78	3.46	Vacuolar protein sorting
Pathr2_30770	7.84	26.35	3.36	Microsomal cytochrome b5
Pathr2_31423	3.12	10.49	3.36	Beclin-1-like protein
Pathr2_44961	3.30	10.44	3.17	unknown function
Pathr2_25956	11.25	35.21	3.13	Phosphoadenosine-reductase
Pathr2_51703	31.03	95.45	3.08	Violaxanthin de-epoxidase
Pathr2_35009	4.82	14.48	3.00	unknown function
Pathr2_37652	36.59	109.44	2.99	Malonyl transacylase
Pathr2_43259	5.43	15.88	2.92	SNARE protein Syntaxin
Pathr2_11230	29.75	84.79	2.85	Stress-induced protein UVI31+
Pathr2_44864	0.08	0.23	2.78	unknown function
Pathr2_13034	29.03	79.81	2.75	Sideroflexin 5
Pathr2_47593	40.90	111.01	2.71	Putative glossy1 protein
Pathr2_10208	845.27	2196.91	2.60	Phosphoribulokinase
Pathr2_10896	7.43	18.64	2.51	Solute carrier protein
Pathr2_14688	2.65	6.56	2.47	Protein Kinase-like
Pathr2_16803	3.63	8.97	2.47	Protein Kinase-like
Pathr2_49354	7.92	19.38	2.45	Centromere autoantigen C
Pathr2_14403	7.62	18.34	2.41	Sodium/proton exchanger 8
Pathr2_32339	0.10	0.24	2.40	Putative transposase
Pathr2_48145	9.97	23.83	2.39	unknown function
Pathr2_45503	2.42	5.55	2.30	Transmembrane protein
Pathr2_48656	7.15	16.33	2.28	Similar to zinc finger
Pathr2_26948	0.10	0.23	2.28	Tripeptidyl peptidase II
Pathr2_48160	14.43	32.03	2.22	WD40 repeat protein
Pathr2_49956	5.02	11.04	2.20	Dystonin
Pathr2_49339	0.16	0.34	2.07	Pyruvate carboxylase
Pathr2_44153	3.06	6.22	2.04	Esterase/lipase/thioesterase
Pathr2_46862	5.05	10.15	2.01	unknown function

Table 3.1. Gene lists obtained from the dark to light transition between wild-type and the *PtCPF1* overexpressing line (c15) were compared. The table shows the 37 most upregulated genes in c15 compared to wild-type in the common list. Target IDs indicated in blue were analyzed by qRT-PCR.

Table 3.2

Target ID	Ratio in WT	Ratio in c15	Ratio c15/WT	Description
Phatr2_10068	103.68	23.46	0.23	Enoyl-acyl carrier reductase
Phatr2_34672	19.74	4.65	0.24	Hydroxyacid dehydrogenase
Phatr2_14090	7.74	2.55	0.33	unknown function
Phatr2_51708	9.87	3.42	0.35	Flavin amine oxidase
Phatr2_36970	10.70	3.78	0.35	FOG: PPR repeat
Phatr2_44438	9.36	3.58	0.38	unknown function
Phatr2_41092	17.15	6.70	0.39	Mitochondrial carrier protein
Phatr2_5142	16.67	6.71	0.40	Mitochondrial carrier protein
Phatr2_45959	10.54	4.28	0.41	Phosphatase
Phatr2_42018	31.06	12.70	0.41	Ferredoxin reductase
Phatr2_48524	11.95	5.03	0.42	unknown function
Phatr2_48534	11.36	4.87	0.43	Nuclear mRNA export factor
Phatr2_43963	12.77	5.49	0.43	unknown function
Phatr2_6834	8.40	3.65	0.43	Phosphoribosyltransferase
Phatr2_36420	30.27	13.49	0.45	LSU ribosomal protein L9P
Phatr2_51092	188.06	87.01	0.46	glutamine synthetase
Phatr2_44320	7.43	3.46	0.47	unknown function
Phatr2_18180	79.25	37.16	0.47	light harvesting protein
Phatr2_47408	86.86	41.83	0.48	unknown function
Phatr2_43120	11.52	5.57	0.48	Phosphatase
Phatr2_4014	12.93	6.27	0.49	NT02FT1074 Peptidase
Phatr2_39615	12.10	5.95	0.49	unknown function
Phatr2_12813	47.88	23.89	0.50	NADP/FAD oxidoreductase
Phatr2_41878	7.43	3.89	0.52	Squalene/phytoene synthase

Table 3.2. Gene lists obtained from the dark to light transition between wild-type and the *PtCPF1* overexpressing line (c15) were compared. The table shows the 24 most downregulated genes in c15 compared to wild-type in the common list. Target ID indicated in blue was analyzed by qRT-PCR.

Table 3.3

Target ID	Ratio c15/WT	Description
Phatr2_50585	74.25	unknown function
Phatr2_32513	6.25	unknown function
Phatr2_49631	5.84	Carbohydrate transport and metabolism
Phatr2_44525	5.79	unknown function
Phatr2_18049	5.54	Fucoxanthin, chlorophyll protein
Phatr2_49612	4.99	unknown function
Phatr2_49699	4.64	unknown function
Phatr2_49594	4.14	Heat shock transcription factor
Phatr2_49556	4.09	unknown function
Phatr2_9180	3.97	ATPase:ABC transporter
Phatr2_55070	3.97	unknown function
Phatr2_54168	3.72	Galactosyl transferase
Phatr2_49680	3.51	unknown function
Phatr2_39253	3.43	unknown function
Phatr2_12322	3.28	hypothetical protein exoribonucleases
Phatr2_16157	3.12	Hydrolase, NUDIX family
Phatr2_22901	3.03	Cathepsin A
Phatr2_3351	2.97	Chloroplast dimethyladenosine synthase
Phatr2_33896	2.83	Homeodomain transcription factor
Phatr2_40322	2.79	unknown function
Phatr2_40744	2.74	Putative Peptidase
Phatr2_18096	2.71	DNAJ domain protein
Phatr2_50431	2.66	unknown function
Phatr2_49564	2.65	unknown function
Phatr2_36175	2.65	Msx-2 interacting nuclear target protein
Phatr2_27709	2.63	unknown (protein for MGC:69147)
Phatr2_30519	2.63	unknown function
Phatr2_44027	2.63	hypothetical protein DDB0215059
Phatr2_47766	2.57	2OG-Fe(II) oxygenase superfamily
Phatr2_47300	2.56	unknown function
Phatr2_49989	2.46	unknown function
Phatr2_48731	2.45	Metalloendopeptidase family
Phatr2_49038	2.44	unknown function
Phatr2_44092	2.40	unknown function
Phatr2_34884	2.36	Hydrolase
Phatr2_49589	2.36	Kinesin (SMY1 subfamily)
Phatr2_16120	2.36	hypothetical protein DDBDRAFT_0188074
Phatr2_47400	2.35	unknown function
Phatr2_49557	2.34	Heat shock transcription factor
Phatr2_49608	2.34	FeS assembly protein SufD
Phatr2_10102	2.34	RNA-binding region RNP-1
Phatr2_12896	2.32	Prolyl 4-hydroxylase
Phatr2_48279	2.31	Glycosyl transferase, family 2

Phatr2_45677	2.31	Putative serine/threonine protein kinase
Phatr2_49610	2.30	Exo-1,3-beta-glucosidase, putative
Phatr2_49651	2.30	unknown function
Phatr2_49227	2.30	unknown function
Phatr2_9312	2.29	RNA polymerase sigma factor
Phatr2_45342	2.29	unknown function
Phatr2_43956	2.28	unknown protein
Phatr2_42635	2.27	Putative oxidoreductase /thioredoxin
Phatr2_49668	2.27	unknown function
Phatr2_47142	2.25	unknown function
Phatr2_49596	2.25	Heat shock transcription factor
Phatr2_46603	2.23	unknown function
Phatr2_45185	2.23	hypothetical protein CPS_0799
Phatr2_48287	2.22	Glycosyl transferase, family 2
Phatr2_40956	2.22	Chaperonin complex component
Phatr2_36970	2.22	Protein with PPR repeat
Phatr2_49626	2.19	unknown function
Phatr2_49223	2.18	Putative Peptidase
Phatr2_49657	2.18	unknown function
Phatr2_14373	2.15	Predicted small membrane protein
Phatr2_26970	2.15	NADH dehydrogenase (ubiquinone)
Phatr2_48329	2.15	unknown function
Phatr2_54405	2.14	unknown function
c80045-79650	2.12	unknown function
Phatr2_49595	2.11	E3 ubiquitin protein ligase
Phatr2_10068	2.10	enoyl-acyl carrier reductase
Phatr2_42949	2.09	unknown function
Phatr2_54952	2.09	Pyridoxamine 5'-phosphate oxidase-related
Phatr2_33757	2.05	Galactosyl transferase
Phatr2_50605	2.04	hypothetical protein
Phatr2_23467	2.04	Methylmalonate-semialdhyde dehydrogenase
Phatr2_48891	2.03	2OG-Fe(II) oxygenase
Phatr2_47200	2.01	unknown function
Phatr2_40831	2.00	F38A5.2a
Pt1_bottom_35003	2.00	unknown function

Table 3.3. Gene list obtained by comparing the dark time point between wild-type and the *PtCPF1* overexpressing line (c15). The table shows the 78 genes upregulated in c15. Target IDs indicated in blue were analyzed by qRT-PCR.

Table 3.4

Target ID	Ratio c15/WT	Description
Phatr2_40650	0.49	Aminotransferase
Phatr2_45417	0.49	RNA-directed RNA polymerase
Phatr2_47761	0.49	Potential chitinase or glucoamylase
Phatr2_43029	0.49	Transcription elongation factor
Phatr2_3131	0.48	Predicted GTP-binding protein
Phatr2_50562	0.48	Splicing coactivator
Phatr2_50470	0.48	Zn-finger, MYND type
Phatr2_50445	0.48	phosphotransferases/phosphoglucomutase
Phatr2_33844	0.48	unknown function
Phatr2_50527	0.47	Nucleolar GTPase/ATPase p130
Phatr2_43234	0.47	hypothetical protein FjohDRAFT_0858
Phatr2_27976	0.47	Phosphoenolpyruvate carboxylase
Phatr2_46746	0.46	unknown function
Phatr2_48756	0.46	Pseudouridine synthase
Phatr2_43339	0.46	unknown function
Phatr2_48218	0.46	unknown function
Phatr2_34321	0.45	Prolyl 4-hydroxylase alpha subunit
Phatr2_34413	0.45	glucose-methanol-choline oxidoreductase
Phatr2_49088	0.44	unknown function
Phatr2_50492	0.44	unknown function
Phatr2_44864	0.44	unknown function
Phatr2_50451	0.44	unknown function
Phatr2_4937	0.44	AAA ATPase
Phatr2_44076	0.44	Putative tyrosine kinase
Phatr2_46543	0.44	Decapping enzyme Dcp2
Phatr2_43046	0.44	G2/Mitotic-specific cyclin A
Phatr2_16991	0.43	Xanthine/uracil transporters
Phatr2_44974	0.43	unknown function
Phatr2_55200	0.42	proton-transporting ATPase
Phatr2_11441	0.42	RNA polymerase II
Phatr2_48011	0.42	unknown function
Phatr2_49133	0.42	Serine/threonine protein kinase
Phatr2_16963	0.42	Glutaminyl-tRNA synthetase
Phatr2_18665	0.41	Serine hydroxymethyltransferase
Phatr2_50456	0.41	Polyadenylate-binding protein
Phatr2_45803	0.41	Nucleolar GTPase/ATPase p130
Phatr2_49308	0.40	unknown function
Phatr2_50153	0.40	hypothetical protein DDBDRAFT_0205657
Phatr2_54265	0.38	calcium/calmodulin-dependent protein kinase
Phatr2_30471	0.38	5,10-methylenetetrahydrofolate reductase
Phatr2_10896	0.38	Mitochondrial carrier protein PET8
Phatr2_55046	0.38	putative DNA binding protein
Phatr2_41478	0.36	hypothetical protein Syncc9902_0102
Phatr2_44310	0.36	Predicted mechanosensitive ion channel
Phatr2_44153	0.35	Esterase/lipase/thioesterase
Phatr2_45475	0.35	20S proteasome, A and B subunits
Phatr2_49343	0.35	unknown function
Phatr2_11230	0.34	unknown function
Phatr2_46703	0.34	unknown function
Phatr2_11740	0.33	synthase/orotate phosphoribosyltransferase
Phatr2_46793	0.32	leucine-rich-repeat protein
Phatr2_43307	0.32	unknown function

Phatr2_12239	0.31	putative adenylate kinase
Phatr2_48055	0.30	Nuclear protein SET
Phatr2_42398	0.29	Malate dehydrogenase
Phatr2_54901	0.28	RabGAP/TBC domain protein
Phatr2_49908	0.27	Molecular chaperone (DnaJ superfamily)
Phatr2_51916	0.27	DNA mismatch repair protein
Phatr2_39421	0.27	Cl- channel, voltage gated
Phatr2_40317	0.26	unknown function
Phatr2_43293	0.25	unknown function
Phatr2_47352	0.16	Protein of unknown function DUF6
Phatr2_43917	0.12	Inositol polyphosphate 5-phosphatase
Phatr2_41413	0.11	Putative NADP-dependent oxidoreductase

Table 3.4. Gene list obtained by comparing the dark time point between wild-type and the *PtCPF1* overexpressing line (c15). The table shows the 64 genes downregulated in c15. Target ID indicated in blue was analyzed by qRT-PCR.

Table 3.5

Target ID	Ratio in WT	Ratio in c15	Ratio c15/WT	EST Expression	Description
Dark-Light					
Phatr_48732	10.78	83.65	7.76	OS, UA, TA	Prolyl 4-hydroxylase
Phatr_51703	31.03	95.45	3.08	BL, LDD, NS, SP, FES, AA	Violaxanthin de-epoxidase
Dark-Light					
Phatr_18180	79.25	37.16	0.47	BL, LDD	Chlorophyll binding protein
Dark-Dark					
Phatr_51916			0.27	LDD, UA, NS, SM, SP, FES, TA, AA	DNA mismatch repair protein
Dark-Dark					
Phatr_49594			4.14	BL, LDD, HDD, SM, SP, UA, FES, TA, OM, OS, TM	Heat shock transcription factor
Phatr_49596			2.25	BL, LDD, NS, SM, SP, FES, TA, OM, OS, TM	Heat shock transcription factor

Table 3.5. Expression data of genes selected for qRT-PCR in microarray analyses. "EST Expression" column shows expression data obtained from EST database (http://www.biologie.ens.fr/diatomics/EST2). BL - Blue Light, LDD - Low Decadienal Treated, HDD - High Decadienal Treated, OS - Original 12000 Standard, UA - Urea Adapted (low urea), NS - Nitrate Starved, SM - Silica Minus, SP - Silica Plus, FES - Iron Starved, AA - Ammonium Adapted (low ammonium), OM - Oval Morphology, TA - Tropical Accession Pt9, TM – Triradiate Morphotype.
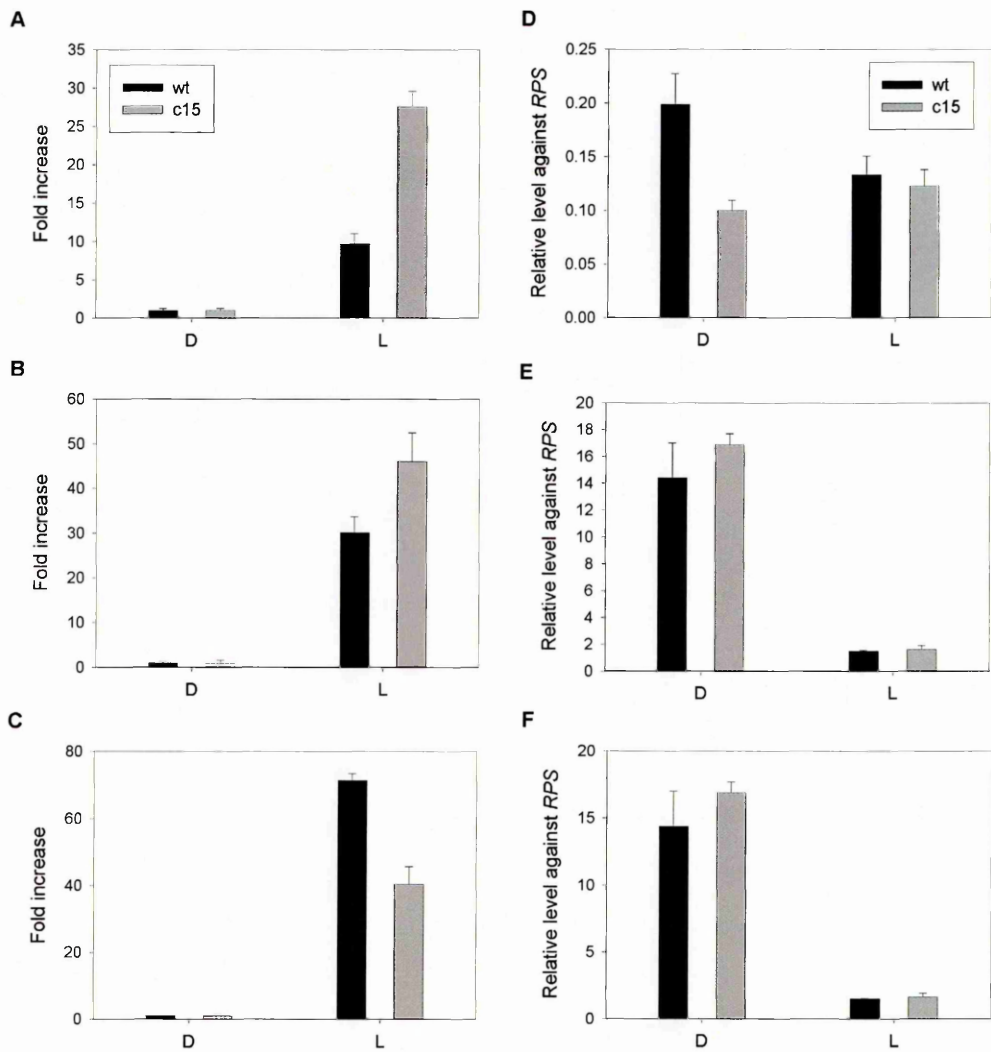
Fig. 34. The effect of a blue light pulse on *Phatr_48732* (A), *Phatr_51703* (B), *Phatr_18180* (C), *Phatr_51916* (D), *Phatr_49594* (E), *Phatr_49596* (F) transcript levels in wild-type and the *PtCPF1* overexpressing line using *RPS* as a reference. The values were normalized to gene expression levels in the dark in panels A, B, C. Data are averages of triplicate measurements of cDNA. The error bars indicate standard deviation.

CHAPTER IV - DISCUSSION

4.1 New Tools for Reverse Genetics in Diatoms

A first step in this thesis project involved the extraction of high molecular weight DNA to be used for sequencing the *P. tricornutum* genome (Fig. 10). The subsequent availability of the whole genome sequence has greatly facilitated the development of reverse genetics in diatoms, and a vast amount of new information has enriched the scientific community. For example, comparisons of pathway-related genes allowed the identification of particular features of diatom cell biology, such as the presence of a urea cycle (see sections 1.1.5). Moreover, comparison of the chloroplast genomes of *P. tricornutum* and *T. pseudonana* compared with other plastid genomes have confirmed that diatoms likely acquired their chloroplast from a red algal endosymbiont (Oudot-Le Secq M.P. *et al.*, 2006). However, these examples represent just the tip of the iceberg compared to what is likely to be discovered in the future. Indeed, at present genomics is a field in continuous expansion, and has now progressed from sequencing single genomes to "community genomics" or "environmental genomics", aimed at identifying all community members and genes in a specific environment, in order to understand an ecological system in all its complexity. Several metagenomic approaches have been applied to marine ecosystems, such as the Sargasso Sea (Venter J.C. *et al.*, 2004) and deep sea sediments (Hallam S.J. *et al.*, 2004). In the near future knowledge derived from diatom genomes will be linked to these metagenomic data from organisms living in the same habitat, offering the possibility to study their biology in its true complexity.

As a complement to the whole genome sequence, we considered it essential to develop molecular tools for the diatom community. The Gateway cloning strategy has thus been optimized for *P. tricornutum*, generating several Destination vectors useful for different purposes (Fig. 12). The Destination vector for overexpression is useful to study the function of a gene of interest. For example, overaccumulation of

a gene product might generate a dominant negative mutant phenotype. On the other hand, enhanced or deregulated responses might help to understand the function of the encoded protein, as has been the case reported here for *PtCPF1*. The availability of this vector is all the more important considering that knockout technology is still missing in diatoms. In addition, diatom Destination vectors containing several epitopes were generated. EYFP or ECFP epitopes are useful for subcellular localization, whereas Destination vectors carrying an HA tag allow immunopurification and other immunodetection-based methods. Moreover, the utility of the Gateway system is that different fusion proteins can easily be obtained with one step cloning. Analysis of a gene product can also be performed using fusions at both extremities, in the case that insufficient information on putative function is available. These new molecular tools will also improve collaboration and dialogue between scientists with different knowledge, such as oceanographers or ecologists that typically consider molecular biology as a too specialized field.

4.2 *PtCPF1* is a novel Cryptochrome/Photolyase Family member

As part of this thesis project, a novel protein of the cryptochrome/photolyase family has been identified that displays a dual function. *PtCPF1* appears to be a blue light photoreceptor, modulating several blue-light-dependent responses, and has a specific DNA repair activity.

4.2.1 Phylogeny of diatom CPFs

Phylogenetic analysis shows that PtCPF1 is more similar to animal cryptochromes than to plant cryptochromes (Fig. 14). The presence of animal-type cryptochromes in diatoms can be explained by the evolutionary history of this group of organisms (see section 1.1.4). Differently from green and red algae and terrestrial

plants that are derived from a primary endosymbiotic event in which a non-photosynthetic eukaryote acquired a chloroplast by engulfing a cyanobacterium, diatoms likely arose after a secondary endosymbiotic event, whereby a eukaryotic algae was engulfed by a second eukaryotic heterotroph (Armbrust E.V. *et al.*, 2004). This second endosymbiotic event likely generated a new and unusual combination of genes. In fact, analysis of diatom genomes indicates that almost 50% of diatom proteins are more similar to animal than to plant counterparts (Armbrust E.V. *et al.*, 2004). Remarkably, phylogenetic analysis of the diatom cry/photolyase members showed another interesting feature, namely that more than one CRY-DASH protein is present in the two diatoms, because only one CRY-DASH member has been identified in other organisms such as *Xenopus*, *Arabidopsis*, zebrafish (Daiyasu H. *et al.*, 2004). Such diversity of putative cryptochromes in marine phytoplankton could find an explanation considering the important role of blue light signals in the water. Blue light has probably played a particularly important role as a driving force in evolution because it is the only component of the sun's spectrum that penetrates to significant depths in aquatic environments (see section 1.2.2) such as those in which life began on earth (Gehring W. & Rosbash M., 2003; Ragni M. & Ribera d'Alcalà M., 2004). It is therefore reasonable to expect a complex variability of blue-light photoreceptor molecules in marine organisms strictly linked to changes in light intensity and quality.

4.2.2 Cryptochrome/Photolyase Chromophores

In this thesis project, a preliminary biochemical characterization of PtCPF1 protein has been performed. The full length protein purified from *E. coli* was found to be suitable for spectroscopic and enzymatic analyses. All cryptochrome/photolyase family members characterized to date contain FAD as

essential chromophore. In addition, these proteins contain a second chromophore, which in the majority of organisms is methenyltetrahydrofolate (MTHF). Spectral analyses of GST:PtCPF1 revealed the presence of a non-covalently bound FAD chromophore in the reduced state (Fig. 17, 18). The presence of the flavin cofactor in the reduced state is of interest because in the majority of cryptochromes isolated (e.g., *A. thaliana* (Lin C. *et al.*, 1995; Malhotra K. *et al.*, 1995), humans (Hsu D.S. *et al.*, 1996), and cyanobacteria (Hitomi K. *et al.*, 2000)), the flavin is in the two-electron oxidized form, whereas the active form of flavin in photolyase is the two-electron reduced form (Hitomi K. *et al.*, 1997; Payne G. *et al.*, 1987). To date, only a few cryptochromes have been purified with the flavin active state, such as *V. cholerae* CRY1 (Worthington E.N. *et al.*, 2003) or CRY3 from *A. thaliana*. In this latter case, spectroscopic studies showed that the dark-adapted state of AtCRY3 contained several FAD forms, such as 40% neutral oxidized FAD (FAD_{ox}), 55% reduced FAD ($FADH_2$) and 5% neutral FAD semiquinone ($FADH^{\bullet}$). Photo-excitation reduced the FAD_{ox} and $FADH^{\bullet}$ to the reduced $FADH_2$ that is the active form (Song S.H. *et al.*, 2006). Therefore, our finding that PtCPF1 contains the flavin in the two-electron reduced form provides further evidence that all other native cryptochromes also contain the flavin in this form, suggesting that cryptochrome functions in a manner similar to photolyase, by photoinduced electron transfer to a substrate. On the other hand, the recombinant diatom protein did not display a clear absorption peak at 375-410 nm that is typical of the MTHF cofactor, indicating that the two cofactors are not present in stoichiometric amounts. However, MTHF is not essential for photolyase function, as was also demonstrated by the *in vitro* enzymatic assay (Fig. 21). Significant amounts of the second cofactor were probably lost during the purification procedure, although we cannot rule out the possibility that PtCPF1 contains a deazaflavin cofactor, which is not synthesized in *E. coli*, as has

been observed for one member of the cryptochrome/photolyase family in cyanobacteria. In fact, *Synechocystis* sp. PCC6803 contains two members of the cryptochrome/photolyase family, *PhrA* and *PhrB* genes. Their products expressed and purified from *E. coli* did not contain the second cofactor (Hitomi H. *et al.*, 2000) but further studies revealed that *PhrA* encodes a cyclobutane photolyase that contains deazaflavin as a second chromophore when purified from its native source (Ng W.O. *et al.*, 2000), whereas the crystal structure of the *PhrB* gene product, that is also known as a CRY-DASH protein, contained FAD but no second chromophore (Brudler R. *et al.*, 2003). The protein used for crystallography was made in *E. coli*, so it is possible that the native protein contains deazaflavin as a second chromophore or that the second chromophore may be distinct from folate and deazaflavin or even absent altogether. The same problems may also apply to PtCPF1 and so it will be important to purify the native protein from diatom cells.

4.2.3 RNA is associated with PtCPF1

A peculiar aspect that was highlighted thanks to the characterization of purified GST:PtCPF1 was the presence of nucleic acids associated with the protein, which were subsequently identified as RNA (Fig. 19). Worthington E.N. *et al.* (2003) observed the same phenomenon when they purified the VcCRY1, a CRY-DASH protein, using a different epitope tag. In addition, these authors have commented that of the 13 recombinant cryptochrome/photolyase family members purified in their laboratory, only VcCRY1 was found to be associated with RNA. We can now conclude that RNA-association is not a CRY-DASH specific feature because it has also been observed for PtCPF1, which is an animal-type cryprochrome. Moreover, a colleague working in the laboratory on two *Ostreococcus* CPF members, an animal-type and a CRY-DASH, has also observed

the presence of nucleic acids associated with both proteins (personal communication). Further characterization of the RNA is clearly needed but a major effort must be to purify PtCPF1 from its native host to determine whether the protein still associates with RNA. At present, diatom overexpressing lines containing STREP:PtCPF1 are available in our laboratory. Strep (Streptavidin) epitope tag is commonly used for protein purification, and preliminary experiments on these clones showed an encouraging binding of the STREP:PtCPF1 protein to the resin. In the near future, a more detailed characterization of the native PtCPF1 protein should therefore be possible.

4.2.4 Photolyase activity *in vitro* and *in vivo*

Recombinant PtCPF1 showed a specific binding to (6-4)photoproducts as well as specific (6-4)photorepair activity in *in vitro* experiments (Fig. 20, 21, respectively), demonstrating that the purified protein not only contains the FAD cofactor attached correctly, but is also able to activate cyclic electron transfer between the catalytic flavin adenine dinucleotide cofactor and the damaged DNA. The absence of a detectable portion of the antenna cofactor in the purified protein did not impair the function of photorepair activity in the *in vitro* assay, as previously discussed. PtCPF1 also displayed (6-4)photolyase activity in an *E. coli in vivo* repair assay (Fig. 22). After UV treatment, a comparable survival rate was observed in bacteria expressing either the PtCPF1 protein or the (6-4)photolyase from *Danio rerio*, which was used as a positive control.

4.2.5 Regulation of *PtCPF1* by light

A thorough analysis of *PtCPF1* transcriptional regulation has been performed, leading to intriguing conclusions. Semi-quantitative RT-PCR

experiments aimed at studying gene expression during a diurnal cycle revealed the presence of a population of transcripts. Each individual mRNA appears to oscillate during the diel cycle (Fig. 25). Subsequent, sequence analysis showed that different intermediate mRNAs are expressed at different times. The most abundant intermediate transcript is a full length mRNA containing the last intron (number 4). Remarkably, the intron sequence contains an in-frame stop codon, which generates a transcript of 1,368 bp from start to stop codon encoding a short protein of 455 amino acids with a molecular weight of 52 kDa. At present, we do not have any proof for the existence of a short protein, although domain analysis illustrated the characteristics of different domains and indicated that the short protein, lacking its C-terminal domain, should it exist, could play a regulatory role (section 3.2.3).

Many examples in the literature have shown the relevant role of alternative splicing for proteomic complexity (Kim E. *et al.*, 2007). Remarkably, in humans this process affects 60% of genes, and thus should be considered more the rule than the exception (Kornblihtt A.R., 2005). To mention one recent example, the different biological function of two human cyclin proteins has been demonstrated, generated by alternative splicing, denoted cyclin D1a and cyclin D1b. The long form (D1a) has a molecular weight of 36 kDa, while the short form (30-31 kDa) lacks the C-terminal moiety required for protein stability and sub-cellular localization (Lévêque C. *et al.*, 2007). The number of alternatively spliced genes reported in plants is currently much smaller than in mammals. Interestingly, a relational database named the Plant Alternative Splicing Database (PASD) has been developed in order to collect and analyze genes subjected to this mechanism (Zhou Y. *et al.*, 2003). Further studies of the transcriptional regulation of *PtCPF1* will be required to clarify the biological role of the intermediate mRNAs. Moreover, qRT-PCR analysis revealed a diurnal oscillation, suggesting that alternative splicing may be necessary to modulate the

pool of *PtCPF1* during the normal light-dark cycle, leading to an increase during the day and a decrease in the night (Fig. 26). Concerning the circadian regulation of *PtCPF1*, we might conclude that a light-independent control of transcript levels is present, as demonstrated by its pattern observed in free-running dark conditions. Nevertheless, this control seems to be absent in free-running light conditions, suggesting that a pleiotropic regulation occurs in the light (Fig. 26). On the other hand, the PtCPF1 protein profile showed a faint oscillation between the light and the dark period, that was not strongly conserved under constant conditions (Fig. 27). However, a light-dependent post-transcriptional regulation may be involved in the shift from an inactive to an active protein. Finally, studies of the acute light response showed a specific induction of *PtCPF1* upon blue light, whereas protein levels were induced both by blue light and white light (Fig. 28). The different behaviour of transcript and protein is a further indication that post-transcriptional regulatory mechanisms modulate *PtCPF1* function.

4.2.6 *PtCPF1* as a potential regulator of the negative feedback loop of the circadian clock

Circadian clocks are endogenous time-keeping devices that regulate daily changes in many aspects of physiology and behaviour. Organisms ranging from bacteria to humans have a clock and many general properties of their function are conserved. Surprisingly, diatom PtCPF1 is able to inhibit CLOCK:BMAL1-mediated transcription in animal cells, suggested that it is able to function in the negative feedback loop of the mammalian circadian clock (Fig. 29, 30). At present, we do not have molecular evidence to explain how the diatom protein mediates this function, but we can speculate on some models.

Cryptochrome proteins have two domains, a core region similar to the photolyases, and a carboxy-terminal tail that varies considerably in length and sequence composition (Green C.B., 2004). Plant and animal cryptochromes possess 30-250 amino acid carboxy-terminal extensions beyond the photolyase-homology region (PHR), which is approximately 500 amino acids. Importantly, these C-terminal domains have been shown to mediate phototransduction by both *Arabidopsis* and *Drosophila* cryptochromes.

In *Arabidopisis*, the carboxy-terminal tail is responsible for transducing the light signal detected by the core PHR domain, by direct interaction and inhibition of its effector protein COP1 upon light activation. The light-dependent inhibition of the COP1 E3 ubiquitin ligase allows accumulation of a set of transcription factors that initiate the photomorphogenic response. Overexpression of the C-terminal domains of either AtCRY1 or AtCRY2 results in a constitutive photomorphogenic phenotype (Yang H.Q. *et al.*, 2000). Therefore, in *Arabidopsis* the core domain acts as a regulator of the carboxy-terminal domain, inhibiting its activity in the dark and/or promoting its activity in the light.

In *Drosophila*, dCRY is a circadian photoreceptor that, when activated by light, interacts with the central clock proteins Timeless (TIM) and Period (PER) (Ceriani M.F. *et al.*, 1999; Rosato E. *et al.*, 2001) and leading to their degradation, which is thought to reset the clock. The C-terminal domain of dCRY is essential for maintaining these light dependent interactions. In fact, a mutant form of *Drosophila* CRY lacking the carboxy-terminal tail interacts in a constitutive, light-independent manner with PER in yeast (Rosato E. *et al.*, 2001). Moreover, overexpression of the truncated CRY in *Drosophila* results in several effects that mimic constant light exposure, including behavioural rhythms with periods longer than 24 hours, altered TIM and PER kinetics and intracellular localization, and changes in dCRY stability

(Busza A. *et al.*, 2004). These results indicate that the core (PHR) domain of *Drosophila* CRY is capable of carrying out its phototransduction functions and that the carboxy-terminal tail must play a regulatory role, normally preventing dCRY from being active in the dark.

In vertebrates, the CRY carboxy-terminal tail has yet a different role, and there is also a high variability between different species, leading to the conclusion that the C-terminal tail is the most creative domain of the molecule. In fact, in *Xenopus* CRYs, the role of the carboxy-terminal tail is to transport the CRY protein into the nucleus, which is required for its transcriptional repressor function. CRY molecules lacking the carboxy-terminal tail are localized in the cytoplasm and do not repress transcription (Zhu H. *et al.*, 2003). Addition of a heterologous nuclear localization signal to the truncated CRY completely restores both repressor activity and nuclear localization, indicating that the core photolyase-like domain is sufficient for repressive activity as long as the protein can make it to the nucleus. It is not known whether this nuclear targeting function of the carboxy terminus is regulated, but it does not appear to be light-dependent. On the contrary, domain analysis of cryptochrome 1a from zebrafish demonstrated that a nuclear localization signal (NLS) is present in the core domain of the protein in a region denoted RD2b. The NLS sequence, that is well conserved among repressor-type CRYs, has also been demonstrated to be functional in mCRY1 (Hirayama J. *et al.*, 2003). Thus, in both *Xenopus* and zebrafish, nuclear localization of CRYs is essential for the interaction with the CLOCK:BMAL heterodimer, but only in *Xenopus* is it mediated by the C-terminal tail, suggesting that that domain plays a different role in zebrafish and as a consequence probably also in mouse.

PtCPF1 constitutes an interesting model to define the function of this domain because it can be considered the missing link in the evolutionary history of the

cryptochrome/photolyase family. An accurate domain analysis (see section 3.2.3) showed that the core domain of PtCPF1 is similar to the animal CRYs, while the protein contains a short C-terminal extension similar to the *Drosophila* cryptochrome. We can speculate that PtCPF1 is derived from a typical (6-4)photolyase that under evolutionary pressure acquired a repressor function necessary for the regulation of the circadian clock. The protein core containing the PHR domain is necessary to repair the (6-4)photoproducts, and could also be sufficient for inhibition of the negative feedback loop, as has been shown for vertebrate cryptochromes. Moreover, since the protein appears to be constitutively nuclear localized, the C-terminal domain might play a regulatory role in switching the two functions, dependent on the input perceived from the environment.

Ishikawa T. *et al.* (2002) showed that zebrafish CRY1a neither disrupts the association between zfCLOCK and zfBMAL nor inhibits binding of the zfCLOCK-zfBMAL1 heterodimer to an E-box-bearing DNA fragment. Instead it binds to the heterodimer to form a stable zCRY1a-zfCLOCK-zfBMAL1-E-box complex, in order to inhibit transcriptional activity of the heterodimer. Results of the transcription assays performed in COS7 and zebrafish cells led to the conclusion that PtCPF1 is able to bind CLOCK and BMAL1 proteins, perhaps forming a stable complex as efficiently as zCRY1a. Although a preliminary analysis to look for CLOCK and BMAL orthologues in the *P. tricornutum* genome did not reveal similar transcription factors, it is possible that molecules with a similar secondary or tridimensional structure are conserved in diatoms, explaining how PtCPF1 can recognize the heterodimer conformation and play a similar role to that of zCRY1a. However, basic helix-loop-helix PAS superfamily domain-containing proteins are encoded in the *P. tricornutum* genome, so in the future it could be possible to identify positive regulators in diatoms that are not orthologous but analogous to

CLOCK and BMAL. Another interesting consideration is that PtCPF1 acts as a general repressor in the negative feedback loop of the mammalian circadian clock. In fact, the protein inhibits CLOCK:BMAL-mediated transcription in two different cellular systems, suggesting also that the action mechanism is light independent. Further studies should clarify this aspect.

Finally, it is important to bear in mind that expression studies of the *PtCPF1* transcript have revealed the presence of intermediate mRNA transcripts during the diel cycle. In particular, an intermediate mRNA containing the last intron was always observed (Fig. 25). This intermediate transcript encodes a short PtCPF1 protein, lacking the C-terminal domain (see red asterisk in Fig. 15). It will be extremely interesting to understand whether this intermediate is just a splicing intermediate, or, on the contrary, whether it encodes a short functional PtCPF1 protein. At present, we have no evidence for the existence of a second form of PtCPF1, because only a single band of 63 kDa, corresponding to the full length protein, is detected by western blotting using the specific CPF1 antibody on diatom cells grown under different light conditions. Moreover, transgenic lines expressing the truncated version of PtCPF1 fused to EYFP at the C-terminal extremity did not reveal a fluorescent signal (data not shown), although negative results were also obtained when using the full length PtCPF1 protein with EYFP at the carboxyl-extremity, suggesting that the EYFP epitope at the C-terminal position might interfere with the correct folding of the protein, leading to degradation. Further characterization of the putatively functional short protein will also elucidate the role of separate domains in the diatom photoreceptor. In fact, the short PtCPF1 form lacks the predicted coiled-coil domain in the C-terminal region (aa 502-524). At present, we propose a model in which the core domain of PtCPF1 contains the nuclear localization signal and might play both functions, the repair activity and the

transcriptional repression. The C-terminal tail of PtCPF1 might play a regulatory role in switching the two functions, or it could be essential for protein stability. The latter consideration could also explain why we did not observe the short PtCPF1 that, lacking the C-terminal domain, is unstable. In conclusion, it seems extremely important to identify the role of separate domains in PtCPF1 to better understand the key changes that have occurred in the protein during evolution.

4.2.7 Transcriptional regulation of genes by *PtCPF1*

Photoreceptors modulate output pathways in a fluence rate dependent manner that depends on the quantity of light perceived. Gene expression studies of acute light responses in which wild-type cells were compared with *PtCPF1* overexpressing lines confirmed this correlation and suggested a possible role of *PtCPF1* in the blue light photoreceptor response (Fig. 32).

Psy and *Pds* gene expression were fluence-modulated, showing the strongest induction at 3.3 $\mu mol.m^{-2}.s^{-1}$ and a partial inhibition at a higher intensity, indicating a threshold-response. Overexpressing lines showed a significant repression of *Psy* and *Pds* gene expression, in agreement with the role of PtCPF1 as a negative regulator demonstrated in the transcriptional assays in animal cells (Fig. 29, 30). On the contrary, the expression of *FcpB* was almost unaffected by fluence rate and no differences were observed between wild-type and overexpressing lines. Previously, studies on *FcpB* gene expression showed a strong induction after a 3 hours treatment with continuous blue light of 25 $\mu mol.m^{-2}.s^{-1}$ (maximum increase >5,000 fold). Data from acute light responses might therefore suggest that *FcpB* gene expression is more likely to be controlled by the photosynthetic redox state of the cells than by blue light photoreceptors. In addition, the fact that there are no differences in the *FcpB* gene expression pattern between wild-type and overexpressing lines might

represent an indirect control that suggests a role for *PtCPF1* principally in blue-light photoreceptor responses.

Finally, the strongest acute responses were observed for *PtCPF25* and *CYC* gene expression. In both cases, a sensitive response to fluence rate was observed although the two genes behaved oppositely to each other. *PtCPF25* showed a typical positive fluence response, whereas *CYC* showed a typical negative fluence response, demonstrating that the control of their expression is under blue light control. Overexpressing lines showed altered responses, indicating that *PtCPF1* is involved in upstream controls of both genes. Moreover, cryptochromes in animals and in plants are key-components of the biological clock, so it will be interesting to investigate the influence of CPF1 overexpression on the cell cycle in *P. tricornutum*.

CONCLUDING REMARKS

Studies of light perception and signaling is a fascinating field because light is a key factor for all living organisms. Diatoms are an intriguing model organism for light-perception studies because of the complexity of light signals underwater. Spectral quality varies with depth due to the absorption properties of the water, with blue light prevailing with increasing depth. It is possible to hypothesize that successful marine algae such as diatoms must have developed sophisticated strategies for responding to variations in light quality and quantity.

In this thesis project a structural and functional characterization of a Cryptochrome/Photolyase Family member (*CPF1*), isolated from the pennate diatom *P. tricornutum*, has been performed. *PtCPF1* is a highly novel and interesting member of this family because it displays a dual activity: a (6-4)photolyase activity, likely of functional relevance for cell survival following UV irradiation, and a photoreceptor activity, controlling gene expression, pigment synthesis and possibly circadian regulated processes in response to blue light. This has been an exciting discovery because no protein with dual function had previously been discovered, and so the sea has provided us with the missing link in the evolution of the Cryptochrome/Photolyase protein family. In addition, the dual activity of the diatom protein opens the way to molecular evolution studies to see how catalytic activity evolved from a photolyase to a photoreceptor. Further characterization of *CPF1* and other members of the Cryptochrome/Photolyase Family identified in *P. tricornutum* and other diatoms will likely provide novel insights for dissecting the molecular secrets underlying the success of diatoms in contemporary oceans.

CHAPTER V - BIBLIOGRAPHY

1.   Ahmad M. & Cashmore A.R. (1993). HY4 gene of *A. thaliana* encodes a protein with characteristics of a blue-light photoreceptor. *Nature*, **366**: 162-166.

2.   Allen A.E., Vardi A. & Bowler C. (2006). An ecological and evolutionary context for integrated nitrogen metabolism and related signaling pathways in marine diatoms. *Curr Opin Plant Biol*, **9**: 264-73.

3.   Armbrust E.V., Berges J.B., Bowler C., Green B.R., Martinez D., Putnam N.H., Zhou S., Allen A.E., Apt K.E., Bechner M., Brzezinski M.A., Chaal B.K., Chiovitti A., Davis A.K., Demarest M.S., Detter J.C., Glavina T., Goodstein D., Hadi M.Z., Hellsten U., Hildebrand M., Jenkins B.D., Jurka J., Kapitonov V.V., Kröger N., Lau V.V.Y., Lane T.W., Larimer F.W., Lippmeier J.C., Lucas S., Medina M., Montsant A., Obornik M., Schnitzler-Parker M., Palenik B., Pazour G.J., Richardson P.M., Rynearson T.A., Saito M.A., Schwartz D.C., Thamatrakoln K., Valentin K., Vardi A., Wilkerson F.P. & Rokhsar D.S. (2004). The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science*, **306**: 79-86.

4.   Aubert C., Vos M.H., Mathis P., Eker A.P. & Brettel K. (2000). Intraprotein radical transfer during photoactivation of DNA photolyase. *Nature*, **405**: 586-90.

5.   Batschauer A. (1993). A plant gene for photolyase: an enzyme catalyzing the repair of UV-light-induced DNA damage. *Plant J*, **4**: 705-9.

6. Baum G., Long J.C., Jenkins G.I. & Trewavas A.J. (1999). Stimulation of the blue light phototropic receptor NPH1 causes a transient increase in cytosolic Ca2+. *Proc Natl Acad Sci USA*, **96**: 13554-9.

7. Beja O., Aravind L., Koonin E.V., Suzuki M.T., Hadd A., Nguyen L.P., Jovanovich S.B., Gates C.M., Feldman R.A., Spudich J.L., Spudich E.N. & DeLong E.F. (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science*, **289**: 1902-6.

8. Bellingham J. & Foster R.G. (2002). Opsins and mammalian photoentrainment. *Cell Tissue Res*, **309**: 57-71.

9. Berger S.A., Diehl S., Stibor H., Trommer G., Ruhenstroth M., Wild A., Weigert A., Jager C.G. & Striebel M. (2007). Water temperature and mixing depth affect timing and magnitude of events during spring succession of the plankton. *Oecologia*, **150**: 643-54.

10. Bernard P. & Couturier M. (1992). Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J Mol Biol*, **226**: 735-45.

11. Berson D.M., Dunn F.A. & Takao M. (2002). Phototransduction by retinal ganglion cells that set the circadian clock. *Science*, **295**: 1070-3.

12. Bhattacharya D. & Medlin L. (1995). Targeting proteins to diatom plastids involves transport through an endoplasmic reticulum. *J Phycol*, **31**: 489-98.

13.   Bhaya D. & Grossman A.R. (1993). Characterization of gene clusters encoding the fucoxanthin chlorophyll proteins of the diatom *Phaeodactylum tricornutum*. *Nucleic Acids Res*, **21**: 4458-66.

14.   Bieszke J.A., Spudich E.N., Scott K.L., Borkovich K.A. & Spudich J.L. (1999). A eukaryotic protein, NOP-1, binds retinal to form an archaeal rhodopsin-like photochemically reactive pigment. *Biochemistry*, **38**: 14138-45.

15.   Bouly J.P., Giovani B., Djamei A., Meuller M., Zeugner A., Dudkin E.A., Baschauer A. & Ahmad M. (2003). Novel ATP-binding and autophosphorylation activity associated with *Arabidopsis* and human cryptochrome-I. *Eur J Biochem*, **270**: 2921-2928.

16.   Brautigam C.A., Smith B.S., Ma Z., Palnitkar M., Tomchick D.R., Machius M. & Deisenhofer J. (2004). Structure of the photolyase-like domain of cryptochrome 1 from *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* **101**: 12142-7.

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

18.   Briggs W.R., Tseng T.-S., Cho H.-Y., Swartz T.E., Sullivan S., Bogomolni R.A., Kaiserli E. & Christie J.M. (2007). Phototropins and Their LOV Domains: Versatile Plant Blue-Light Receptors. *J Integr Plant Biol*, **49**: 1-4

19.   Brudler R., Hitomi K., Daiyasu H., Toh H., Kucho K., Ishiura M., Kanehisa M., Roberts V.A., Todo T., Tainer J.A. & Getzoff E.D. (2003).

Identification of a new cryptochrome class. Structure, function, and evolution. *Mol Cell*, **11**: 59-67.

20. Busza A., Emery-Le M., Rosbash M. & Emery P. (2004). Roles of the two *Drosophila* CRYPTOCHROME structural domains in circadian photoreception. *Science*, **304**: 1503-6.

21. Cashmore A.R. (2003). Cryptochromes: enabling plants and animals to determine circadian time. *Cell*, **114**: 537-43.

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

23. Cavalier-Smith T. (2000). Membrane heredity and early chloroplast evolution. *Trends Plant Sci*, **5**: 174-82.

24. Ceriani M.F., Darlington T.K., Staknis D., Mas P., Petti A.A., Weitz C.J. & Kay S.A. (1999). Light-dependent sequestration of timeless by cryptochrome. *Science*, **285**: 553-6.

25. Chaves I., Yagita K., Barnhoorn S., Okamura H., van der Horst G.T. & Tamanini F. (2006). Functional evolution of the photolyase/cryptochrome protein family: importance of the C terminus of mammalian CRY1 for circadian core oscillator performance. *Mol Cell Biol*, **26**: 1743-53.

26. Chen M., Chory J. & Fankhauser C. (2004). Light signal transduction in higher plants. *Annu Rev Genet*, **38**: 87-117.

27. Chew A.G. & Bryant D.A. (2007). Bacteriochlorophyll and Carotenoid Biosynthesis. *Annu Rev Microbiol*, in press.

28. Christie J.M., Swartz T.E., Bogomolni R.A. & Briggs W.R. (2002). Phototropin LOV domains exhibit distinct roles in regulating photoreceptor function. *Plant J*, **32**: 205-19.

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

30. Cutler S.R., Ehrhardt D.W., Griffitts J.S. & Somerville C.R. (2000). Random GFP::cDNA fusions enable visualization of subcellular structures in cells of *Arabidopsis* at a high frequency. *Proc Natl Acad Sci USA*, **97**: 3718-23.

31. Daiyasu H., Ishikawa T., Kuma K., Iwai S., Todo T. & Toh H. (2004). Identification of cryptochrome DASH from vertebrates. *Genes Cells*, **9**: 479-95.

32. De Martino A., Meichenin A., Shi J., Pan K. & Bowler C. (2007). Genetic and Phenotypic Characterisation of *Phaeodactylum tricornutum* (Bacillariophyceae) Accessions. *J Phycol*, in press.

33. Derelle E., Ferraz C., Rombauts S., Rouze P., Worden A.Z., Robbens S., Partensky F., Degroeve S., Echeynie S., Cooke R., Saeys Y., Wuyts J., Jabbari K., Bowler C., Panaud O., Piegu B., Ball S.G., Ral J.P., Bouget F.Y., Piganeau G., De Baets B., Picard A., Delseny M., Demaille J., Van de Peer Y. & Moreau H. (2006). Genome analysis of the smallest free-living

eukaryote *Ostreococcus tauri* unveils many unique features. *Proc Natl Acad Sci USA*, **103**: 11647-52.

34.   Devlin P.F. (2002). Signs of the time: environmental input to the circadian clock. *J Exp Bot*, **53**: 1535-50.

35.   Douglas S., Zauner S., Fraunholz M., Beaton M., Penny S., Deng L.T., Wu X., Reith M., Cavalier-Smith T. & Maier U.G. (2001). The highly reduced genome of an enslaved algal nucleus. *Nature*, **410**: 1091-6.

36.   Duek P.D. & Fankhauser C. (2003). HFR1, a putative bHLH transcription factor, mediates both phytochrome A and cryptochrome signalling. *Plant J*, **34**: 827-36.

37.   Dunlap J.C. (1999). Molecular bases for circadian clocks. *Cell*, **96**: 271-90.

38.   Eide E.J., Vielhaber E.L., Hinz W.A. & Virshup D.M. (2002). The circadian regulatory proteins BMAL1 and cryptochromes are substrates of casein kinase Iε. *J Biol Chem*, **277**: 17248-54.

39.   Emery P., So W.V., Kaneko M., Hall J.C. & Rosbash M. (1998). CRY, a *Drosophila* clock and light-regulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity. *Cell*, **95**: 669-79.

40.   Emery P., Stanewsky R., Helfrich-Forster C., Emery-Le M., Hall J.C. & Rosbash M. (2000). *Drosophila* CRY is a deep brain circadian photoreceptor. *Neuron*, **26**: 493-504.

41.   Ermilova E.V., Zalutskaya Z.M., Huang K. & Beck C.F. (2004). Phototropin plays a crucial role in controlling changes in chemotaxis

during the initial phase of the sexual life cycle in *Chlamydomonas*. *Planta*, **219**: 420-7.

42.   Falciatore A. & Bowler C. (2002). Revealing the molecular secrets of marine diatoms. *Annu Rev Plant Biol*, **53**: 109-30.

43.   Falciatore A. & Bowler C. (2005). The evolution and function of blue and red light photoreceptors. *Curr Top Dev Biol*, **68**: 317-50.

44.   Falciatore A., Casotti R., Leblanc C., Abrescia C. & Bowler C. (1999). Transformation of nonselectable reporter genes in marine diatoms. *Marine Biotechnology*, **1**: 239-251.

45.   Falkowski P.G., Barber R.T. & Smetacek V.V. (1998). Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science*, **281**: 200-7.

46.   Felsenstein J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**: 783-791.

47.   Fujihashi M., Numoto N., Kobayashi Y., Mizushima A., Tsujimura M., Nakamura A., Kawarabayasi Y. & Miki K. (2007). Crystal structure of archaeal photolyase from *Sulfolobus tokadaii* with two FAD molecules: implication of a novel light-haversting cofactor. *J Mol Biol*, **365**: 903-910.

48.   Galston A.W. (1950). Riboflavin, light, and the growth of plants. *Science*, **111**: 619-24.

49.   Gehring W. & Rosbash M. (2003). The coevolution of blue-light photoreception and circadian rhythms. *J Mol Evol*, **57**: S286-9.

50.   Genick U.K., Soltis S.M., Kuhn P., Canestrelli I.L. & Getzoff E.D. (1998). Structure at 0.85 Å resolution of an early protein photocycle intermediate. *Nature*, **392**: 206-9.

51.   Gibbs S.P. (1981). The chloroplasts of some algal groups may have evolved from endosymbiotic eukaryotic algae. *Ann N Y Acad Sci*, **361**: 193-208.

52.   Giliberto L., Perrotta G., Pallara P., Weller J.L., Fraser P.D., Bramley P.M., Fiore A., Tavazza M. & Giuliano G. (2005). Manipulation of the blue light photoreceptor cryptochrome 2 in tomato affects vegetative development, flowering time, and fruit antioxidant content. *Plant Physiol*, **137**: 199-208.

53.   Gomelsky M. & Klug G. (2002). BLUF: a novel FAD-binding domain involved in sensory transduction in microorganisms. *Trends Biochem Sci*, **27**: 497-500.

54.   Gordon R. & Parkinson J. (2005). Potential roles for diatomists in nanotechnology. *J Nanosci Nanotechnol*, **5**: 35-40.

55.   Goud B., Salminen A., Walworth N.C. & Novick P.J. (1988). A Gtp-Binding Protein Required for Secretion Rapidly Associates with Secretory Vesicles and the Plasma-Membrane in Yeast. *Cell*, **53**: 753-768.

56.   Green C.B. (2004). Cryptochromes: tail-ored for distinct functions. *Curr Biol*, **14**: R847-9.

57.   Griffin E.A. Jr, Staknis D. & Weitz C.J. (1999). Light-independent role of CRY1 and CRY2 in the mammalian circadian clock. *Science*, **286**: 768-71.

58. Grossman A.R. (2003). A molecular understanding of complementary chromatic adaptation. *Photosynth Res*, **76**: 207-15.

59. Gu Y.Z., Hogenesch J.B. & Bradfield C.A. (2000). The PAS superfamily: sensors of environmental and developmental signals. *Annu Rev Pharmacol Toxicol*, **40**: 519-61.

60. Guillard R.R.L. (1975). Culture of phytoplankton for feeding marine invertebrates. In: Smith, W.L. and Chanley, M.H. (Eds), Culture of Marine Invertebrates Animals. Plenum Press, New York, USA, 26-60.

61. Guo H., Duong H., Ma N. & Lin C. (1999). The Arabidopsis blue light receptor cryptochrome 2 is a nuclear protein regulated by a blue light-dependent post-transcriptional mechanism. *Plant J*, **19**: 279-87.

62. Guo H., Mockler T., Duong H. & Lin C. (2001). SUB1, an *Arabidopsis* Ca^{2+}-binding protein involved in cryptochrome and phytochrome coaction. *Science*, **291**: 487-90.

63. Hallam S.J., Putnam N., Preston C.M., Detter J.C., Rokhsar D., Richardson P.M. & DeLong EF. (2004). Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science*, **305**: 1457-62.

64. Hamm-Alvarez S.F., Sancar A., Rajagopalan K.V. (1990). The presence and distribution of reduced folates in *Escherichia coli* dihydrofolate reductase mutants. *J Biol Chem*, **265**: 9850-6.

65. Heim R. & Tsien R.Y. (1996). Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr Biol*, **6**: 178-82.

66.   Hirayama J., Nakamura H., Ishikawa T., Kobayashi Y. & Todo T. (2003). Functional and structural analyses of cryptochrome. Vertebrate CRY regions responsible for interaction with the CLOCK:BMAL1 heterodimer and its nuclear localization. *J Biol Chem*, **278**: 35620-8.

67.   Hitomi K., Kim S.T., Iwai S., Harima N., Otoshi E., Ikenaga M. & Todo T. (1997). Binding and catalytic properties of Xenopus (6-4) photolyase. *J Biol Chem*, **272**: 32591-8.

68.   Hitomi K., Okamoto K., Daiyasu H., Miyashita H., Iwai S., Toh H., Ishiura M. & Todo T. (2000). Bacterial cryptochrome and photolyase: characterization of two photolyase-like genes of *Synechocystis* sp. PCC6803. *Nucl Acids Res*, **28**: 2353-2362.

69.   Hoffman E.C., Reyes H., Chu F.F., Sander F., Conley L.H., Brooks B.A. & Hankinson O. (1991). Cloning of a factor required for activity of the Ah (dioxin) receptor. *Science*, **252**: 954-8.

70.   Hsu D.S., Zhao X., Zhao S., Kazantsev A., Wang R.P., Todo T., Wei Y.F. & Sancar A. (1996). Putative human blue-light photoreceptors hCRY1 and hCRY2 are flavoproteins. *Biochemistry*, **35**: 13871-7.

71.   Huala E., Oeller P.W., Liscum E., Han I.S., Larsen E. & Briggs W.R. (1997). *Arabidopsis* NPH1: a protein kinase with a putative redox-sensing domain. *Science*, **278**: 2120-3.

72.   Huang K. & Beck C.F. (2003). Phototropin is the blue-light receptor that controls multiple steps in the sexual life cycle of the green alga *Chlamydomonas reinhardtii*. *Proc Natl Acad Sci USA*, **100**: 6269-74.

73.   Huang Y., Baxter R., Smith B.S., Partch C.L., Colbert C.L. & Deisenhofer J. (2006). Crystal structure of cryptochrome 3 from *Arabidopsis thaliana* and its implications for photolyase activity. *Proc Natl Acad Sci USA*, **103**: 17701-6.

74.   Imaizumi T., Tran H.G., Swartz T.E., Briggs W.R. & Kay S.A. (2003). FKF1 is essential for photoperiodic-specific light signaling in *Arabidopsis*. *Nature*, **426**: 302–6.

75.   Inada S., Ohgishi M., Mayama T., Okada K. & Sakai T. (2004). RPT2 is a signal transducer involved in phototropic response and stomatal opening by association with phototropin 1 in *Arabidopsis thaliana*. *Plant Cell*, **16**: 887-96.

76.   Iseki M., Matsunaga S., Murakami A., Ohno K., Shiga K., Yoshida K., Sugai M., Takahashi T., Hori T. & Watanabe M. (2002). A blue-light-activated adenylyl cyclase mediates photoavoidance in *Euglena gracilis*. *Nature*, **415**: 1047-51.

77.   Ishikawa T., Hirayama J., Kobayashi Y. & Todo T. (2002). Zebrafish CRY represses transcription mediated by CLOCK-BMAL heterodimer without inhibiting its binding to DNA. *Genes Cells*, **7**: 1073-86.

78.   Jiao Y., Lau O.S. & Deng X.W. (2007). Light-regulated transcriptional networks in higher plants. *Nat Rev Genet*, **8**: 217-30.

79.   Jones D.T., Taylor W.R. & Thornton J.M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci*, **8**: 275-82.

80. Jorns M.S., Wang B.Y., Jordan S.P. & Chanderkar L.P. (1990). Chromophore function and interaction in *Escherichia coli* DNA photolyase: reconstitution of the apoenzyme with pterin and/or flavin derivatives. *Biochemistry*, **29**: 552-61.

81. Jung K.H., Trivedi V.D. & Spudich J.L. (2003). Demonstration of a sensory rhodopsin in eubacteria. *Mol Microbiol*, **47**: 1513-22.

82. Kanai S., Kikuno R., Toh H., Ryo H. & Todo T. (1997). Molecular evolution of the photolyase-blue-light photoreceptor family. *J Mol Evol*, **45**: 535-48.

83. Kenrick P. & Crane P.R. (1997). The origin and early evolution of plants on land. *Nature*, **389**: 33-39.

84. Kim E., Magen A. & Ast G. (2007). Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res*, **35**: 125-31.

85. Kim S.-T., Sancar A., Essenmacher C. & Babcock G.T. (1993). Time-resolved EPR studies with DNA photolyase: excited-state $FADH^0$ abstracts an electron form Trp-306 to generate $FADH^-$, the catalytically active form of the cofactor. *Proc Natl Acad Sci USA*, **90**: 8023-8027.

86. Kirk J.T.O. (1994). Light and photosynthesis in aquatic ecosystems. Second ed. Cambridge University Press, New York, N.Y.

87. Klar T., Pokorny R., Moldt J., Batschauer A. & Essen L.O. (2007). Cryptochrome 3 from *Arabidopsis thaliana*: structural and functional analysis of its complex with a folate light antenna. *J Mol Biol*, **366**: 954-64.

88. Kleine T., Lockhart P. & Batschauer A. (2003). An *Arabidopsis* protein closely related to *Synechocystis* cryptochrome is targeted to organelles. *Plant J*, **35**: 93-103.

89. Kobayashi K., Kanno S., Smit B., van der Horst G.T., Takao M. & Yasui A. (1998). Characterization of photolyase/blue-light receptor homologs in mouse and human cells. *Nucleic Acids Res*, **26**: 5086-5092.

90. Kobayashi Y., Ishikawa T., Hirayama J., Daiyasu H., Kanai S., Toh H., Fukuda I., Tsujimura T., Terada N., Kamei Y., Yuba S., Iwai S. & Todo T. (2000). Molecular analysis of zebrafish photolyase/cryptochrome family: two types of cryptochromes present in zebrafish. *Genes Cells*, **5**: 725-38.

91. Komori H., Masui R., Kuramitsu S., Yokoyama S., Shibata T., Inoue Y. & Miki K. (2001). Crystal structure of thermostable DNA photolyase: pyrimidine-dimer recognition mechanism. *Proc Natl Acad Sci USA*, **98**: 13560-5.

92. Kooistra W.H.C.F., De Stefano M., Mann D.G. & Medlin L.K. (2003). The phylogeny of the diatoms. *Prog Mol Subcell Biol*, **33**: 59-97.

93. Kornblihtt A.R. (2005). Promoter usage and alternative splicing. *Curr Opin Cell Biol*, **17**: 262-8.

94. Kume K., Zylka M.J., Sriram S., Shearman L.P., Weaver D.R., Jin X., Maywood E.S., Hastings M.H. & Reppert S.M. (1999). mCRY1 and mCRY2 are essential components of the negative limb of the circadian clock feedback loop. *Cell*, **98**: 193-205.

95.   Leblanc C., Falciatore A., Watanabe M. & Bowler C. (1999). Semi-quantitative RT-PCR analysis of photoregulated gene expression in marine diatoms. *Plant Mol Biol*, **40**: 1031-44.

96.   Lee M.L., Kuo F.C., Whitmore G.A. & Sklar J. (2000). Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc Natl Acad Sci USA*, **97**: 9834-9.

97.   Lee R.E. & Kugrens P. (2000). Commentary: ancient atmospheric CO_2 and the timing of evolution of secondary endosymbioses. *Phycologia*, **39**: 167-172.

98.   Lévêque C., Marsaud V., Renoir J.M. & Sola B. (2007). Alternative cyclin D1 forms a and b have different biological functions in the cell cycle of B lymphocytes. *Exp Cell Res*, Apr 24; [Epub ahead of print].

99.   Levine J.S. & MacNichol Jr E.F. (1982). Color vision in fishes. *Scientific American*, **246**: 140-149.

100.   Li L. & Deng X.W. (2003). The COP9 signalosome: an alternative lid for the 26S proteasome? *Trends Cell Biol*, **13**: 507-509.

101.   Li Y.F. & Sancar A. (1990). Active site of *Escherichia coli* DNA photolyase: mutations at Trp277 alter the selectivity of the enzyme without affecting the quantum yield of photorepair. *Biochemistry*, **29**: 5698-706.

102.   Lin C. (2000). Plant blue-light receptors. *Trends Plant Sci*, **8**: 337-42.

103. Lin C., Robertson D.E., Ahmad M., Raibekas A.A., Jorns M.S., Dutton P.L. & Cashmore A.R. (1995). Association of flavin adenine dinucleotide with the *Arabidopsis* blue light receptor CRY1. *Science*, **269**: 968-70.

104. Lin C., Ahmad M., Chan J. & Cashmore A.R. (1996). CRY2: a second member of the *Arabidopsis* cryptochrome gene family. *Plant Physiol*, **110**: 1047-1048.

105. Lin C. & Shalitin D. (2003). Cryptochrome Structure and Signal Transduction. *Annual Review of Plant Biology*, **54**: 469-496.

106. Lin C. & Todo T. (2005). The cryptochromes. *Genome Biol*, **6**: 220.

107. Lin F.J., Song W., Meyer-Bernstein E., Naidoo N. & Sehgal A. (2001). Photic signaling by cryptochrome in the *Drosophila* circadian system. *Mol Cell Biol*, **21**: 7287-94.

108. Liscum E., Hodgson D.W. & Campbell T.J. (2003). Blue light signaling through the cryptochromes and phototropins. So that's what the blues is all about. *Plant Physiol*, **133**: 1429-36.

109. MacIntyre H.L., Kana T.M. & Geider R.J. (2000). The effect of water motion on short-term rates of photosynthesis by marine phytoplankton. *Trends Plant Sci*, **5**: 12-7.

110. Maheswari U., Montsant A., Goll J., Krishnasamy S., Rajyashri K.R., Patell V.M. & Bowler C. (2005). The Diatom EST Database. *Nucleic Acids Res*, **33**: D344-7.

111. Malhotra K., Kim S.T., Batschauer A., Dawut L. & Sancar A. (1995). Putative blue-light photoreceptors from *Arabidopsis thaliana* and *Sinapis alba* with a high degree of sequence homology to DNA photolyase contain the two photolyase cofactors but lack DNA repair activity. *Biochemistry*, **34**: 6892-9.

112. Mann D.G. (1993). Patterns of sexual reproduction in diatoms. *Hydrobiologia*, 269-270: 11-20.

113. Más P. (2005). Circadian clock signaling in *Arabidopsis thaliana*: from gene expression to physiology and development. *Int J Dev Biol*, **49**: 491-500.

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

115. Masuda S. & Bauer C.E. (2002). AppA is a blue light photoreceptor that antirepresses photosynthesis gene expression in *Rhodobacter sphaeroides*. *Cell*, **110**: 613-23.

116. Matsuoka D. & Tokutomi S. (2005). Blue light-regulated molecular switch of Ser/Thr kinase in phototropin. *Proc Natl Acad Sci USA*, **102**: 13337-42.

117. Matsuzaki M., Misumi O., Shin-I T., Maruyama S., Takahara M., Miyagishima S.Y., Mori T., Nishida K., Yagisawa F., Yoshida Y., Nishimura Y., Nakao S., Kobayashi T., Momoyama Y., Higashiyama T., Minoda A., Sano M., Nomoto H., Oishi K., Hayashi H., Ohta F., Nishizaka S., Haga S., Miura S., Morishita T., Kabeya Y., Terasawa K., Suzuki Y., Ishii Y., Asakawa S., Takano H., Ohta N., Kuroiwa H., Tanaka

K., Shimizu N., Sugano S., Sato N., Nozaki H., Ogasawara N., Kohara Y., & Kuroiwa T. (2004). Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature*, **428**: 653-7.

118. Miki T., Park J.A., Nagao K., Murayama N. & Horiuchi T. (1992). Control of segregation of chromosomal DNA by sex factor F in *Escherichia coli*. Mutants of DNA gyrase subunit A suppress letD (ccdB) product growth inhibition. *J Mol Biol.* **225**: 39-52.

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

120. Miyamoto Y. & Sancar A. (1998). Vitamin B2-based blue-light photoreceptors in the retinohypothalamic tract as the photoactive pigments for setting the circadian clock in mammals. *Proc Natl Acad Sci USA*, **95**: 6097-102.

121. Møller S.G., Kim Y.S., Kunkel T. & Chua N.H. (2003). PP7 is a positive regulator of blue light signaling in *Arabidopsis*. *Plant Cell*, **15**: 1111-9.

122. Montsant A., Allen A.E., Coesel S., De Martino A., Falciatore A., Heijde M., Jabbari K., Maheswari U., Mangogna M., Rayko E., Siaut M., Vardi A., Apt K.E., Berges J.B., Chiovitti A., Davis A.K., Hadi M.Z., Lane T.W., Lippmeier J.C., Martinez D., Schnitzler-Parker M., Pazour G.J., Saito M.A., Thamatrakoln K., Rokhsar D.S., Armbrust E.V. & Bowler, C. (2007). Identification and comparative genomic analysis of signaling and regulatory components in the diatom *Thalassiosira pseudonana*. *J of Phycology*, **43**: 585-604.

123. Motchoulski A. & Liscum E. (1999). *Arabidopsis* NPH3: A NPH1 photoreceptor-interacting protein essential for phototropism. *Science,* **286**: 961-4.

124. Nagel G., Ollig D., Fuhrmann M., Kateriya S., Musti A.M., Bamberg E. & Hegemann P. (2002). Channelrhodopsin-1: a light-gated proton channel in green algae. *Science,* **296**: 2395-8.

125. Nakasako M., Matsuoka D., Zikihara K. & Tokutomi S. (2005). Quaternary structure of LOV-domain containing polypeptide of *Arabidopsis* FKF1 protein. *FEBS Lett,* **579**: 1067–71.

126. Nambu J.R., Lewis J.O., Wharton K.A. Jr & Crews S.T. (1991). The *Drosophila* single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. *Cell,* **67**: 1157-67.

127. Naruse Y., Oh-hashi K., Iijima N., Naruse M., Yoshioka H. & Tanaka M. (2004). Circadian and light-induced transcription of clock gene Per1 depends on histone acetylation and deacetylation. *Mol Cell Biol,* **24**: 6278-87.

128. Ng W.O., Zentella R., Wang Y., Taylor J.S. & Pakrasi H.B. (2000). PhrA, the major photoreactivating factor in the cyanobacterium Synechocystis sp. strain PCC 6803 codes for a cyclobutane-pyrimidine-dimer-specific DNA photolyase. *Arch Microbiol,* **173**: 412-7.

129. Ninu L., Ahmad M., Miarelli C., Cashmore A.R. & Giuliano G. (1999). Cryptochrome 1 controls tomato development in response to blue light. *Plant J,* **18**: 551-556.

130. Norton T.A., Melkonian M. & Andersen R.A. (1996). Algal biodiversity. *Phycologia,* **35**: 308-26.

131. Ormo M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y. & Remington S.J. (1996). Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science,* **273**: 1392-5.

132. Osterlund M.T., Hardtke C.S., Wei N. & Deng X.W. (2000). Targeted destabilization of HY5 during light-regulated development of *Arabidopsis. Nature,* **405**: 462-6.

133. Oudot-Le Secq M.P., Loiseaux-de Goer S., Stam W.T. & Olsen J.L. (2006). Complete mitochondrial genomes of the three brown algae (Heterokonta: Phaeophyceae) *Dictyota dichotoma, Fucus vesiculosus* and *Desmarestia viridis. Curr Genet,* **49**: 47-58.

134. Palenik B., Grimwood J., Aerts A., Rouze P., Salamov A., Putnam N., Dupont C., Jorgensen R., Derelle E., Rombauts S., Zhou K., Otillar R., Merchant S.S., Podell S., Gaasterland T., Napoli C., Gendler K., Manuell A., Tai V., Vallon O., Piganeau G., Jancek S., Heijde M., Jabbari K., Bowler C., Lohr M., Robbens S., Werner G., Dubchak I., Pazour G.J., Ren Q., Paulsen I., Delwiche C., Schmutz J., Rokhsar D., Van de Peer Y., Moreau H. & Grigoriev I.V. (2007). The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. *Proc Natl Acad Sci USA,* **104**: 7705-10.

135. Park H.W., Kim S.T., Sancar A. & Deisenhofer J. (1995). Crystal structure of DNA photolyase from *Escherichia coli. Science,* **268**: 1866-72.

136. Partch C.L., Clarkson M.W., Ozgur S., Lee A.L. & Sancar A. (2005). Role of structural plasticity in signal transduction by the cryptochrome blue-light photoreceptor. *Biochemistry*, **44**: 3795-805.

137. Payne G., Heelis P.F., Rohrs B.R. & Sancar A. (1987). The active form of *Escherichia coli* DNA photolyase contains a fully reduced flavin and not a flavin radical, both *in vivo* and *in vitro*. *Biochemistry*, **26**: 7121-7.

138. Pickett-Heaps J., Schmid A.M. & Edgar L.A. (1990). The cell biology of diatom valve formation. In: *Progress in Phycological Research*, Vol . 7. Round F.E. & Chapman D.J. (Eds). Biopress Ltd., Bristol, UK, pp. 1-168.

139. Pondaven P., Gallinari M., Chollet S., Bucciarelli E., Sarthou G., Schultes S. & Jean F. (2007). Grazing-induced changes in cell wall silicification in a marine diatom. *Protist*, **158**: 21-8.

140. Poulsen N.C., Spector I., Spurck T.P., Schultz T.F. & Wetherbee R. (1999). Diatom gliding is the result of an actin-myosin motility system. *Cell Motil Cytoskeleton*, **44**: 23-33.

141. Provencio I., Jiang G., De Grip W.J., Hayes W.P. & Rollag M.D. (1998). Melanopsin: An opsin in melanophores, brain, and eye. *Proc Natl Acad Sci USA*, **95**: 340-5.

142. Provencio I., Rodriguez I.R., Jiang G., Hayes W.P., Moreira E.F. & Rollag M.D. (2000). A novel human opsin in the inner retina. *J Neurosci*, **20**: 600-5.

143. Quail P.H. (2002). Phytochrome photosensory signalling networks. *Nat Rev Mol Cell Biol*, **3**: 85-93.

144. Ragni M. & Ribera d'Alcalà M. (2004). Light as an information carrier underwater. *J Plankton Res*, **26**: 433-443.

145. Reddy P., Jacquier A.C., Abovich N., Petersen G. & Rosbash M. (1986). The period clock locus of *D. melanogaster* codes for a proteoglycan. *Cell*, **46**: 53-61.

146. Reinfelder J.R., Kraepiel A.M. & Morel F.M. (2000). Unicellular C4 photosynthesis in a marine diatom. *Nature*, **407**: 996-9.

147. Rensing L. & Ruoff P. (2002). Temperature effect on entrainment, phase shifting, and amplitude of circadian clocks and its molecular bases. *Chronobiol Int*, **19**: 807-64.

148. Reppert S.M. & Weaver D.R. (2001). Molecular analysis of mammalian circadian rhythms. *Annu Rev Physiol*, **63**: 647-76.

149. Reppert S.M. & Weaver D.R. (2002). Coordination of circadian timing in mammals. *Nature*, **418**: 935-41.

150. Rönneberg T. & Foster R.G. (1997). Twilight times: light and the circadian system. *Photochem Photobiol*, **66**: 549-61.

151. Rosato E., Codd V., Mazzotta G., Piccin A., Zordan M., Costa R. & Kyriacou C.P. (2001). Light-dependent interaction between Drosophila CRY and the clock protein PER mediated by the carboxy terminus of CRY. *Curr Biol*, **11**: 909-17.

152. Saitou N. & Nei M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*, **4**: 406-25.

153. Sakamoto K. & Nagatani A. (1996). Nuclear localization activity of phytochrome B. *Plant J,* **10**: 859-68.

154. Salomon M., Knieb E., von Zeppelin T. & Rudiger W. (2003). Mapping of low- and high-fluence autophosphorylation sites in phototropin 1. *Biochemistry,* **42**: 4217-4225.

155. Salomon M., Lempert U. & Rudiger W. (2004). Dimerization of the plant photoreceptor phototropin is probably mediated by the LOV1 domain. *FEBS Lett,* **572**: 8-10.

156. Sambrook J., Fritsch E.F. & Maniatis T. (1989). Molecular cloning: a Laboratory Manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

157. Sanada K., Harada Y., Sakai M., Todo T. & Fukada Y. (2004). Serine phosphorylation of mCRY1 and mCRY2 by mitogen-activated protein kinase. *Genes Cells,* **9**: 697-708.

158. Sancar A. (2003). Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors. *Chem Rev,* **103**: 2203-37.

159. Sancar A. (2004). Photolyase and cryptochrome blue-light photoreceptors. *Adv Protein Chem,* **69**: 73-100.

160. Sancar A. & Sancar G.B. (1984). *Escherichia coli* DNA photolyase is a flavoprotein. *J Mol Biol,* **172**: 223-7.

161. Sancar G.B. & Sancar A. (2006). Purification and characterization of DNA photolyases. *Methods Enzymol,* **408**: 121-56.

162. Sang Y., Li Q.H., Rubio V., Zhang Y.C., Mao J., Deng X.W. & Yang H.Q. (2005). N-terminal domain-mediated homodimerization is required for photoreceptor activity of Arabidopsis CRYPTOCHROME 1. *Plant Cell*, **17**: 1569-84.

163. Scala S., Carels N., Falciatore A., Chiusano M.L. & Bowler C. (2002). Genome Properties of the Diatom *Phaeodactylum tricornutum*. *Plant Physiol*, **129**: 993-1002.

164. Schwerdtfeger C. & Linden H. (2003). VIVID is a flavoprotein and serves as a fungal blue light photoreceptor for photoadaptation. *EMBO J*, **22**: 4846-55.

165. Selby C.P. & Sancar A. (1999). A third member of the photolyase/blue-light photoreceptor family in *Drosophila*: a putative circadian photoreceptor. *Photochem Photobiol*, **69**: 105-7.

166. Selby C.P. & Sancar A. (2006). A cryptochrome/photolyase class of enzymes with single-stranded DNA-specific photolyase activity. *Proc Natl Acad Sci USA*, **103**: 17696-700.

167. Selby C.P., Thompson C., Schmitz T.M., Van Gelder R.N. & Sancar A. (2000). Functional redundancy of cryptochromes and classical photoreceptors for nonvisual ocular photoreception in mice. *Proc Natl Acad Sci USA*, **97**: 14697-702.

168. Senger H. (1980). The Blue Light Syndrome. Springer-Verlag, Berlin.

169. Shalitin D., Yang H., Mockler T.C., Maymon M., Guo H., Whitelam G.C. & Lin C. (2002). Regulation of *Arabidopsis* cryptochrome 2 by blue-light-dependent phosphorylation. *Nature*, **417**: 763-7.

170. Siaut M., Heijde M., Mangonga M., Montsant A., Coesel S., Allen A., Falciatore A. & Bowler C. (2007). Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. *Gene*, in press.

171. Sineshchekov O.A. & Govorunova E.G. (2001). Rhodopsin receptors of phototaxis in green flagellate algae. *Biochemistry (Mosc)*, **66**: 1300-10.

172. Sineshchekov O.A., Jung K.H. & Spudich J.L. (2002). Two rhodopsins mediate phototaxis to low- and high-intensity light in *Chlamydomonas reinhardtii*. *Proc Natl Acad Sci USA*, **99**: 8689-94.

173. Smetacek V. (1999). Diatoms and the ocean carbon cycle. *Protist*, **150**: 25-32.

174. Smetacek V. (2001). A watery arms race. *Nature*, **411**: 745.

175. Song S.H., Dick B., Penzkofer A., Pokorny R., Batschauer A. & Essen L.O. (2006). Absorption and fluorescence spectroscopic characterization of cryptochrome 3 from *Arabidopsis thaliana*. *J Photochem Photobiol B*, **85**: 1-16.

176. Spalding E.P. (2000). Ion channels and the transduction of light signals. *Plant Cell Environ*, **23**: 665-74.

177. Sprenger W.W., Hoff W.D., Armitage J.P. & Hellingwerf K.J. (1993). The eubacterium *Ectothiorhodospira halophila* is negatively phototactic, with a

wavelength dependence that fits the absorption spectrum of the photoactive yellow protein. *J Bacteriol*, **175**: 3096-104.

178. Stanewsky R., Kaneko M., Emery P., Beretta B., Wager-Smith K., Kay S.A., Rosbash M. & Hall J.C. (1998). The *cry*[b] mutation identifies cryptochrome as a circadian photoreceptor in *Drosophila*. *Cell*, **95**: 681-92.

179. Steinbrenner J. & Linden H. (2003). Light induction of carotenoid biosynthesis genes in the green alga *Haematococcus pluvialis*: regulation by photosynthetic redox control. *Plant Mol Biol*, **52**: 343-356.

180. Suter B., Livingstone-Zatchej M. & Thoma F. (1997). Chromatin structure modulates DNA repair by photolyase *in vivo*. *EMBO J*, **16**: 2150-60.

181. Tamada T., Kitadokoro K., Higuchi Y., Inaka K., Yasui A., de Ruiter P.E., Eker A.P. & Miki K. (1997). Crystal structure of DNA photolyase from *Anacystis nidulans*. *Nat Struct Biol*, **4**: 887-91.

182. Taylor J.-S. & Cohrs M.P. (1987). DNA, light and Dewar pyrimidinones: the structure and biological significance of TpT3. *J Am Chem Soc*, **109**: 2834-2835.

183. Thompson J.D., Higgins D.G. & 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 Acids Res*, **22**: 4673-80.

184. Todo T., Takemori H., Ryo H., Ihara M., Matsunaga T., Nikaido O., Sato K. & Nomura T. (1993). A new photoreactivating enzyme that

specifically repairs ultraviolet light-induced (6-4)photoproducts. *Nature*, **361**: 371-4.

185. Todo T., Ryo H., Yamamoto K., Toh H., Inui T., Ayaki H., Nomura T. & Ikenaga M. (1996). Similarity among the Drosophila (6-4)photolyase, a human photolyase homolog, and the DNA photolyase-blue-light photoreceptor family. *Science*, **272**: 109-12.

186. Ueda T., Kato A., Kuramitsu S., Terasawa H. & Shimada I. (2005). Identification and characterization of a second chromophore of DNA photolyase from *Thermus thermophilus* HB27. *J Biol Chem*, **280**: 36237-36243.

187. Vaistij F.E., Boudreau E., Lemaire S.D., Goldschmidt-Clermont M. & Rochaix J.D. (2000). Characterization of Mbb1, a nucleus-encoded tetratricopeptide-like repeat protein required for expression of the chloroplast *psbB/psbT/psbH* gene cluster in *Chlamydomonas reinhardtii*. *Proc Natl Acad Sci USA*, **97**: 14813-8.

188. Valverde F., Mouradov A., Soppe W., Ravenscroft D., Samach A. & Coupland G. (2004). Photoreceptor Regulation of CONSTANS Protein in Photoperiodic Flowering. *Science*, 303: 1003-1006.

189. Van Den Hoeck C., Mann D.G. & Johns H.M. (1997). Algae. An Introduction to Phycology. Cambridge, UK: Cambridge Univ. Press.

190. van der Horst G.T., Muijtjens M., Kobayashi K., Takano R., Kanno S., Takao M., de Wit J., Verkerk A., Eker A.P., van Leenen D., Buijs R., Bootsma D., Hoeijmakers J.H. & Yasui A. (1999). Mammalian Cry1 and

Cry2 are essential for maintenance of circadian rhythms. *Nature*, **398**: 627-30.

191. van der Horst M.A. & Hellingwerf K.J. (2004). Photoreceptor proteins, "star actors of modern times": a review of the functional dynamics in the structure of representative members of six different photoreceptor families. *Acc Chem Res*, **37**: 13-20.

192. Van Gelder R.N., Wee R., Lee J.A. & Tu D.C. (2003). Reduced pupillary light responses in mice lacking cryptochromes. *Science*, **299**: 222.

193. Vande Berg B.J. & Sancar G.B. (1998). Evidence for dinucleotide flipping by DNA photolyase. *J Biol Chem*, **273**: 20276-84.

194. Vaulot D., Olson R.J. & Chisholm S.W. (1986). Light and dark control of the cell cycle in two marine phytoplankton species. *Exp Cell Res*, **167**: 38-52.

195. Venter J.C., Remington K., Heidelberg J.F., Halpern A.L., Rusch D., Eisen J.A., Wu D., Paulsen I., Nelson K.E., Nelson W., Fouts D.E., Levy S., Knap A.H., Lomas M.W., Nealson K., White O., Peterson J., Hoffman J., Parsons R., Baden-Tillson H., Pfannkoch C., Rogers Y.H. & Smith H.O. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, **304**: 66-74.

196. Villareal T.A. (1989). Division cycles in the nitrogen-fixing *Rhizozlenia* (Bacillariophyceae)-*Richelia* (Nostocaceae) symbiosis. *Br Phycol J*, **24**: 357-65.

197. Villareal T.A. & Carpenter E.J. (2003). Buoyancy regulation and the potential for vertical migration in the oceanic cyanobacterium trichodesmium. *Microb Ecol*, **45**: 1-10.

198. Vitaterna M.H., Selby C.P., Todo T., Niwa H., Thompson C., Fruechte E.M., Hitomi K., Thresher R.J., Ishikawa T., Miyazaki J., Takahashi J.S. & Sancar A. (1999). Differential regulation of mammalian period genes and circadian rhythmicity by cryptochromes 1 and 2. *Proc Natl Acad Sci USA*, **96**: 12114-9.

199. Wada M., Kagawa T. & Sato Y. (2003). Chloroplast movement. *Annu Rev Plant Biol*, **54**: 455-68.

200. Walhout A.J., Temple G.F., Brasch M.A., Hartley J.L., Lorson M.A., van den Heuvel S. & Vidal M. (2000). GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol*, **328**: 575-92.

201. Walsh C.T. (1986). Naturally occurring 5-deazaflavin coenzymes: biological redox roles. *Acc Chem Res*, **19**: 216-221.

202. Weber G. (1959). Fluorescence of riboflavin and flavin adenine dinucleotide. *Biochem J*, **47**: 114-121

203. Worthington E.N., Kavakli I.H., Berrocal-Tito G., Bondo B.E. & Sancar A. (2003). Purification and characterization of three members of the photolyase/cryptochrome family glue-light photoreceptors from *Vibrio cholerae*. *J Biol Chem*, **278**: 39143-54.

204. Yang H.Q., Wu Y.J., Tang R.H., Liu D., Liu Y. & Cashmore A.R. (2000). The C termini of *Arabidopsis* cryptochromes mediate a constitutive light response. *Cell*, **103**: 815-27.

205. Yang H.Q., Tang R.H. & Cashmore A.R. (2001). The signaling mechanism of Arabidopsis CRY1 involves direct interaction with COP1. *Plant Cell*, **13**: 2573-87.

206. Yang J., Lin R., Sullivan J., Hoecker U., Liu B., Xu L., Deng X.W. & Wang H. (2005). Light regulates COP1-mediated degradation of HFR1, a transcription factor essential for light signaling in Arabidopsis. *Plant Cell*, **17**: 804-21.

207. Yanovsky M.J. & Kay S.A. (2003). Living by the calendar: how plants know when to flower. *Nat Rev Mol Cell Biol*, **4**: 265-75.

208. Yasui A., Eker A.P., Yasuhira S., Yajima H., Kobayashi T., Takao M. & Oikawa A. (1994). A new class of DNA photolyases present in various organisms including aplacental mammals. *EMBO J*, **13**: 6143-51.

209. Yu X., Shalitin D., Liu X., Maymon M., Klejnot J., Yang H., Lopez J., Zhao X., Bendehakkalu K.T. & Lin C. (2007). Derepression of the NC80 motif is critical for the photoactivation of *Arabidopsis* CRY2. *Proc Natl Acad Sci USA*, **104**: 7289-94.

210. Zhao S. & Sancar A. (1997). Human blue-light photoreceptor hCRY2 specifically interacts with protein serine/threonine phosphatase 5 and modulates its activity. *Photochem Photobiol*, **66**: 727-31.

211. Zhao X., Liu J., Hsu D.S., Zhao S., Taylor J.S. & Sancar A. (1997). Reaction mechanism of (6-4) photolyase. *J Biol Chem*, **272**: 32580-90.

212. Zhou Y., Zhou C., Ye L., Dong J., Xu H., Cai L., Zhang L. & Wei L. (2003). Database and analyses of known alternatively spliced genes in plants. *Genomics*, **82**: 584-95.

213. Zhu H., Conte F. & Green C.B. (2003). Nuclear localization and transcriptional repression are confined to separable domains in the circadian protein CRYPTOCHROME. *Curr Biol*, **13**: 1653-8.

214. Zurzolo C. & Bowler C. (2001). Exploring bioinorganic pattern formation in diatoms. A story of polarized trafficking. *Plant Physiol*, **127**: 1339-45.